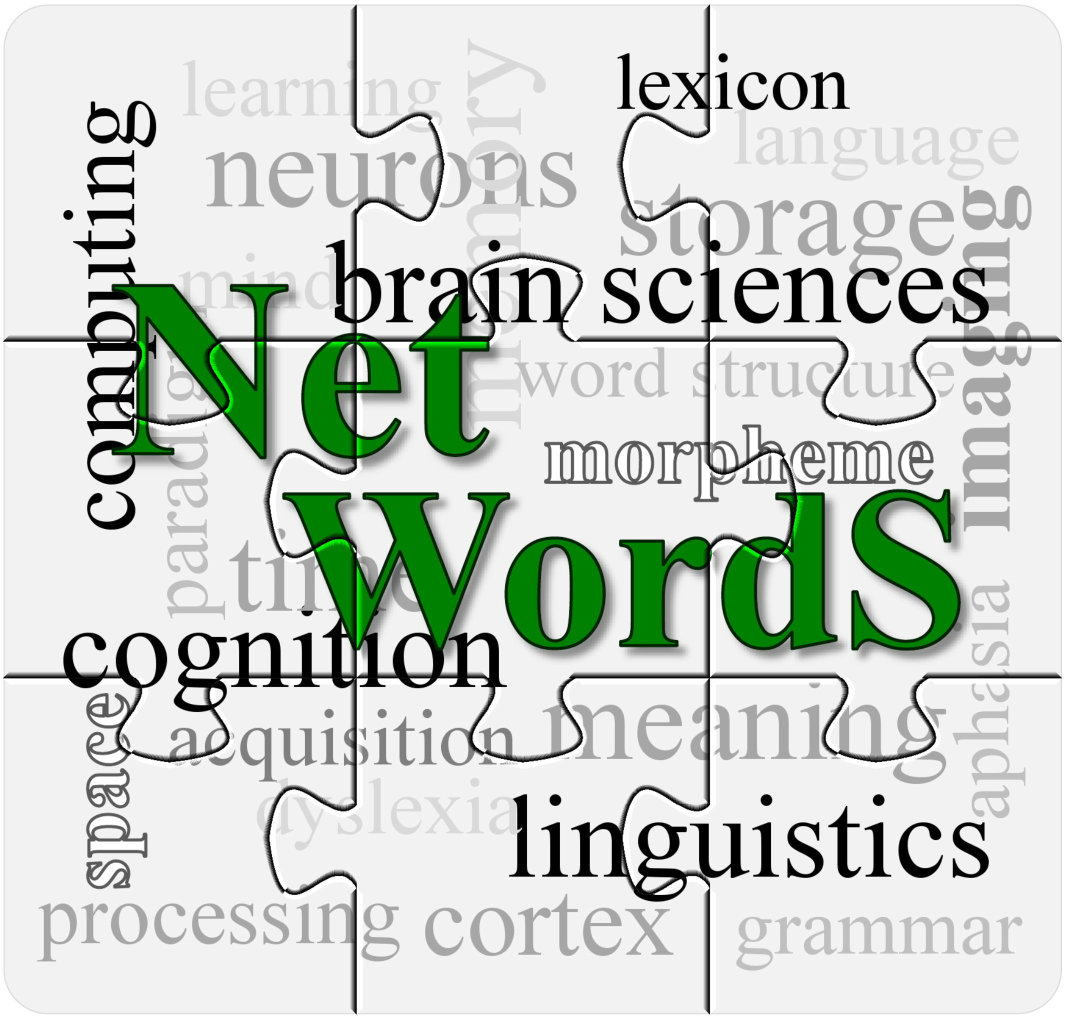


NetWordS 2015

Word Knowledge and Word Usage

Representations and Processes in the Mental Lexicon

March 30th - April 1st, 2015
Scuola Normale Superiore, Pisa - Italy

CONFERENCE PROCEEDINGS

http://www.networds-esf.eu/

SCUOLA NORMALE SUPERIORE

Istituto di Linguistica Computazionale

ComPhys physiology of communication

European Science Foundation

Vito Pirrelli, Claudia Marzi, and Marcello Ferro (eds.)

NetWordS 2015

Word Knowledge and Word Usage

Representations and Processes in the Mental Lexicon

Pisa, Italy, March 30th - April 1st, 2015
Conference proceedings

Editors' address:
Consiglio Nazionale delle Ricerche
Istituto di Linguistica Computazionale
via G. Moruzzi, 1
56124 Pisa, Italy

{vito.pirrelli, claudia.marzi, marcello.ferro}@ilc.cnr.it

Table of contents

Foreword

This international conference "Word Knowledge and Word Usage: Representations and processes in the mental lexicon" is the final outcome of 4 years of intense multi-disciplinary research networking and cooperation funded by the European Science Foundation within the framework of the NetWordS programme (May 2011 - April 2015).

NetWordS' mission was to bring together experts of various research fields (from brain sciences and computing to cognition and linguistics) and of different theoretical inclinations, to advance the current awareness of theoretical, typological, psycholinguistic, computational and neurophysiological evidence on the structure and processing of words, with a view to developing novel research paradigms and bringing up a new generation of language scholars. The conference was intended to provide a first forum for assessing current progress of cross-disciplinary research on language architecture and usage, and discussing prospects of future synergy.

People are known to memorise, parse and access words in a context-sensitive and opportunistic way, by caching their most habitual and productive processing patterns into routinized behavioural schemes. Speakers not only take advantage of token-based information such as frequency of individual, holistically stored words, but they are also able to organise stored words through paradigmatic structures (or word families) whose overall size and frequency is an important determinant of ease of lexical access and interpretation. Accordingly, lexical organisation is not necessarily functional to descriptive economy and minimisation of storage, but to more performance-oriented factors such as efficiency of memorisation, access and recall. Usage-based approaches to word processing lend support to this view, to promote explanatory frameworks that aim to investigate the stable correlation patterns linking distributional entrenchment of lexical units with productivity, internal structure and ease of interpretation. Ultimately, this is intended to establish a deep interconnection between performance-oriented, low-level lexical functions such as memorisation, rehearsal, access and recall, and their neuroanatomical correlates.

The impressive wealth of data and approaches reported in 23 oral presentations and 19 posters (selected from 84 original submissions), and the broader perspectives broached by Wolfgang U. Dressler, Marta Kutas, Gabriella Vigliocco and Michael Zock, provided compelling evidence that the time has now come for this area to make a significant methodological leap towards tighter and targeted synergy. The overall conference message was clear. Interdisciplinarity should be coupled with both theoretical modelling and quantitative analysis of empirical evidence. Any truly interdisciplinary effort must take advantage of the many methodological caveats that psycholinguists, neurolinguists, theoretical and cognitive linguists, historical linguists, typologists and computational linguists have developed over many years of relatively independent work. Integration of their data and approaches will necessary mean more complex models, far more constrained, explanatory and comprehensive than any other account put forward so far. There is general consensus that joining forces in this research area will not only lead to considerable progress in our theoretical understanding of the physiology of communication, but will also be conducive to more effective ways to help real people engaged in their daily communicative exchanges.

The conference was held in Pisa at the Scuola Normale Superiore, between March 30th and April 1st 2015, and benefited from the invaluable support and advice of Prof. Pier Marco Bertinetto and his team, to whom our warmest thanks go.

Pisa, April 2nd, 2015

Vito Pirrelli
Claudia Marzi
Marcello Ferro

Psycholinguistic illusions in and on morphology

Wolfgang U. Dressler
University of Vienna
wolfgang.dressler@univie.ac.at

Fruitful interdisciplinary contact between specialists in theoretical morphology and in various branches of psycholinguistics (my examples will come from acquisition, processing, aphasia) is hampered by reciprocal illusions, some of them rarely criticised explicitly. Often ecological validity is dubious.

The bridge of iconicity: from a world of experience to experience of language

Gabriella Vigliocco
University College London
g.vigliocco@ucl.ac.uk

Arbitrariness between linguistic form and meaning is taken as foundational in language studies and the question of how linguistic form links to meaning is central to language development, processing and evolution. But, languages also display iconicity in addition to arbitrariness. This is especially evident in sign languages. This, what if the study of language started from signed rather than spoken languages? In the talk I will explore this question.

Needles in a haystack and how to find them. Can neuroscientists, psychologists and computational linguist help us (to build a tool) to overcome the "tip of the tongue" problem?

Michael Zock
LIF-CNRS, University of Marseille
michael.zock@lif.univ-mrs.fr

Whenever we speak, read or write we always use words, the exchange money of concepts they are standing for. No doubt, words ARE important. Yet having stored "words" does not guarantee that we can access them under all circumstances. Some forms may refuse to come to our mind when we need them most, the moment of speaking or writing. This is when we tend to reach for a dictionary, hoping to find the token we are looking for.

The problem is that most dictionaries, be they in paper or electronic form, are not well suited to support the language producer. Hence the questions, why is this so and what does it take to enhance existing resources? Can we draw on what is known about the human brain or its externalized form (texts)? Put differently, what kind of help can we expect by looking at the work done by neuroscientists, psycholinguists or computational linguistics? These are some of the questions I will briefly touch upon, by ending with a concrete proposal (roadmap), outlining the majors steps to be performed in order to enhance an existing electronic resource.

Content and organization of knowledge and its use in language comprehension

Marta Kutas
University of California, San Diego
mkutas@ucsd.edu

Significant work takes place at the language-memory interface that supports word and sentence processing. Both the content and the functional organization of our world knowledge impact language comprehension in real time. Each cerebral hemisphere is involved, albeit in different ways. The nature of knowledge organization (associative, categorical, events, perceptuomotor) and their use in predictive and/or integrative language processing have been revealed via investigations employing event-related brain potentials (ERPs). I will review some of our electrophysiological work supporting the idea that language processing is immediate and incremental, contextual, sometimes predictive, multi-modal, and bi-hemispheric.

Implicative structure and joint predictiveness

Olivier Bonami
Université Paris-Sorbonne
Laboratoire de linguistique formelle
(U. Paris Diderot & CNRS)
olivier.bonami@paris-sorbonne.fr

Sacha Beniamine
Université Paris Diderot
Laboratoire de linguistique formelle &
Alpage (Inria & U. Paris Diderot)

1 Introduction

(Ackerman et al., 2009) define the PARADIGM CELL FILLING PROBLEM (PCFP), which we paraphrase in (1), as the cornerstone of the study of inflectional paradigms.

(1) How do speakers know how to inflect the full paradigm of a lexeme on the basis of exposure to only some of its forms?

(Ackerman et al., 2009) go on to argue that speakers rely on knowledge of the IMPLICATIVE STRUCTURE of paradigms (Wurzel, 1984): paradigms are structured in such a way that there are reliable correlations between the form filling one paradigm cell A and the form filling another cell B. The reliability of these correlations depends on the particular pair of cells A and B under scrutiny; it can be assessed quantitatively by examining the statistical distribution of operations required to go from A to B in the lexicon.

This presentation focuses on one particular aspect of implicative structure, which we call JOINT PREDICTIVENESS. In some situations, joint knowledge of two paradigm cells A and B provides more information on cell C than could be inferred from knowledge of either A or B. Table 1 below provides a simple example from French, using lexemes illustrating 7 patterns corresponding to of 95% of the verbs documented in the *Flexique* phoneticized lexicon (Bonami et al., 2014). In French conjugation, predicting the past participle from the infinitive is hard, because of the opacity between second conjugation infinitives, such as BÂTIR, and some third conjugation infinitives, such as TENIR, OUVRIR, MOURIR. Predicting the past participle from present SG forms is also hard, this time because some first conjugation verbs with a stem in *-i* (e.g. RELIER) are not distinguished from second conjugation verbs. A different subset of first conjugation verbs (e.g. RATISSER) raises similar problems for PL forms.

Overall, no other cell in the paradigm is a very good predictor of the past participle. However, joint knowledge of some pairs of paradigm cells radically improves the quality of prediction. For instance, joint knowledge of the infinitive and some present plural form removes all uncertainty in the sample in Table 1: knowledge of the infinitive form partitions the set of lexemes in two classes within which the PRS.3PL is fully predictive of the past participle.

Although the existence of joint predictiveness is acknowledged in the literature (Matthews, 1972; Thymé et al., 1994; Ackerman et al., 2009; Stump and Finkel, 2013; Blevins, forthcoming; Sims, forthcoming), little attention has been given to quantifying its importance. In this paper we first give further arguments that joint predictiveness is a crucial aspect of implicative structure, and that a careful empirical examination of joint predictiveness is essential to both linguistic and psycholinguistic assessment of the PCFP and related issues. We then propose and illustrate a method for the quantitative evaluation of joint predictiveness. We end with a discussion of principal part systems.

2 The relevance of joint predictiveness

We start by establishing that speakers do have the opportunity to use joint predictiveness. Figure 1 plots how the number of forms per lemma evolves when walking through the 1.6 billion words of the *FrWaC* web corpus (Baroni et al., 2009), restricting attention to the 6847 verbs documented in the *Lefff* lexicon (Sagot, 2010) to compensate for tagging errors.[1] Note that 1.6 billion words is

[1] Note that this restriction leads to overestimating the average number of forms per lemma, as neologisms, very rare words and hapaxes not present in the lexical resource are not included. We are counting distinct forms rather than distinct paradigm cells, as there is currently no tagger for French that reliably disambiguates homographic forms of the same lexeme. French verbs have 51 paradigm cells, and the average number of distinct forms per verb in the *Lefff* lexicon is 35.8.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

Lexeme	INF	PRS.3SG	PRS.3PL	PST.PTCP	#
LIVRER 'deliver'	livʁe	livʁ	livʁ	livʁe	4108
RELIER 'link'	ʁəlje	ʁəli	ʁəli	ʁəlje	210
RATISSER 'rake'	ʁatise	ʁatis	ʁatis	ʁatise	22
BÂTIR 'build'	batiʁ	bati	batis	bati	327
TENIR 'hold'	təniʁ	tjɛ̃	tjɛn	təny	37
OUVRIR 'open'	uvʁiʁ	uvʁ	uvʁ	uvɛʁ	8
MOURIR 'die'	muʁiʁ	mœʁ	mœʁ	mɔʁ	1

Table 1: Exemplary paradigms for inflection patterns for 4-cell subparadigms of French verbs (data from *Flexique* — 5% of the lemmas illustrating minor patterns have been excluded)

in the order of magnitude of the overall linguistic exposure of an adult speaker. The distribution strongly suggests that, as speakers get exposed to more words, paradigms fill slowly on average, so that predicting unknown forms stays relevant; at the same time, speakers are massively exposed to multiple forms of the same lexemes, which makes knowledge of joint predictiveness relevant to addressing the PCFP.

A second relevant observation is that speakers do manifest knowledge of joint predictiveness. Although this topic deserves dedicated experimental studies that are beyond the scope of this paper, circumstantial evidence from speech errors is easy to find. One common conjugation error in French (Kilani-Schoch and Dressler, 2005) is to use *mouru* as the past participle of MOURIR, whereas *mouri* is almost never used (140 relevant occurrences of *mouru* in the full *FrWaC* corpus, 0 or *mouri*). This would be surprising if speakers were analogizing from a single paradigm cell: given knowledge of the sole infinitive, *mouri* would be the most likely regularization; given knowledge of some present form, *mour* or *meur* would be expected.[2] Thus the property speakers seem to be sensitive to is the existence of an allomorphic relation between the infinitive and the present stem—hence, employing joint predictiveness from two cells to infer the likely form of the participle.

The final observation is that there are important linguistic generalizations that can only be obtained by looking at joint predictiveness. To supplement the French data presented in the introduction, let us consider a spectacular example from European Portuguese, concerning the prediction of the form of the infinitive from those of the present singular. Table 2 presents relevant data. Because it does not contain a theme vowel, the present 1SG is a bad predictor of the infinitive: a priori, any present 1SG could correspond to a first, second or third conjugation verb. 2SG and 3SG forms are slightly better predictors, as they distinguish first conjugation endings (-ɐʃ,-ɐ) from second/third conjugation endings (-əʃ,-ə); the distinction between the two last conjugations is still neutralized. However, if a verb has a mid prethematic vowel in the 2SG and 3SG, the shape of that vowel is raised to high-mid in the 1SG in the second conjugation (witness RECEBER, RECORRER), and to high in the third conjugation (witness SEGUIR, SUBIR). Whether one sees this phenomenon as the result of a synchronic vowel harmony in the 1SG operating prior to theme vowel deletion (Mateus and d'Andrade, 2000) or as a historical remnant with no synchronic motivation, it remains that on the surface, for verbs with a mid prethematic vowel in the 2SG and 3SG, knowledge of the 1SG disambiguates whether the verb belongs to the second or third conjugation and thus helps predict the infinitive.

3 Quantifying joint predictiveness

To assess the importance of joint predictiveness, we build on previous proposals by (Bonami and Boyé, 2014) and (Bonami and Luís, 2014) on the evaluation of predictiveness from a single paradigm cell, themselves improving on (Acker-

[2] A reviewer points out that if speech errors are due to analogy to the nearest (frequent) neighbor, *mouru* is unsurprising, as *courir* (past participle *couru*) is the most frequent of the verbs whose infinitive is at a minimal edit distance from *mourir*. This assumption however is not plausible. Witness the case of the verb *dire*, whose present 2PL *dites* is very commonly overregularized to *disez*. The most frequent phonological neighbor of *dire* is *lire*; however, according to the *lexique* database (New et al., 2007), *dire* is 8 times more frequent than *lire* in written French, and 17 times in spoken French. It is thus not plausible that analogical regularization is driven by the closest neighbor; rather, it is driven by general patterns applying across lexemes—for instance, *dire* is one of a handful of exceptions to the regular $Xons \sim Xez$ alternation between 1PL and 2PL, that is overwhelmingly prevalent both in type and token frequency.

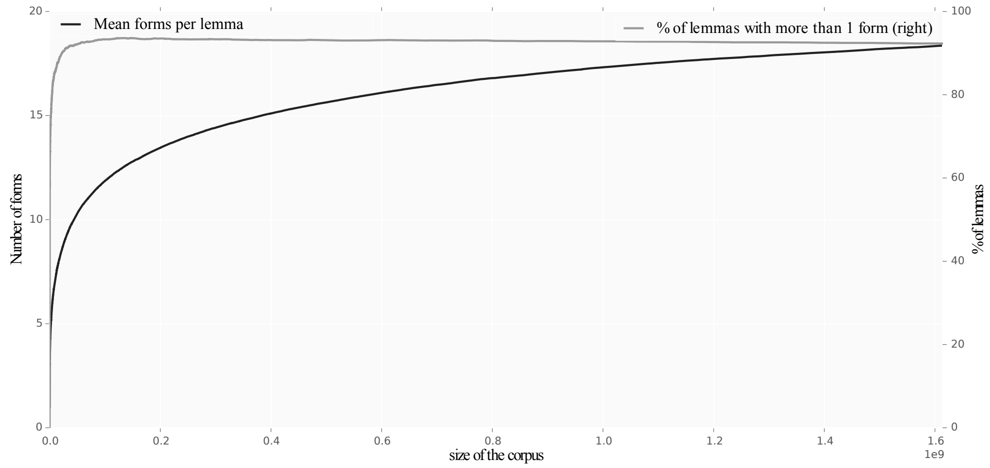


Figure 1: Mean number of forms per lemma and proportion of lemmas with multiple forms as a function of vocabulary size (*FrWaC* corpus)

	INF	1SG	2SG	3SG	1PL	2PL	3PL
LEVAR	ləˈvaɾ	ˈlɛvu	ˈlɛvɐʃ	ˈlɛvɐ	ləˈvɐmuʃ	ləˈvaiʃ	ˈlɛvɐ̃u
NOTAR	nuˈtaɾ	ˈnɔtu	ˈnɔtɐʃ	ˈnɔtɐ	nuˈtɐmuʃ	nuˈtaiʃ	ˈnɔtɐ̃u
RECEBER	rəsəˈbeɾ	rəˈsebu	rəˈsɛbəʃ	rəˈsɛbə	rəsəˈbemuʃ	rəsəˈbɐiʃ	rəˈsɛbɐ̃ĩ
RECORRER	rəkuˈreɾ	rəˈkoru	rəˈkɔrəʃ	rəˈkɔrə	rəkuˈremuʃ	rəkuˈrɐiʃ	rəˈkɔrɐ̃ĩ
SEGUIR	səˈgiɾ	ˈsigu	ˈsɛgəʃ	ˈsɛgə	səˈgimuʃ	səˈgiʃ	ˈsɛgɐ̃ĩ
SUBIR	suˈbiɾ	ˈsubu	ˈsɔbəʃ	ˈsɔbə	suˈbimuʃ	suˈbiʃ	ˈsɔbɐ̃ĩ

Table 2: Selected European Portuguese verbs in the infinitive and present indicative

man et al., 2009) and (Ackerman and Malouf, 2013). Specifically, for every pair of paradigm cells A and B, we infer a classification of patterns of alternation relating these two cells. These patterns are then used to define a random variable $A{\sim}B$ over pairs of forms corresponding to the distribution of patterns, and a random variable $A_{A\sim B}$ classifying possible form for A on the basis of the patterns they could possibly instantiate. For instance, going back to the data in Table 1, INF∼PST.PTCP partitions the set of pairs in 5 subsets corresponding to the patterns Xe∼Xe, Xiʁ∼Xi, Xiʁ∼Xy, Xʁiʁ∼Xɛʁ and Xuʁiʁ∼Xɔʁ, while $\text{INF}_{\text{INF}\sim\text{PST.PTCP}}$ partitions the set of infinitive forms in 4 sets, depending on whether they end in -e, -uʁiʁ, -Vʁiʁ with $V \neq$ u, or -Xiʁ with $X \neq$ ʁ.

$H(A{\sim}B \mid A_{A\sim B})$, the conditional entropy of the pattern relating A and B given relevant features of the form filling A, evaluates how well A predicts B.

Crucial to this computation is the choice of a strategy of exhaustive classification of patterns of alternation between pairs of forms. Since the design of an algorithm finding an optimal such classification from raw data is an open research question,[3] we opportunistically use the algorithm sketched in (2) that we know to give satisfactory results for the languages at hand.

(2) a. For any pair of strings $\langle\phi_1, \phi_2\rangle$, find strings $\alpha, \gamma, \beta_1, \beta_2, \delta_1$ and δ_2 such that $\phi_1 = \alpha\beta_1\gamma\delta_1$ and $\phi_2 = \alpha\beta_2\gamma\delta_2$, where β_1 and β_2 have the same length; segments in β_1 and β_2 (resp. δ_1 and δ_2) match in category (vowel vs. consonant), starting from the left; and the length of α is maximal. Classify the pair as instantiating pattern $[X\beta_1 Y\delta_1{\sim}X\beta_2 Y\delta_2 \;/\; \alpha\_\gamma\_]$.

b. For all patterns instantiating the same alternation $[x{\sim}y \,/\, \alpha_1\_\gamma_1\_], \ldots, [x{\sim}y \,/\, \alpha_n\_\gamma_n\_]$, determine maximally specific feature descriptions of sets of strings $\{\alpha_1, \ldots, \alpha_n\}$

[3]The problem can be presented as that of finding, for any set of pairs of forms, a minimal set of subsequential finite-state transducers such that one of the transducers maps each input form to the correct output. Even if that problem were solved, it is entirely possible for there to be more than one such minimal set, leading to competing classifications of the pairs and thus to different assessments of predictiveness.

and $\{\gamma_1, \ldots, \gamma_n\}$, using (Albright, 2002)'s Minimal Generalization strategy.

Joint predictiveness can then be assessed looking at joint random variables: predicting C from A and B is evaluated by (3): we assess the uncertainty associated with predicting both the pattern relating A to C and the pattern relating B to C, given knowledge of relevant properties of A, relevant properties of B, and the pattern relating A and B. Notice that this easily generalizes to prediction given joint knowledge of n different cells.

(3) $H(A{\sim}C, B{\sim}C \mid A_{A\sim C}, B_{B\sim C}, A{\sim}B)$

Table 3 shows the average entropy from 1 or 2 cells for 5000 French verbs and 2000 European Portuguese verbs respectively.[4] In both languages, knowing a second cell significantly reduces uncertainty on average.

# of predictor cells	French	Portuguese
1	0.1670	0.1649
2	0.0540	0.0818

Table 3: Average conditional entropy when predicting from 1 or 2 cells

4 Principal part systems

A system of principal parts is a set of paradigm cells such that knowledge of the forms filling these cells is sufficient to derive the rest of the paradigm (Hockett, 1967; Matthews, 1972; Finkel and Stump, 2007; Stump and Finkel, 2013).[5] The validity of a principal part system thus rests on the existence of systematic categorical joint predictiveness; and the evaluation method outlined in the preceding section may be used to infer sets of principal parts.

Exploring this issue on the European Portuguese dataset, we find that there are 177 such systems for Portuguese. All these systems include a form with a 3-way contrast of theme vowels, such as the infinitive, and a form with stress on the prethematic vowel, such as the present 3SG. This corresponds to the observation in (Bonami and Luís, 2014) that such pairs of cells have complementary predictive power. The sheer number of alternative principal part systems highlights the arbitrariness of the choice of a particular set of principal parts (Matthews, 1972; Ackerman et al., 2009; Blevins, forthcoming).

Turning to French, we found no set of principal parts of cardinality 2, as already observed by (Stump and Finkel, 2013). This is testament to the prevalence of erratic stem allomorphy in French conjugation, leading to numerous situations of unpredictibility local to a small subpart of the paradigm (Bonami and Boyé, 2002). However, this observation should be modalized in two ways.

First, our method yields 396 sets of principal parts of cardinality 3, whereas (Stump and Finkel, 2013) found no set of cardinality smaller than 5. This difference seems to be due to the fact that, under the methodology used here, the applicability of a pattern of alternation is sensitive to phonotactic properties of the stem (thanks to the use of the Minimal Generalization strategy in (2b)), whereas (Stump and Finkel, 2013) only look at exponence. Arguably then, the present method provides a superior evaluation of the diagnostic value of paradigm cells.

Second, although there is no pair of cells with categorical diagnostic value, some come very close. There are 25 pairs of cells (among which pairs of very frequent cells such as the present 3PL and the infinitive) such that predicting any other cell from this pair yields an entropy below 0.005. This means that given knowledge of these two cells, trying to guess any other cell will be about as hard as predicting an event with a 99.95% probability of occurrence.[6] This casts doubts both on the pedagogical value of categorical principal part systems and on the usefulness of principal part systems, as opposed to graded evaluations of joint predictiveness, for the study of morphological competence.

Acknowledgments

This work was partially supported by a public grant overseen by the French National Research

[4]The French dataset was extracted from *Flexique* (Bonami et al., 2014). The Portuguese dataset was derived from the University of Coimbra pronunciation dictionary (Veiga et al., 2012) for the purpose of (Bonami and Luís, 2013).

[5]We focus here on traditional static principal part systems. See (Bonami and Boyé, 2007; Finkel and Stump, 2007; Stump and Finkel, 2013) for alternative formulations of the notion of principal part where different sets of paradigm cells serve as predictor depending on the lexeme.

[6]If X is a binary random variable one of whose values has a probability of 0.9995, $H(X) > 0.0062$.

Agency (ANR) as part of the "Investissements d'Avenir" program (reference: ANR-10-LABX-0083).

References

[Ackerman and Malouf2013] Farrell Ackerman and Robert Malouf. 2013. Morphological organization: the low conditional entropy conjecture. *Language*, 89:429–464.

[Ackerman et al.2009] Farrell Ackerman, James P. Blevins, and Robert Malouf. 2009. Parts and wholes: implicative patterns in inflectional paradigms. In James P. Blevins and Juliette Blevins, editors, *Analogy in Grammar*, pages 54–82. Oxford University Press, Oxford.

[Albright2002] Adam C. Albright. 2002. *The Identification of Bases in Morphological Paradigms*. Ph.D. thesis, University of California, Los Angeles.

[Baroni et al.2009] Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. 2009. The wacky wide web: A collection of very large linguistically processed web-crawled corpora. In *Language Resources and Evaluation*, volume 43, pages 209–226.

[Blevinsforthcoming] James P. Blevins. forthcoming. *Word and Paradigm Morphology*. Oxford University Press, Oxford.

[Bonami and Boyé2002] Olivier Bonami and Gilles Boyé. 2002. Suppletion and stem dependency in inflectional morphology. In Franck Van Eynde, Lars Hellan, and Dorothee Beerman, editors, *The Proceedings of the HPSG '01 Conference*, pages 51–70. CSLI Publications, Stanford.

[Bonami and Boyé2007] Olivier Bonami and Gilles Boyé. 2007. Remarques sur les bases de la conjugaison. In Elisabeth Delais-Roussarie and Laurence Labrune, editors, *Des sons et des sens*, pages 77–90. Hermès, Paris.

[Bonami and Boyé2014] Olivier Bonami and Gilles Boyé. 2014. De formes en thèmes. In Florence Villoing, Sarah Leroy, and Sophie David, editors, *Foisonnements morphologiques. Etudes en hommage à Françoise Kerleroux*, pages 17–45. Presses Universitaires de Paris Ouest.

[Bonami and Luís2013] Olivier Bonami and Ana R. Luís. 2013. Causes and consequences of complexity in portuguese verbal paradigms. In *9th Mediterranean Morphology Meeting*, Dubrovnik, septembre.

[Bonami and Luís2014] Olivier Bonami and 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.

[Bonami et al.2014] Olivier Bonami, Gauthier Caron, and Clément Plancq. 2014. Construction d'un lexique flexionnel phonétisé libre du français. In Franck Neveu, Peter Blumenthal, Linda Hriba, Annette Gerstenberg, Judith Meinschaefer, and Sophie Prévost, editors, *Actes du quatrième Congrès Mondial de Linguistique Française*, pages 2583–2596.

[Finkel and Stump2007] Raphael Finkel and Gregory T. Stump. 2007. Principal parts and morphological typology. *Morphology*, 17:39–75.

[Hockett1967] Charles F. Hockett. 1967. The Yawelmani basic verb. *Language*, 43:208–222.

[Kilani-Schoch and Dressler2005] Marianne Kilani-Schoch and Wolfgang Dressler. 2005. *Morphologie naturelle et flexion du verbe français*. Gunter Narr Verlag, Tübingen.

[Mateus and d'Andrade2000] Maria Helena Mateus and Ernesto d'Andrade. 2000. *ThePhonology of Portuguese*. Oxford University Press, Oxford.

[Matthews1972] P. H. Matthews. 1972. *Inflectional Morphology. A Theoretical Study Based on Aspects of Latin Verb Conjugation*. Cambridge University Press, Cambridge.

[New et al.2007] Boris New, Marc Brysbaert, Jean Veronis, and Christophe Pallier. 2007. The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics*, 28:661–677.

[Sagot2010] Benoît Sagot. 2010. The *Lefff*, a freely available and large-coverage morphological and syntactic lexicon for French. In *Proceedings of LREC 2010*.

[Simsforthcoming] Andrea Sims. forthcoming. *Inflectional defectiveness*. Cambridge University Press, Cambridge.

[Stump and Finkel2013] Gregory T. Stump and Raphael Finkel. 2013. *Morphological Typology: From Word to Paradigm*. Cambridge University Press, Cambridge.

[Thymé et al.1994] Ann Thymé, Farrell Ackerman, and Jeff Elman. 1994. Finnish nominal inflection: Paradigmatic patterns and token analogy. In Susan D. Lima, Roberta Corrigan, and Gregory K. Iverson, editors, *The Reality of Linguistic Rules*. John Benjamins.

[Veiga et al.2012] Arlindo Oliveira da Veiga, Sara Candeias, and Fernando Perdigão. 2012. Generating a pronunciation dictionary for european portuguese using a joint-sequence model with embedded stress assignment. *Journal of the Brazilian Computer Society*, 88.

[Wurzel1984] Wolfgang Ulrich Wurzel. 1984. *Flexionsmorphologie und Natürlichkeit. Ein Beitrag zur morphologischen Theoriebildung*. Akademie-Verlag, Berlin. Translated as (Wurzel, 1989).

[Wurzel1989] Wolfgang Ulrich Wurzel. 1989. *Inflectional Morphology and Naturalness*. Kluwer, Dordrecht.

Of crowds and corpora: A marriage of measures

Emmanuel Keuleers, Paweł Mandera, Michaël Stevens, & Marc Brysbaert
Department of Experimental Psychology
Ghent University
Henri Dunantlaan 2, 9000 Gent, Belgium
{emmanuel.keuleers, pawel.mandera,
michael.stevens, marc.brysbaert}@ugent.be

Abstract

We discuss the relationship between a word's corpus frequency and its prevalence –the proportion of people who know the word– and show that they are complementary measures. We show that adding word prevalence as a predictor of lexical decision reaction time in the Dutch lexicon project increases explained variance by more than 10%. In addition, we show that, for the same dataset, word prevalence is the best independent predictor of word processing time.

1 Introduction

Word frequency is one of the most important measures in the cognitive study of word processing, both theoretically and methodologically. Its contribution in explaining behavioural measures such as reaction time is so large that researchers take great care in collecting large and reliable corpora and in applying the best possible word frequency estimates in their research.

1.1 Where the corpus is weak the crowd is strong

A drawback of frequency counts is that, regardless of corpus size, lower counts are unreliable. As an example, consider asking a random sample of 100 people whether they know each of the word types that occur just once in a large corpus. Although frequency for all these types is equal, the number of judges knowing each word will vary from zero to one hundred and, as the judges are language users, words known to many of them may be considered to occur more often in language than words which are known by fewer of them. Following this reasoning, the estimate of *the number of language users who know a word*, or *word prevalence* may give a better indication of occurrence than corpus frequency counts.

1.2 Where the corpus is strong the crowd is weak

On the other hand, consider presenting the same random sample of people with words from the language's core vocabulary. Since these words will be known to all of the judges, *prevalence* will be singularly high and uninformative. In this case corpus counts should be a much better estimate of occurrence.

2 Testing the prevalence measure

To test the complementarity of prevalence and frequency as measures of occurrence, we used prevalence norms for Dutch collected through a lexical decision task presented as an online vocabulary test (Keuleers, Stevens, Mandera, & Brysbaert, in press). Each participants saw 100 stimuli (about 70 words and 30 nonwords) selected randomly from a list of 54,319 words and 21,734 nonwords. In the current analysis, we used the data of 190,771 participants who indicated that they were living in Belgium, giving us about 250 observations per word. The score for a word obtained by fitting a Rasch model –a mathematical model simultaneously ranking participants by ability and test-items by difficulty– to the data was considered an operationalization of its prevalence.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

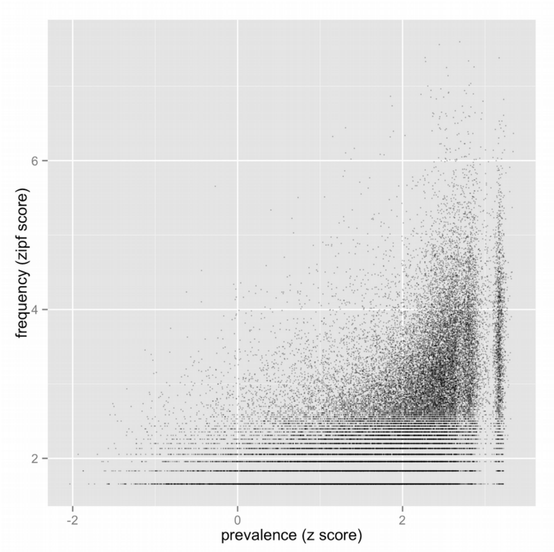


Figure 1: The relationship between frequency and prevalence. Word frequency is displayed as Zipf-score (log frequency per billion words; Van Heuven et al., 2014).

Figure 1 shows the complementary relation between the SUBTLEX-NL word frequencies (based on 42 million word corpus of film and television subtitles; see Keuleers, Brysbaert, & New, 2012) and the prevalence measure obtained from the online vocabulary test. Higher z-scores indicate more prevalent words. The dark lines at the bottom half of the plot indicate words with singularly low frequencies over a large range of prevalence. The elongated cluster at the right side of the plot shows words with nearly full prevalence over large frequency ranges.

In addition, we investigated the relationship between prevalence and other typical measures of word frequency. Table 1 gives an overview of these correlations.

	Frequency	Prevalence	OLD 20	Length
Frequency	1.00	0.35	-0.34	-0.37
Prevalence	0.35	1.00	0.00	0.07
OLD20	-0.34	0.00	1.00	0.74
Length	-0.37	0.07	0.74	1.00
Contextual Diversity	0.98	0.36	-0.34	-0.35

Table 1: Correlations between main predictors of Lexical Decision RT in the Dutch Lexicon ProjectTable 1 shows that the correlation between prevalence and frequency was relatively low (.34), giving further evidence that prevalence is distinct from word frequency and contextual diversity –a word's document count– which correlates very highly with word frequency.

Finally, we used the data from the 7,885 items in the Dutch Lexicon Project (Keuleers et al., 2010) for which both frequency and prevalence were available to examine the contributions of Dutch corpus word frequency (SUBTLEX-NL, Keuleers et al., 2010) and word prevalence on average reaction times.

In single variable analyses, log word frequency explained about 36.13% of the variance in reaction times and prevalence explained about 33.03% of the variance in reaction times.

This was also made clear when both measures were considered in the same analysis, where both measures jointly explained 51.37 % of the variance in reaction times. The unique contributions to explained variance (eta-squared) were 27.39% for frequency and 23.87% for prevalence. In further analyses, we found that including the quadratic trend of word frequency and contextual diversity did not substantially alter this pattern of results.

3 Conclusion

The results show that, next to word frequency, prevalence is by far the most important independent contributor to visual word recognition times, suggesting that prevalence should be included in any analysis where word corpus frequency is considered to be relevant. However, several questions remain open. First, what is the influence of corpus size on the relation between corpus word frequency and prevalence and on the contribution of prevalence to lexical processing? Second, how well does prevalence perform on others tasks and in other languages? Finally, does the effect of prevalence on word processing truly lie in a better measurement of word occurrence or does it partly reflect an independent property associated with the learnability of a word?

Acknowledgments

The text of this abstract is an early summary of findings from a larger study reported in the Quarterly Journal of Experimental Psychology as *Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment.* (Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M., in press).

References

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, 39(3), 445–459.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643–650. doi:10.3758/BRM.42.3.643

Keuleers, E., Diependaele, K., & Brysbaert, M. (2010). Practice Effects in Large-Scale Visual Word Recognition Studies: A Lexical Decision Study on 14,000 Dutch Mono- and Disyllabic Words and Nonwords. *Frontiers in Psychology, 1*. doi:10.3389/fpsyg.2010.00174

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2011). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods, 44*(1), 287–304. doi:10.3758/s13428-011-0118-4

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (in press). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2015.1022560

Perception of gesturally distinct consonants in Persian

Reza Falahati
Laboratorio di Linguistica
Scuola Normale Superiore
Piazza dei Cavalieri 7
56126 Pisa, Italy
reza.falahati@sns.it

Chiara Bertini
Laboratorio di Linguistica
Scuola Normale Superiore
Piazza dei Cavalieri 7
56126 Pisa, Italy
chiara.bertini@sns.it

Abstract

This study explores the sensitivity of the individuals to the residual gestures remaining after the simplification of consonant clusters. Three sets of target stimuli having full, reduced, and zero alveolar gestures along with the control stimuli were used in a perceptual identification task. The results of the experiment showed that subjects reliably distinguished the three target sets with varying residual gestures from the control. The results also showed that the degree of residual gestures affects the rate of [t] perception by the subjects; however, this was not statistically significant. The results are discussed in the context of different theories of speech perception.

1 Introduction

This study investigates the perception of three categories of consonant clusters that are perceptually similar but gesturally distinct. In Persian, word-final coronal stops are optionally deleted, when they are preceded by obstruents or the homorganic nasal /n/. For example, the final clusters in the words /ræft/ "went", /duχt/ "sew" and /qæsd/ "intension" are optionally simplified[1] in fast/casual speech, resulting in: [ræf], [duχ], and [qæs], respectively. The articulatory study conducted on this process in Persian by Falahati (2013) has shown that the gestures of the deleted segments are often still present. More specifically, the findings showed that of the clusters that sounded simplified, some had no alveolar gesture, some had gestural overlap that masked at least some of the acoustic information for [t], and some had reduced alveolar gestures. The current study tests listeners' sensitivity to these three types of /t/ realizations.

[1] The term "simplification" is used here for the acoustic and perceptual consequence of apparent coronal consonant deletion, regardless of whether there is a residual articulatory gesture.

2 Background

Choosing the basic units or building blocks by which the phenomena in a discipline could be explained is fundamentally important. Due to the "complex" nature of language, there is no consensus among linguists as to the nature of this basic unit in the field. The controversy over choosing the building blocks extends to the domain of speech perception where different models have postulated various basic units of processing and storage.

In general, there are two major theoretical approaches to speech perception: gesturalist theories versus auditory and exemplar theories. The two main gestural theories of speech perception are Motor Theory and Direct Realism (MT and DR, henceforth). In motor theories, the intended phonetic gestures of the speaker are considered to be the objects of speech perception. These gestures are "represented in the brain as invariant motor commands that call for movements of the articulators through certain linguistically significant configurations" (Liberman and Mattingly 1985, p. 2). The main motivation for choosing such basic unit by MT, among other factors, is mainly because of patterns where different acoustic cues could give rise to the same phonetic percept or where variant phonetic percepts were found for the same synthetic speech across different contexts (Delattre et al., 1955, 1964; Liberman 1957; Liberman and Mattingly 1985). Despite of the fact that this theory has gone through significant changes from its inception, all the versions share the idea that the objects of speech perception are articulatory events rather than acoustic or auditory events.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

An intended gesture is produced by a number of muscles that act in concert sometimes ranging over more than one articulator. For instance, constriction needed for producing coronal stops involves the action of the tip/blade of the tongue and the jaw; however, such a constriction is considered one gesture. According to MT, the orchestration among gestures is quite systematic and listeners can use the systematically varying acoustic cues for coronal stops as information to detect the related consonant gestures.

MT assumes a biological link between perception and production. According to this perspective both speech perception and speech production share the same set of invariants and are governed by auditory principles. “The motivation for articulatory and coarticulatory maneuvers is to produce just those acoustic patterns that fit the language-independent characteristics of the auditory system” (Liberman and Mattingly, 1985, p. 6). The acoustic signal only serves as a source of information about the gestures. It is the gestures which define the phonetic category.

The other main gestural theory to speech perception is direct realism. Both DR and MT share the claim that listeners to speech perceive vocal tract gestures. However, in DR it is the phonological gestures of the vocal tract, rather than the intended gestures, which are the perceptual objects (Fowler 1981, 1984, 1996). According to DR, “the temporal overlap of vowels and consonants does not result in a physical merging or assimilation of gestures; instead, the vowel and consonant gestures are coproduced. That is, they remain, to a considerable extent, separate and independent events...” (Diehl et al., 2004, p. 153). If we could extend this to the gestures of two adjacent consonants, one should expect that the gestures related to them also remain separate and distinct from each other.

In contrast to gestural theories, the auditory theories assume that speech sounds are perceived via general cognitive and learning mechanisms. In this view, speech is not special and listeners do not perceive gestures. The auditory approach to perception mainly considers general auditory mechanisms responsible for perceptual performance. According to this view, the speech and nonspeech stimuli do not invoke a special or speech-specific module. Gestures have no mediatory role as to the perception of speech sounds in this approach. Listeners use multiple imperfect acoustic cues in order to categorize the complex stimuli with structured variance (Diehl et al., 2004). According to this approach, the phonological representations are assumed to be speaker independent and they are associated with each word in the listener’s mental lexicon. The proponents of this approach take, for example, categorical perception of non-speech sounds or categorical-like perception by non-human animals as evidence for their argument. They also consider some of the cross-linguistic sound patterns and the “maximal auditory dispersion” in vowel systems as further support for their claim (Ohala 1990, 1995).

Exemplar theories form another approach to speech perception where words and frequently-used grammatical constructions are represented in memory as large sets of exemplars containing fine phonetic information. Listeners are sensitive to phonetic details existing in the speech signal. In such a speech perception model, a mechanism is needed for gradiently changing the lexical representations over time. In order to do so, the perceptual system must be capable of making fine phonetic distinctions (Johnson 1997).

These different approaches to speech perception have been tested in different studies. Beddor et al., (2013), for example, used eye-tracking to assess listeners' use of coarticulatory vowel nasalization as that information unfolded in real time. In the experiment, subjects heard the nasalized vowels with two different time latencies. The prediction was that subjects will fixate on the related image sooner when they hear the nasalized vowel earlier. The results showed that listeners use relevant acoustic cues, which was argued to allow listeners to track the gestural information. Nalon (1992) in an identification task tested whether participants could identify different degrees of velar assimilation. He used four different articulation types called full alveolar, residual alveolar, zero alveolar (i.e., full assimilation to the following velar), and nonalveolar (i.e., velar in underlying representation). The results of his study showed that the participants perceived full alveolar tokens with 100% accuracy with /d/ responses while less than half the tokens with residual alveolar were identified with /d/ responses. In another study, Pisoni showed that the nonspeech analogs of VOT stimuli are perceived categorically. Similar studies like this were taken as evidence against MT which claimed categorical perception as a specific feature of the speech mode of perception.

In this study, I will use three sets of simplified consonant clusters which are auditorily similar but gesturally different. The consonant clusters (i.e., C_1C_2#) happen in the coda of the words followed by another word which also starts with a consonant, therefore giving us three consonants in a row in an intervocalic environment (i.e., $V_1C_1C_2\#C_3V_2$). The prediction is that if subjects are sensitive, they should have different judgment for the stimuli. The stimuli set with no coronal gesture is expected to show the same pattern as the control (with zero coronal gesture in the underlying representation). The stimuli with overlapped gestures and reduced gestures are predicted to show a pattern different both from control and the stimuli with zero residual gestures. The following section introduces the methodology of the study.

3 Methodology

3.1 Participants

Thirty-two Persian-speaking students from the Università di Pisa and Sant'Anna, seventeen females fifteen males, aged 18-38 participated in this study. The results of eight of them are not considered for analysis because they reported to be bilinguals and mainly used a language rather than Persian at home or with their close friends. This resulted in twenty-four, twelve females twelve males. None of them reported any hearing problem.

3.2 Stimuli

Three sets of target words varying in only the degree/amount of alveolar residual gestures and one control stimuli set were used in the experiment. The three target categories are mainly the same except for the degree of alveolar residual gestures. Target Full_G category has full coronal gesture but has overlap hence marked with two superscript [^{tt}]. Target Partial_G category has partial residual gesture marked via superscript [^{t}] whereas Target Zero_G has no gestural leftover. The stimuli in the control are used as the baseline since they don't have any underlying coronal stop in the coda position of the first word. Some examples of the target and control words are given below:

Target Full_G: [æχ^{tt} kɑ], [æf^{tt} bæ], [uf^{tt} bɑ]

Target Partial_G: [æχ^{t} kɑ], [æf^{t} bæ], [uf^{t} bɑ]

Target Zero_G: [æχ kɑ], [æf bæ], [uf bɑ]

Control: [æχ ke], [æf bæ], [uf bɑ]

The four sets of target and control nonwords presented above are the excised tokens taken from the full words presented below:

Target: /sæχt kɑr/ "hard-working", /næft bærɑje/ "oil for", /kuft bɑʃeh/ "be cheap"

Control: /næχ ke/ "thread that", /sæf bærɑje/, "cue for" / mæruf bɑʃeh / "be famous"

3.3 Procedure

All the participants listened to forty stimuli (10 stimuli in each category) with eight repetitions. (total of 320 tokens) in a sound booth located at the linguistics laboratory in Scuola Normale Superiore. The software Presentation was used to present the stimuli to the listeners as an identification task. The participants were asked to listen very carefully and decide as quickly as possible whether it is likely that there has been a [t] at the end of the first part of each stimuli. For each stimulus, the participants were asked to press either the green or the blue button on a Cedrus response pad. On the screen of a computer, listeners could also see "T" or "NO T" corresponding to the response buttons. The stimuli were shuffled and presented in blocks in a way that participants could either begin by hearing all the tokens with [f] or [χ]. They also had the choice of taking a break after listening to every 80 tokens. All the participants received a short training before the start of the experiment. The following section contains the results of the study.

4 Results

The main goal of this study is to test listeners' sensitivity to different degrees of residual gestures remaining after the simplification of consonant clusters. The response type and reaction time are the dependent variables in this study; however, only the results related to response type are presented here. Figure 1 below shows the perception rate of [t] by all subjects across the four conditions. According to this, the subjects show the highest rate of [t] perception in tokens with full alveolar gesture (i.e., 59.69%) and the lowest for the ones in the control (i.e., 36.09%). The condition with partial alveolar gestures shows the rate of 56.20% which is very close to the full condition. The stimuli in zero alveolar condition show an intermediate level between the control and the other two target conditions with the rate of 49.84%. This shows almost a similar pattern between the two target conditions with full and partial gestures, an intermediate situation for the target condition with zero gesture, and a pattern for the control which is different from the three target conditions.

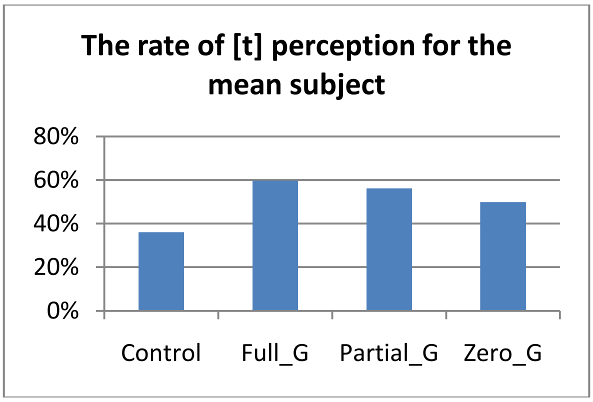


Figure 1: The Rate of [t] Perception by all Subjects

In order to examine the relation between the two categorical variables in the study, namely the response type and stimuli condition, a Pearson chi-square test was run. The null hypothesis is that there is no relation in the [t] perception and the four conditions in the study. The results of the test with [t] perception as the dependent variable found significant main effect of conditions χ^2 (3, N = 960) = 46.2, p < 0.001. This shows that there is a significant relation between the stimuli conditions and response type. In order to determine whether the difference in the perception of [t] across four categories is really significant or it is due to chance variation, a column proportions test was performed. This test uses z-test to make the comparisons. The result showed that the perception of [t] in the control was significantly different from the all target categories. The next section presents the discussion and concluding remarks of the study.

5 Discussion and Conclusion

This research investigated listeners' sensitivity to three types of /t/ realizations as target and compared the results with the control. The target categories included simplified consonant clusters with full, partial, and zero alveolar gestures. The stimuli used as the baseline in the control had no alveolar gesture in the underlying form. The general results of the study showed that subjects reliably distinguished the three target sets with varying residual gestures from the control. This could be due to more similarity in tongue configuration in realizing these varying degrees of coronal stop articulation compared to the control condition where there is no alveolar gesture in the underlying form. Any articulatory modification is expected to trigger acoustic changes. The acoustic results of the stimuli used in this study by Falahati (2013) showed no significant difference between the simplified tokens (i.e., the three target sets with varying degrees of residual gestures labeled all together as simplified) and control tokens. The acoustic parameters used in the analysis were V_1 duration, consonant clusters duration, and formant transitions. Despite of the fact that the results did not show any significant difference between simplified and control conditions, the duration of V_1 and consonant clusters in the simplified condition was always higher than the control condition. It could be the case that these acoustic cues, although not very strong, are enough for human's auditory system to trigger the presence of a segment.

The results of the current study also showed that participants perceived almost 36% of the tokens with no underlying coronal stop as having [t]. This is very similar to the results of the study reported by Nalon (1992) where 20% of the control nonalveolar tokens were perceived as having [d]. In his study, however, the control tokens showed similar pattern to that of the target with zero alveolar (i.e., full assimilation). He attributes this to both subjects' natural language experience as well as the inherent ambiguity in the stimuli. He states that subjects are "willing to "undo" its effects" and therefore, in the case of the current study, report coronal stops even where there is no evidence for them.

The results of our study also showed that participants perceived more [t] in the tokens with full and partial alveolar gestures compared to the ones with zero alveolar gestures. The difference between the three categories, however, did not reach the significance level. Such result could shed more light on the theories of speech perception discussed earlier in this paper. In order to discuss this issue, first we need to further explore the nature of the three categories in the target stimuli. From the three groups in the target stimuli, one group categorically had no alveolar gesture while the other two had different degrees of the gesture either as a result of overlap or reduction. We argue that the gradient gestural reduction and overlap are due to low-level phonetic and mechanical reasons while the categorical deletion, which results in tokens with zero gestures, is caused by the cognitive system. In the former groups, speakers neither intend to reduce nor plan to overlap gestures while the latter process is intended by the speaker.

According to MT and DR, listeners' target in speech perception is the intended or phonological gestures. Therefore, the overlapped and reduced stimuli should show different perceptual pattern compared to the stimuli with no residual gesture. The results in this study did not show a striking difference between these three target sets. The existence of acoustic cues pertaining to the presence of gestures is a prerequisite to their perception by the listener. If distinguishing acoustic details could be found between these three categories, then this would not support the gesturalist approach to speech perception. However, with the current results, such a claim cannot be made. Further acoustic analysis between these three target sets is needed to examine this idea further.

The findings in our experiment could be best explained by referring to exemplar models of speech perception. In such models, the lexical representations of words change in a gradient way over time. This is due to the nature of some phonological processes in languages which are not categorical. According to this view, the perceptual mechanism is capable to make fine phonetic distinctions. However, it is the mapping between the gradient stimuli and the auditory system which fails and does not result in nonvariant forms.

The lack of such a one-to-one mapping will bring variation across subjects in the speech community. The degree of such variation is determined by the *amount* of individual's exposure to the *specific* variants. A closer look at the results for individual subjects showed that all twenty-four participants in the study could fall into three or four dominant patterns based on their perception of [t]. The variation across individuals regarding speech perception could be a good source of information for the specialists in the field. Moreover, the degree to which an individual's speech production could map to his/her perception is an interesting topic which remains to be explored.

Acknowledgments

We are very grateful to Patrice Beddor for her comments and suggestions on this study.

Reference

Patrice S. Beddor, Kevin B. McGowan, Julie Boland, Andries W. Coetzee, and Anna Brasher. 2013. The perceptual time course of coarticulation. *Journal of the Acoustical Society of America*, 133, 2350-2366.

Pierre Delattre, Alvin M. Liberman, and Franklin S. Cooper. 1955. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 27, 769-773.

Pierre Delattre, Alvin M. Liberman, and Franklin S. Cooper. 1964. Formant transition and loci as acoustic correlates of place of articulation in American fricatives. *Stud. Linguist*. 18, 104-121.

Randy L. Diehl, Andrew J. Lotto, and Lorri L. Holt . 2004. Speech perception. *Annual Review of psychology.* 55, 149-179.

Alvin M. Liberman and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. Cognition, 21: 1-36.

Reza Falahati. 2013. *Gradient and Categorical Consonant Cluster Simplification in Persian: An Ultrasound and Acoustic Study*, Ph. D Dissertation, University of Ottawa, Ottawa.

Carol C. Fowler. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research,* 46, 127-139.

Carol C. Fowler. 1984. Segmentation of coarticulated speech in perception. *Perception & Psychophysics,* 36, 359-368.

Carol C. Fowler. 1996. Listeners do hear sounds, not tongues. *Journal of the Acoustical Society of America,* 99, 1730-1741.

Keith Johnson. 1997. Speech perception without speaker normalization: an exemplar model. In K. Johnson and J. W. Mullennix (eds.), *Talker variability in speech processing*, 145-165. San Diego: Academic Press.

Alvin M. Liberman and Ignatius G. Mattingly. 1985. The motor theory of speech perception revised. *Cognition*, 21: 1-36.

Francis Nalon, 1992. The descriptive role of segments: evidence from assimilation. In G. J. Docherty and R. Ladd (eds.), *Papers in Laboratory Phonology IV*, 261-289. Cambridge: Cambridge University Press.

John Ohala. 1990. Respiratory activity in speech. In W. J. Hardcastle & A. Marchal (eds.), *Speech Production and Speech Modeling*, 23-53. Netherlands: Kluwer Academic Publishers.

John Ohala. 1995. The perceptual basis of some sound patterns. In B. Connell and A. Arvaniti (eds.), *Phonology and phonetic evidence, Papers in Laboratory Phonology IV*, 87-92. Cambridge: Cambridge University Press.

Words matter more than morphemes: Evidence from masked priming with bound-stem stimuli

Hélène Giraudo
Laboratoire CLLE (Equipe ERSS)
CNRS & Université Toulouse Jean Jaurès
giraudo@univ-tlse2.fr

Madeleine Voga
University Paul-Valéry Montpellier III
madeleine.voga@univ-montp3.fr

Abstract

Five masked priming studies were carried in order to shed light on the processing of bound-stem words (e.g., *terr-* in *terrible*). Both orthographic (e.g., *termite*) and unrelated (e.g., *montagne* 'montain') conditions stand as baselines for controlling morphological effects. The results of the experiments using unrelated word controls suggest that in the particular case of bound-stem words, only genuinely derived word primes (*terrible*) produce positive effects differing from formal overlap effects. Morphological effects are interpreted as resulting from both "morceme" and "base-lexeme" activations.

1 Introduction

As is broadly admitted, morphologically related words prime each other in various languages (Arabic: Boudelaa & Marslen-Wilson, 2001; English: Rastle, Davis, Marslen-Wilson & Tyler, 2000; French: Giraudo & Grainger, 2000; German and Dutch: Drews and Zwitserlood, 1995; Hebrew: Frost, Deutsch & Forster, 1997) thus suggesting the existence of a morphological level of processing. This kind of study has used different types of materials, words or pseudowords, as well as multiple settings: for the masked priming technique (Forster & Forster, 2003), widely used to shed light on morphological processes as well as in this study, the distinction can be made between designs using only unrelated controls and those using both unrelated and orthographic/phonological controls, as suggested by Giraudo & Grainger (2001) or Pastizzo & Feldman (2002).

Even though the existence of a morphological level of processing is unanimously acknowledged, the exact nature, locus and the role of these representations within the mental lexicon is a matter of ongoing research. Two hypotheses can be drawn: according to the first, morphemic units correspond to concrete pieces of words (i.e., stems and affixes). Complex words are therefore processed through a decomposition mechanism that strips off the affix in order to isolate the stem. The morphemic nature of the remaining letters is then verified by the system and access to word representations (i.e., word forms coded in the orthographic lexicon) operates via the pre-activation of their constituent morphemes, i.e., morphemic representations stand as access units. This mechanism is exemplified by Taft's model (1994), the basic principles of which are followed by many psycholinguistic studies (e.g., Crepaldi, Rastle & Davis, 2010). Morphemic units are situated between the level of letters/syllables and the word level; consequently, they can only be matched to concrete letter clusters (i.e., bound-stems, free-stems and affixes) that constitute words. This decompositional mechanism is also insensitive to any semantic characteristics of words (i.e., transparent vs. opaque morphological formation) or to their lexical environment (in terms of orthographic neighborhood or family size). One of the strong predictions of the decompositional approach is that morphological priming effects should vary following the ease with which constituent morphemes can be identified/extracted.

According to the second hypothesis, morphology is coded at the interface of word and semantic representations and corresponds rather to lexemes (Aronoff, 1994). Lexeme units are coded at the interface of the word and the semantic level, organizing the lexicon in terms of morphological families. The recognition of any complex word triggers first the activation of all word forms that can match with it; a competition is then engaged between the pre-activated forms (forms matching the input, i.e., those who are morphologically

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

related but also those who are only orthographically related) until the right lexical representation reaches its recognition threshold (determined by its surface frequency). During this competition phase, morphologically related words send positive activation to their respective base lexeme, feeding back activation to them. Morphological priming effects result from this mechanism of co-activation. Following this supralexical approach (Giraudo & Grainger, 2000; 2001), complex words are not "decomposed", but are able to trigger the activation of their constituent morphemes. In this kind of architecture, lexeme units are supposed to be abstract enough to tolerate variation induced by derivation and inflection (i.e., allomorphy, suppletion, phonological/ morphological truncation, haplology, verb-noun conversion). In other words, a morphological unit does not necessarily need to surface in the real world in order to be coded in long-term memory. This organisation, compatible with recent neuroimaging data (Lévy, Hagoort, Démonet, 2014), also implies that all morphemes of a given language are not necessarily represented within the mental lexicon: units such as neologisms, hapaxes and nonce words are not necessarily directly connected with existing morphological units; bound-stem words could be such a case.

2 The study

The present paper focuses on the processing of bound-stem words by opposition to free-stem words. For ex., on one hand, the word *viral* composed of the bound-stem *vir-,* also present in *virus*, *virulent*, *virulence*, *virology* and *virologist* and, on the other hand the word *singer* composed with the free-stem *sing* that forms *singing*, *song*, etc. Both are defined as being morphologically complex but while it is evident for the standard speaker/reader that the complex word *singer* derives from the root *sing*, it is less evident to say from which root the complex word *viral* derives. The morpheme *vir-*, which does not have any clear meaning in English, can be considered as a bound-stem whereas *sing-* in *singer* is a free-stem. From a processing point of view, the *vir-viral* example can be viewed as a case where the lexical unit is not directly connected to the morphological unit, by virtue of its twofold handicap: the first aspect is semantic interpretability, i.e., derivations composed with a bound-stem could be less interpretable than those with a free-stem. Psycholinguists tested this difference and found that processing for free and bound-stems may differ but both produce significant priming effects (Forster & Azuma, 2000; Järvikivi & Niemi, 2002; Marslen-Wilson, Tyler, Waksler & Older, 1994; Pastizzo & Feldman, 2004). Of great importance to our study, Pastizzo and Feldman (2004) observed that the magnitude of facilitation varied following the baseline used in the experiments: equivalent magnitudes of priming for free and bound-stems were obtained relative to an unrelated baseline; with an orthographic control however, free-stems produced systematically greater priming than bound-stems. The interpretation of this line of research suggests that morphological priming effects are not directly constrained by semantic similarity between prime and target. The second handicap, in terms of surface analysis, consists in the difficulty in segmenting the word forms into morphemes. At this point, the two different models presented above give rise to different predictions: according to the morpheme-based approach all complex forms (free-stem as well as bound-stem words) are first analyzed in morpheme fragments and then access word representation, in other words, the lexicality of the base doesn't matter. This approach predicts morphological priming between derivations (e.g., *virus-viral*) as well as between the base and its derivation (e.g., *vir-viral*).

According to the supralexical approach, the members of a morphological family are linked together by virtue of their common base at the lexeme level; however, the base of bound-stem words is not represented at the word level. In this case, priming effects between related derived words (e.g., *virus-viral*) are expected but no effect should be observed using their bound-stems as primes, the access to the base lexeme being conditioned by the prior activation of a word form at the word level.

Taft and Kougious (2004) investigated this issue in English through a masked priming experiment. They compared both semantically and orthographically related words (e.g., *virus-viral*) to merely orthographically related words (e.g., *future-futile*) and, unsurprisingly, found facilitation in the former case but not in the latter. Nevertheless, the design of this study is not very informative with respect to the decomposition issue, given that the critical condition examining the effect of the bound-stem on its derivations has not been considered.

Our study aims to fill this gap through five visual masked priming experiments with native French speakers. In this kind of protocol, subjects are unaware of the presence of the prime, which allows minimizing strategy use and examining automatic processing during the early stages of word identification: all five experiments use a within-priming (Latin square) design, in which we directly compare the effects of different primes on the same target. A 57ms prime duration) was used and the task was lexical decision.

Exp. 1 examined morphological effects induced by words sharing the same bound-stem, e.g., *terrible – terreur 'terrible-terror'* relative to an orthographic control baseline, e.g., *termite – terreur*' *'termite-terror'* (where *'termite'* is a monomorphemic word), as well as an unrelated baseline (*montagne – terreur* 'mountain-terror'). Results show that only truly derived word primes produce facilitation, relative to unrelated (36ms of facilitation) as well as orthographic controls (35ms). However, this first result does not inform us about how derived words constructed with a bound-stem are processed: are they analyzed in terms of stem + affix or are they globally processed? Exp. 2 examined the extent to which the facilitation we take as morphological could be due to formal overlap: this is done by using non-existent orthographic controls, sharing all but one letter with the 'true' bound-stem, e.g., for the target *terreur,* the first possible prime is the true bound-stem *terr-* presented in isolation (e.g., *terr* - *terreur*); the second priming condition is the non-existing bound-stem condition (orthographic control) *tarr-* (e.g., *tarr* – *terreur*); the third condition is an unrelated baseline (e.g., *montag* – *terreur*). Although only true bound-stems induced significant facilitation relative to the unrelated baseline (28ms), the non-existing stem condition (e.g., *tarr-*) exhibited reaction times (RTs) that did not differ significantly from those of the true bound-stem condition. This result highlights the fact, already pointed out by Forster (1999), that there is an influence of formal factors in this kind of protocol, as well as the need to include orthographic controls in the design. Experiment 3 directly compared the effects of complex word primes to those of bound-stem primes: the targets were the same as in Exp. 1 and the three levels of the prime type factor were the following: a morphologically related suffixed word sharing the same bound-stem, e.g., *terrible* – *terreur 'terrible-terror'*; its bound-stem, e.g., *terr* - *terreur*; an unrelated control, e.g., *montagne* – *terreur*. Results showed that only complex word primes (e.g., *terrible*) produced significant priming effects (33ms), though these conditions did not significantly differ from the bound-stem condition (e.g., *terr-*) whose effect (18ms) did not manage to reach significance.

Exp. 4 was designed to see if the advantage for the complex word sharing the same bound-stem found in exp. 3 holds up to the comparison with non-word primes constructed with the same bound-stem and an existing suffix. The three priming conditions were the following: the morphologically related word sharing its bound-stem with the target, e.g., *terrible – terreur*; a non-word made of the same bound-stem and a suffix different to that of the target, e.g., *terrage* – *terreur* (where *-age* corresponds to an existing morpheme); an unrelated control, e.g., *montagne* – *terreur*. The statistical analysis of the results revealed that only related word primes (e.g., *terrible*) produced significant morphological priming (40ms) relative to the unrelated controls. Even if the non-word prime condition (e.g., *terrage*) led to quicker reaction times compared to the unrelated baseline (688 vs 703ms), it didn't differ significantly from it. More importantly, the 25ms difference between the word prime condition and the non-word one is statistically significant. This suggests that it takes a real word to induce morphological priming, independently and above orthographic low-level perceptual influences, to which the masked priming technique is known to be sensitive. Our results show that the presence of an existing bound-stem in a non-word does not suffice to produce morphological priming, a finding which contradicts those published by Longtin and Meunier (2005) as we shall see in the discussion. Experiment 5 examined the extent to which the morphological facilitation found in exp. 4 could be due to formal factors: in order to test this, we replaced the morphologically related word primes by non-words constructed with a bound-stem and a final letter sequence that does not correspond to any suffix in French. The following three prime conditions defined the three levels of the prime type factor: a complex non-word formed by a bound-stem and a suffix, e.g., *terrage* – *terreur* (where *terr-* and *–age* correspond to existing morphemes); a simplex non-word formed by a bound-stem and a non-existing ending, e.g., *terryme* – *terreur,* in which *–yme* is not a suffix; finally, an unrelated non-word, e.g., *moitagne* – *terreur*. The statistical analysis of the results revealed that both complex and simplex non-word primes produced shorter RTs than unrelated primes (31 and 27ms of effect respectively): both types of prime are able to facilitate target recognition and produce thus morphological-like facilitation. Nevertheless, the fact that the effects produced by complex primes (e.g., *terrage*) did not differ from those produced by simplex non-word primes (e.g., *terryme*) leads us to reject any interpretation based on pre-lexical morphological decomposition. We suggest interpreting this pattern of results on the basis of formal criteria: for real words it takes a real word to facilitate processing (exp. 4), but for non-words, given the absence of representation in the word-level, morphological-like priming does nothing but reflect low-level perceptual similarities, such as between the two non-words (both complex and simplex, *terrage/terryme*) and the target *terreur*. Besides the role attributed to formal factors, the point that should be stressed in the interpretation of exp. 4 and 5 is that while in exp. 4 the nonword made up from an existing bound-stem and an existing suffix (*terrage*) seems to interfere with processing of the target (*terreur*) by virtue of its morphological structure, in exp. 5 this interference disappears. The fact that, in the '*terrage'* condition (exp. 4) we observe RTs that are not significantly quicker than the unrelated condition, despite the existence of a formal overlap combined with morphological-like structure (*terr-age/terr-eur*), can only be due to some kind of interference, otherwise we should observe at least a small formal effect. This interference nevertheless disappears in exp. 5, since both types of non-words (with existing suffix, e.g., *terrage,* as well as well as non-existing suffix, i.e., simplex non-words such as *terryme*) lead to significant facilitation. We therefore obtain a different pattern of priming for words (exp. 4) and for non-words (exp. 5) which leads us towards an approach where lexicality of the prime does matter in the overall pattern of results. Even if the processing system can take advantage of orthographic similarities between prime and target (and will not prevent itself from doing so, as exp. 2 showed) this does not tell the whole story, and it certainly does not tell a morphological story: it is just another demonstration of a fact that researchers working with masked priming are familiar with, namely that this technique is sensitive to formal factors (Forster, Mohan & Hector, 2003). The experiments presented here provide evidence that we can use this valuable technique in order to shed light on truly morphological effects, as opposed to morphological-like effects.

Taken together, the results of the experiments using unrelated word controls (exp. 1, 3 and 4) suggest that in the particular case of bound-stem words, only genuinely derived word primes (*terrible*) produce positive effects differing from formal overlap effects. This is true with the exception of exp. 3, where the effect of genuinely derived word primes did not differ from bound-stem primes (*terr-*); note however that in this experiment, the bound-stem condition did not differ from the unrelated condition, while the derived word condition did. This is a demonstration of the fact that "nonwords would be always better form-primes than words, even when masked. The reason is simply because a related word prime will compete more vigorously with the target than a related nonword prime" (Forster, 1999: 8). These results are not in accordance with those found by Longtin and Meunier (2005) using roughly the same priming conditions. In their study, derived non-word primes (e.g., *garagité*) systematically produced significant priming effects on target recognition relative to unrelated word controls, while non-morphological non-word primes (e.g., *rapiduit*) yielded a 29 ms non-significant effect. Two factors can explain these contradictory results: a) the type of unrelated controls: contrary to Longtin and Meunier, we examined priming effects relative to unrelated non-word primes when the prime conditions included non-words and word primes when the prime conditions included words; b) the type of word targets: given that our study focuses on bound-stem words, our targets are mandatorily complex words, and not bare-bases, as in the Longtin & Meunier study. Bare-bases are by definition more frequent, and, subsequently, easier to activate because of their lower activation threshold (due to their residual activation; for a discussion on this point based on McClelland & Rumelhart 1981, see Voga & Giraudo, 2009; Giraudo & Voga, 2014).

3 Discussion: On the representation of bound-stem words

On the basis of the above results, we can conclude that recognition of complex words benefits from two springs of facilitation: a bottom-up excitation from a sublexical level and a top-down facilitation from a supralexical level. The idea of a double representation for morphology was recently expressed by Diependaele, Sandra & Grainger (2005), suggesting that the morphological level should be situated both above and below the word-form level. Subsequently, morphological representations would be either defined as morphologically constrained orthographic representations (depending on frequencies) or as morphologically constrained semantic representations (coded in terms of regularities in the mapping of word forms onto semantics). In the same line, Crepaldi et al. (2010) proposed an extension of Taft's (1994) sublexical model integrating a lemma level comprised between an orthographic lexicon and the semantic system. However, these two models consider the two morphological levels equivalent, given that they both contain units corresponding to concrete morphemes. One may nevertheless assume that different locations imply different contents: the hybrid model we propose (Giraudo & Voga, 2014) is based exactly on this assumption. Within this model, morphological complex words are coded according to two dimensions, their surface form and their internal structure. The first level captures the statistical regularities of morphemes translated in terms of perceptual saliency in the language. At this level, morphologically complex and pseudo-derived words as well as non-words whose surface structure can be divided into distinct morphemes, are equally processed. This level is not a morphological level but rather a sub-orthographic level containing "morcemes". The second level, i.e., the morphological level is paradigmatically oriented, it deals with the construction of words according to morphological rules (Booij, 2005; Corbin, 1987/1991); it contains "base-lexemes", units abstract enough to tolerate orthographic and phonological variations produced by derivation and inflection processes and connected to their related word forms on the basis of semantic transparency.

References

Aronoff, M. (1994). *Morphology by itself.* Cambridge: MIT Press.

Boudelaa, S., & Marslen-Wilson, W. D. 2001. Morphological units in the Arabic mental lexicon. *Cognition*, 81:65-92.

Booij, Geert. 2005. Compounding and derivation: evidence for Construction Morphology. In W.U. Dressler, D. Kastovsky and F. Rainer (eds) *Demarcation in Morphology*. Amsterdam / Philadelphia: John Benjamins. 109-132.

Corbin, D. (1987/1991). *Morphologie derivationnelle et structuration du lexique*, vol. 2. Tubingen/Villeneuve d ' Ascq: Max Niemeyer Verlag / Presses Universitaires de Lille.

Crepaldi, D., Rastle, K., Davis, C. 2010. Morphemes in their place: Evidence for position-specific identification of suffixes. *Memory & Cognition,* 38:312-321.

Drews, E. & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance,* 21:1098-1116.

Forster, K. I. (1999). The microgenesis of priming effects in lexical access. *Brain and Language*, 68:5–15.

Forster, K.I. and Azuma, T. 2000 Masked priming for prefixed words with bound-stems: Does submit prime permit? *Language and Cognitive Processes,* 15(4-5):539-561.

Forster, K.I., & Forster, J.C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavioral Research Methods: Instruments & Computers,* 35:116-124.

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

Giraudo, H. & Grainger, J. 2000. Prime word frequency in masked morphological and orthographic priming. *Language and Cognitive Processes,* 15:421-444.

Giraudo, H., & Grainger, J. 2001. Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review,* 8(1):127-131.

Giraudo, H. & Voga, M. (2014). Measuring Morphology: the tip of the iceberg ? A retrospective on 10 years of morphological processing. *Carnets de Grammaire*, 22:136-167.

Järvikivi, J., & Niemi, J. 2002. Form-based representation in the mental lexicon: Priming (with) bound stem allomorphs in Finnish. *Brain & Language,* 81:412-423.

Lévy, J., Hagoort, P., & Démonet, J.F. 2014. A neuronal gamma oscillatory signature during morphological unification in the left occipitotemporal junction. *Human Brain Mapping*, 35(12):5847-60.

Longtin, M.-C. & Meunier, F. 2005. Morphological decomposition in early visual word processing, *Journal of Memory and Language,* 53(1), 26-41.

McClelland, J. L., D.E. Rumelhart. 1981. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review,* 88: 375-407.

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review,* 101:3-33.

Pastizzo, M. J. and Feldman, L. B. 2004. Morphological processing: A comparison between free and bound-stem facilitation. *Brain & Language,* 90:31-39.

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. 2000. Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes,* 15:507-537.

Taft, M. 1994. Interactive activation as a framework for understanding morphological processing. *Language & Cognitive Processes,* 9:271-294.

Taft, M., & Kougious, P. 2004. The processing of morpheme-like units in monomorphemic words. *Brain & Language,* 90:9-16.

Voga, M. & H. Giraudo. 2009. Pseudo-family size influences processing of French inflections: evidence in favor of a supralexical account in F. Montermini, G. Boyé, J. Tseng (eds), *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux.* Somerville, MA: Cascadilla Proceedings Project, 148-155.

Phonotactic probabilities in Italian simplex and complex words: a fragment priming study

Giulia Bracco
Università di Salerno
Via Giovanni Paolo II 132
Fisciano (SA)
gcbracco@unisa.it

Basilio Calderone
CNRS & Université de Toulouse II
5 allées Antonio Machado
Toulouse
basilio.calderone@univ.tlse2.fr

Chiara Celata
Scuola Normale Superiore
P.zza dei Cavalieri 7
Pisa
celata@sns.it

1 Introduction

Phonotactics refers to the sequential organization of phonological units that are legal in a language (Crystal 1992). However, legal sound sequences do not all occur with the same probability in a language. *Phonotactic probability* is most often measured in terms of transitional probabilities (TPs) of biphones and has been shown to influence a large range of processes, including infants' discrimination of native language sounds, adults' ratings of the wordlikeness of nonwords (Vitevitch et al. 1997), speech segmentation (Pitt & McQueen 1998, Mattys & Jusczyk 2001), word acquisition (Storkel 2001) and recognition (Luce & Large 2001). Specifically, in the domain of word recognition, high TPs facilitate word and nonword identification in speeded same-different matching tasks, but slow down identification in lexical decision tasks due to the inhibitory effects of a large neighborhood (e.g. Vitevitch & Luce 1999, Luce & Large 2001). Most of the studies on the role of TPs in speech production and perception have been conducted on English.

In this paper we focus on the role of phonotactic probabilities in priming morphologically simplex and complex words in Italian. We investigate whether biphone TPs affect the recognition of word targets after exposure to fragment primes differing in the probability with which the fragment-final consonant predicts the consecutive segment in the target.

We opted for a non-factorial, regression design including lexical and sub-lexical frequency and distributional variables as predictors (see Baayen 2010). In this paper, we report on the results of the study on simplex words only; we however discuss the implications of the current findings for the processing of complex words.

2 Experiment

2.1 Materials and procedure

Forty-two native Italian speakers participated in a speeded lexical decision task in a fragment priming paradigm. Thirty bi- or tri-syllabic Italian nouns containing a biphonemic consonant cluster in internal position (e.g. *borsa*, 'bag') served as targets. Each target was primed by a sequence corresponding to an initial fragment of the target (e.g. *bor-borsa*). The fragment prime could consist of 3 o 4 phonemes and always ended with the first consonant of the cluster. The average length ratio between prime and target was 0.49. The clusters were different across words and each cluster could occur in only one target (although more than one fragment could end in a given consonant). 12 were heterosyllabic (e.g. *bor-sa* 'bag'), 12 tautosyllabic (e.g. *deg-rado* 'decay') and 6 ambisyllabic clusters (e.g. *dis-tanza* 'distance').

Another set of 30 Italian nouns matching for average length, frequency and prime/target length ratio, in which the fragment prime ended in a syllable onset consonant followed by a vowel (e.g. *tuc-tucano* 'toucan'). The same proportion of fragment-final consonants was maintained in the two sets of words.

Sixty pseudowords matching for average length and properties of the fragment were added. Pseudowords were obtained by changing one letter of existing words (belonging to the same frequency range of the experimental words), for

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

1/3 in their initial part, 1/3 in their central part and 1/3 in their final part. The 30 clusters used for pseudowords did not appear in the words' list.

In the lexical decision task, participants were asked to press a button corresponding to their dominant hand as soon as the orthographically presented target was judged as a word, and a different button for targets judged as nonwords. All the stimuli appeared in Courier New font, 18 point size in the center of the computer screen. In order to avoid allographic effects, primes were displayed in uppercase and targets in lowercase. The fixation was 200 ms, followed by a 50 ms pause. Primes appeared for 150 ms, followed by a 50 ms pause. The targets remained on the computer screen for a maximum of 1 sec. If the participants did not produce any answer within that time, the feedback *Fuori tempo* ('Out of time') appeared on the screen. Reaction times (RTs) and the number of errors (Nerr) constituted the dependent variables. The reaction times were measured from target onset to subject's response, and responses given after the deadline were scored as errors.

The Experiment was preceded by a practice session. When the participants reached the 70 % of valid responses the experiment started.

2.2 Experimental variables

Several statistical and distributional properties of word primes, targets and clusters were derived from the CoLFIS corpus (Bertinetto et al., 2005).

For each prime-target pair, we calculated (i) the token frequency of the target ('TargetFreq'), (ii) the N of words beginning with the prime fragment ('PrimeTypeFreq'), (iii) the cumulated frequency of the words in (ii) ('PrimeTokenFreq'), (iii) the length of the target (in N graphemes), (iv) the length of the prime (in N graphemes), (v) the prime/target length ratio.

For each cluster, we calculated (vi) the TP value, i.e. the probability with which the first consonant of the cluster predicts the occurrence of the following consonant, calculated over the corpus word tokens ('BigramTP'), (vii) the N of words containing the cluster ('BigramTypeFreq'), (viii) the cumulated frequency of the words in (vii) ('BigramTokenFreq'), (ix) the TP between the fragment prime and the second consonant of the cluster, e.g. P(*s*|*bor*) in *borsa* 'bag' ('SequenceTP'), (x) the N of words containing the sequence of the prime followed by the second C of the cluster ('SequenceTypeFreq'), (xi) the cumulated frequency of the words in (x) ('SequenceTokenFreq').

2.3 Analysis and results

Fixed and mixed models with subject and prime as random variables were used.

For the purposes of the present study, we tested two different models, both including frequency variables and phonotactic probability variables; they are shown in Table 1. The two models differed for the presence, in model II, of a measure of prime frequency, which was not included in model I, and for being focused either on sequence and bigram token frequencies (model I), or on sequence and bigram type frequencies. Both models were tested for CC items (e.g. *borsa*, 'bag') and CV items (e.g. *tuc-ano* ' toucan') separately.

	Model I	*Model II*
Fixed effects	TargetFreq LenghRatio SequenceTokenFreq BigramTokenFreq SequenceTP BigramTP	TargetFreq PrimeTokenFreq LengthRatio SequenceTypeFreq BigramTypeFreq SequenceTP BigramTP
Random effects	Subject Fragment prime	Subject Fragment prime

Table 1. Fixed and random effects for the CC and CV items.

The results of the fixed effects analyses for the relevant models are summarized in Table 2 (dependent variable: RTs) and Table 3 (dependent variable: Nerr).

According to model I, with RTs as the dependent variable, the sequence's TP (i.e., the TP between the fragment prime and the second consonant of cluster) turned out to be the most significant predictor, even outranking the contribution of frequency values (for the target, the sequence and the bigram), which all concurred to the intercept. A different picture emerged however for the CV items, for which no probability variables turned out to significantly predict the subjects' response times; on the contrary, the target frequency, with the secondary contribution of the frequency of the cluster, appeared to play a role for this subset of items.

According to model II, for CC items the role of the target frequency turned out to be very important, and the only additional effect was generated by the sequence's TP. Thus the two models were similar in emphasizing the role of the probability with which a given C follows the prime sequence. As for CV items, model II returned a picture very similar to the one that emerged in model I, with target frequency and bigram type frequency as the only significant predictors.

	Estimate	Std.	Error	p-value	Adjusted R^2
CC item model I					
Intercept	643.492	30.563	21.054	<0.001	0.433
TargetFreq	-12.008	5.062	-2.372	<0.05	
SequenceTokenFreq	-7.487	3.105	-2.412	<0.05	
BigramTokenFreq	-7.483	3.038	-2.463	<0.05	
SequenceTP	64.248	18.907	3.398	<0.01	
CV item model I					
	Estimate	Std.	Error	p-value	Adjusted R^2
Intercept	911.829	96.470	9.452	<0.001	0.47
TargetFreq	-15.849	3.860	-4.106	<0.001	
BigramTokenFreq	-24.888	8.513	-2.923	<0.01	
CC item model II					
	Estimate	Std.	Error	p-value	Adjusted R^2
Intercept	579.715	23.303	24.878	<0.001	0.3
TargetFreq	-17.809	4.867	-3.659	<0.01	
SequenceTP	43.518	18.859	2.308	<0.05	
CV item model II					
	Estimate	Std.	Error	p-value	Adjusted R^2
Intercept	838.777	90.123	9.307	<0.001	0.41
TargetFreq	-16.816	4.043	-4.160	<0.001	
BigramTypeFreq	-24.021	10.364	-2.318	<0.05	

Table 2. Fixed effects coefficients for the two models, CC and CV items (RTs=dependent variable).

When subject and prime were included as random factors, the pairwise comparison in the likelihood ratio test confirmed that the contribution of the sequence's TP increased significantly the predictability of the RTs patterns: $\chi^2(1)= 11.184$, $p= 0.0008$ in model I, $\chi^2(1)= 5.4403$, $p= 0.019$ in model II.

The average reaction times and the number of errors were positively and significantly correlated, though with an intermediate correlation coefficient ($r = .648$, $p < .01$). We thus tested the two models with Nerr as the dependent variable, in order to determine if the error rate was influenced by frequencies and probabilities to a different extent than response latencies.

With Nerr as the dependent variable, R^2 values were consistently lower than in the RTs simulations (Table 3), thus indicating that the error patterns were accounted for by our frequency and probability variables to a more limited extent. In particular, both model I and model II emphasized for the CC items the role of target frequency as the only significant predictor of errors, while for CV items an additional role of bigram frequencies (by token and by type, respectively) was found. Thus for the CV items, RTs and error rate produced consistent results.

	Estimate	Std.	Error	p-value	Adjusted R^2
CC item model I					
Intercept	7.714	1.349	5.717	<0.001	0.2
TargetFreq	-1.207	0.462	2.613	<0.05	
CV item model I					
	Estimate	Std.	Error	p-value	Adjusted R^2
Intercept	61.5379	16.8337	3.656	<0.01	0.39
TargetFreq	-1.7089	0.6736	-2.537	<0.05	
BigramTokenFreq	-4.4807	1.4855	-3.016	<0.01	
CC item model II					
	Estimate	Std.	Error	p-value	Adjusted R^2
Intercept	7.714	1.349	5.717	<0.001	0.19
TargetFreq	-1.207	0.462	-2.613	<0.05	
CV item model II					
	Estimate	Std.	Error	p-value	Adjusted R^2
Intercept	55.4970	15.0637	3.684	<0.01	0.33
TargetFreq	-1.8956	0.6757	-2.805	<0.01	
BigramTypeFreq	-5.1472	1.7322	-2.971	<0.01	

Table 3. Fixed effects coefficients for the two models, CC and CV items (Nerr=dependent variable).

3 Discussion

This work aimed to shed light on the role of TPs in a so far unstudied experimental environment, i.e., a lexical decision task with fragment priming. As the large part of studies on phonotactic probabilities focused on English, this work also added to the field with evidence from a poorly investigated language, Italian.

Fragment priming is known to be modulated not only by word frequency and the frequencies of words matching the fragment but also by top-down information conveyed by the prime: a fragment prime matching a unique morpho-lexical family is as effective as a stem prime, thus showing that priming acts as a cue for the properties displayed in the target (see e.g. Laudanna & Bracco, 2006, for Italian).

This study has shown that the priming effect when an initial fragment is available is influenced also by bottom-up variables; in particular, it depends on the probability with which the segments composing the fragment or the fragment-final consonant predict the occurrence of the consecutive consonant. Although to a lesser extent, the frequency with which bigrams and sequences occur (as types or tokens) in the lexicon also predict the subjects' behavior. Phonotactic probabilities thus turned out to predict the subjects' response to a large degree for many of the phonological environments tested in the current experiment, sometimes outperforming target frequencies, and consistently overtaking the contribution of the prime/target length ratio and of the prime frequency.

The results however suggested that the phonotactic probabilities in the case of consonant clusters were overall more important than in the case of consonant-vowel sequences; thus it must be concluded that the contribution of TPs in lexical recognition is not the same across phonological environments. Consonant clusters might play a particularly relevant role in lexical access, compared to CV sequences, as contemporary theories based on the principles of phonological and morphological naturalness also seems to predict (see e.g. Dressler & Dziubalska-Kolaczyk, 2006; Korecky-Kroell et al. 2014).

Additionally, for CC sequence the token frequencies (of the bigram and of the prime + C sequence) turned out to be relatively more important than the corresponding type frequencies, thus suggesting that the exposure to the number of occurrence of a cluster or of a segment sequence may be more important in lexical access than the exposure to the individual items containing them.

An additional issue concerns the role of TPs in morphologically complex words. According to some models, morphological parsing is necessary for lexical access and the prefix (in the case of prefixed words) has to be stripped away in order for the word to be recognized (from Taft & Forster, 1975 onwards). Assuming a condition in which the fragment prime coincides with a prefix, TPs would play the additional role of marking the morphological boundary during the priming event. According to the results of the current study, it appears to be of utmost importance to further verify whether prefixed and pseudo-prefixed words behave in the same way. In fact, models postulating morphologicl pre-parsing (e.g. Schreuder & Baayen, 1995) would suggest that high TPs will codetermine latencies for prefixed targets only, while if morphology does not affect word recognition, then the TPs between the fragment prime and the following segment composing the target will modulate latencies in prefixed and pseudo-prefixed words to the same extent.

A follow-up experiment will therefore test the contribution of phonotactic statistical knowledge in native speakers' access to complex word forms (specifically, prefixed nouns). Prefixed and pseudo-prefixed words will be used for that purpose. In particular, fragment primes will be selected according to two different conditions: in condition a) the targets are prefixed words and the fragment prime coincides with the prefix (e.g. *bis-bisnonna* 'grandmother'); in condition b) the targets are pseudo-prefixed words and no morphological boundary occurs between the initial fragment and the second part of the word (e.g. *per-perdente* 'loser'). Together with the current experiment, the experiment on prefixed and pseudo-prefixed words will determine whether or not the role of TPs is different when the target is a simplex word compared to when it is a prefixed word, and to when it is a pseudo-prefixed word. Different hypotheses may be put forward here, according to whether or not morphological boundaries affect the processing of consonant clusters (e.g., Calderone et al. 2014, Celata et al. 2015 in press), and according to the likelihood that a given sequence occurs as morpheme or as homographic non-morphological pattern (see Laudanna et al., 1994).

By describing phonotactic probability and frequency effects during word recognition, this study offers arguments to models of lexical access based on bottom-up processes such as cohort models for orthographic stimuli (see e.g. Johson & Pugh, 1994). The property of single consonants to predict the following segment then speeding up the recognition of the whole word, as an additional if not independent way to access words and their subparts, might also be discussed with reference to models that associate orthographic input units to semantic and lexical knowledge (from connectionist models such as in Harm & Seidenberg, 1999, to amorphous models such as in Baayen et al. 2011).

References

Harald R. Baayen. 2010. A real experiment is a factorial experiment? *The Mental Lexicon*, 5(1): 149-157.

Harald R. Baayen, Petar Milin, Dusica Filipovic Durdevic, Peter Hendrix and Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension on naive discriminative learning. *Psychological Review*, 118: 438-482.

Pier Marco Bertinetto, Cristina Burani, Alessandro Laudanna, Lucia Marconi, Daniela Ratti, C. Rolando and Anna Maria Thornton. 2005. *Corpus e Lessico di Frequenza dell'Italiano Scritto CoLFIS).* http://linguistica.sns.it/CoLFIS/Home.net

Basilio Calderone, Chiara Celata, Katharina Korecky-Kroell and Wolfgang U. Dressler. 2014. A computational approach to (mor)phonotactics: Evidence from German. *Language Sciences*, 46 (part A): 59-70.

Chiara Celata, Katharina Korecky-Kroell, Irene Ricci, and Wolfgang U. Dressler. 2015 (in press). Online processing of German (mor)phonotactic clusters by adults and adolescents. *Italian Journal of Linguistics*, 27(1).

Wolfgang U. Dressler and Katarzyna Dziubalska-Kolacyk. 2006. Proposing Morphonotactics. *Italian Journal of Linguistics*, 18: 249-266.

Katharina Korecky-Kroell, Wolfgang U. Dressler, Eva Maria Freiberger, Eva Reinisch, Karlheinz Moerth and Gary Libben. 2014. Phonotactic and morphonotactic processing in German-speaking adults. *Language Sciences*, 46 (part A): 48-58.

N.F. Johnson and K.R. Pugh. 1994. A cohort model of visual word recognition. *Cognitive Psychology*, 26: 240-346.

Alessandro Laudanna, Cristina Burani and Antonella Cermele. 1994. Prefixes as processing units. *Language and Cognitive Processes*, 9, 295-316.

Alessandro Laudanna and Giulia Bracco. 2006. Stem and fragment priming on verbal forms of Italian. *In Proceedings of the 5th International Conference on the Mental Lexicon* (Montreal, Canada, 11-13 October, 2006): 26.

Paul A. Luce and Nathan R. Large. 2001. Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, 16: 565-581.

Sven L. Mattys and Peter W. Jusczyk. 2001. Phonotactic cues for segmentation of fluent speech by infants. *Cognition*, 78: 91-121.

Mark Pitt and James McQueen. 1998. Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language*, 39: 347-370.

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

Holly L. Storkel. 2001. Learning nonwords: Phonotactic probabilities in language development. *Journal of Speech, Language, and Hearing Research*, 44: 1321–1337

Marcus Taft and Kenneth I. Forster. 1975. Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14: 638-647.

Michael Vitevitch, Paul Luce, J. Charles-Luce and D. Kemmerer. 1997. Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech*, 40: 47–62.

Michael S. Vitevitch and Paul A. Luce. 1999. Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language*, 40: 374-408.

Michael Vitevitch, Paul A. Luce, David B. Pisoni and Edward T. Auer. 1999. Phonotactics, neighborhood activation and lexical access for spoken words. *Brain and Language*, 68: 306-311.

Zipfian discrimination

Jim Blevins
University of Cambridge
`jpb39@cam.ac.uk`

Petar Milin
University of Novi Sad
Eberhard Karls Universität Tübingen
`petar.milin@uni-tuebingen.de`

Michael Ramscar
Eberhard Karls Universität Tübingen
`michael.ramscar@uni-tuebingen.de`

This talk outlines how form variation can be modelled in terms of equilibria between two dominant communicative pressures. The pressure to **discriminate** forms of a language enhances differences between expressions. Unchecked, this pressure can in principle lead to suppletion of the kind reported in languages such as Yélî Dnye (Henderson 1995). However, in most languages, the pressure towards maximally discriminative expressions is countered by the need to **extrapolate** from sparse input. It has long been known that corpora provide only a partial coverage of the forms of a language (inflectional and derivational). This talk presents evidence that the shortfall is far greater and far more systematic than previously appreciated, and that the coverage of the form variation remains sparse in corpora of up to one billion words. The sampling reported in this talk suggests that the forms in a corpus or encountered by a speaker exhibit a Zipfian distribution at all sample sizes.

The interaction of these pressures also accounts for the role of lexical neighbourhoods. Since most paradigms will be only partially attested, the organization of paradigms into neighbourhoods provides an analogical base for extrapolation.

The status of regularity

It is usually assumed that regularity in a linguistic system is desirable or normative and that suppletion and other irregularities represent deviations from the uniform patterns that systems (or their speakers) strive to maintain. From a discriminative perspective, the situation is exactly reversed. To the extent that patterns like suppletion enhance the discriminability of forms, they contribute to the communicative efficiency of a language. In a discriminative model, such as that of Ramscar et al. (2013), the only difference between overtly suppletive forms such as *mouse/mice* and more regular forms such as *rat/rats* is that the former serve to accelerate the rate at which a speakers' representation of a specific form/meaning contrast becomes discriminated from the form classes that express similar contrasts. Thus all learning serves to increase the level of suppletion in form-meaning mappings.

Moreover, standard cases of 'suppletion' are merely extreme instances of discriminative contrasts that seem ubiquitous at the sub-phonemic level. In the domain of word formation, Davis et al. (2002) found suggestive differences in duration and fundamental frequency between a word like *captain* and a morphologically unrelated onset word such as *cap*. Of more direct relevance are studies of inflectional formations. Baayen et al. (2003) found that a sample of speakers produced Dutch nouns with a longer mean duration when they occurred as singulars than as when they occurred as the stem of the corresponding plural. In a follow-up study, Kemps et al. (2005) tested speakers' sensitivity to prosodic differences, and concluded that "acoustic differences exist between uninflected and inflected forms and that listeners are sensitive to them" (Kemps et al. 2005: 441). Recent studies by Plag et al. (2014) find similar contrasts between phonemically identical affixes in English.

The role of discriminability

From a discriminative perspective, it is **regularity** that stands in need of explanation. Learning models offer a solution here as well. Unlike derivational processes, inflectional processes are traditionally assumed to be highly productive, defining uniform paradigms within a given class. Lemma size is thus not expected to vary, except where forms are unavailable due to paradigm 'gaps' or 'defectiveness'. Yet corpus studies suggest that this expectation is an idealization. Many potentially available inflected forms are unattested in corpora. As corpora increase in size, they do not converge on uniformly populated paradigms. Instead, they reinforce previously attested forms and classes while introducing progressively fewer new units. As shown in

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): **Word Structure and Word Usage**. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

Figure 1, the number of attested inflected noun variants decreases in all random samples, ranging from 1-million to 15-million hits, at which point the 850-million word StdeWaC corpus is essentially exhausted. As sample size increases, there is a marked attenuation in the steepness of the slope steepness, though it never becomes completely flat. This trend is extracted and presented in Figure 2, which plots number of attested forms on the X-axis and slopes of six trends from Figure 1 on the Y-axis. From this relationship we can infer that even if the corpus size were increased to infinity, it would never contain all possible inflected forms of every German noun. As shown in Figure 3, the forms of a language obey Zipf's law at all sample sizes. Speakers must be able to extrapolate from a partial – often sparse – sample of their language, and regular patterns subserve this need.

It takes a neighbourhood

In order for a collection of partial samples to allow the generation of unattested forms, the forms that speakers do know must be organized into systematic structures that collectively enable the scope of possible variations to be realized. These structures correspond to lexical neigbourhoods, whose effects have been investigated in a wide range of psycholinguistic studies (Baayen et al. 2006; Gahl et al. 2011). From the present perspective, neighbourhoods are not independent dimensions of lexical organization but, rather, constitute the creative engine of the morphological system, permitting the extrapolation of the full system from partial patterns. Interesting support for this perspective comes from the study reported in Milin et al. (2011). In this study, analogical extrapolation from a small set of nearest neighbors allowed a system to model the choice of masculine instrumental singular allomorph by Serbian speakers presented with nonce words. Regular paradigms thus enable language users to generate previously unencountered forms, not because they are the product of an explicit rule, or of any kind of explicit grammatical knowledge, but rather they are implicit in the distribution of forms and semantics in the language as a system, much as suggested by Hockett (1967: 221).

> in his analogizing … [t]he native user of the language … operates in terms of all sorts of internally stored paradigms, many of them doubtless only partial

References

Baayen, R. H., Feldman, L. B. & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language* **53**, 496–512.

Baayen, R. H., McQueen, J. M., Dijkstra, T. & Schreuder, R. (2003). Frequency effects in regular inflectional morphology: Revisiting Dutch plurals. In Baayen, R. H. & Schreuder, R. (eds.), *Morphological Structure in Language Processing*, Berlin: Mouton de Gruyter, 355–370.

Davis, M., Marslen-Wilson, W. D. & Gaskell, M. (2002). Leading up the lexical garden-path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception & Performance* **28**, 218–244.

Gahl, S., Yao, Y. & Johnson, K. (2011). Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech. *Journal of Memory and Language* **66(4)**, 789–806.

Henderson, J. E. (1995). *Phonology and Grammar of Yele, Papua New Guinea*. Pacific Linguistics B-112, Camberra: Pacific Linguistics.

Hockett, C. F. (1967). The Yawelmani basic verb. *Language* **43**, 208–222.

Kemps, J. J. K., Rachèl, Ernestus, M., Schreuder, R. & Baayen, R. H. (2005). Prosodic cues for morphological complexity: The case of Dutch plural nouns. *Memory & Cognition* **33(3)**, 430–446.

Milin, P., Keuleers, E. & Filipović Đurdjević, D. (2011). Allomorphic responses in Serbian pseudo-nouns as a result of analogical learning. *Acta Linguistica Hungarica* **58**, 65–84.

Plag, I., Homan, J. & Kunter, G. (2014). Homophony and morphology: The acoustics of word-final S in English. Ms, Heinrich-Heine-Universität, Düsseldorf.

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.

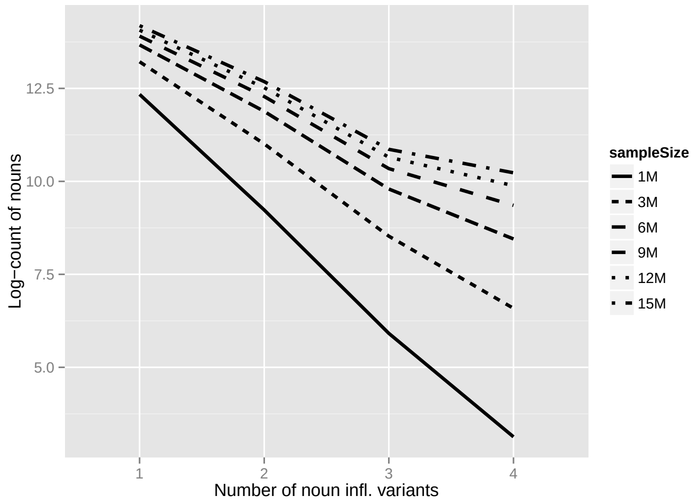


Figure 1: The paradigm non-filling pattern

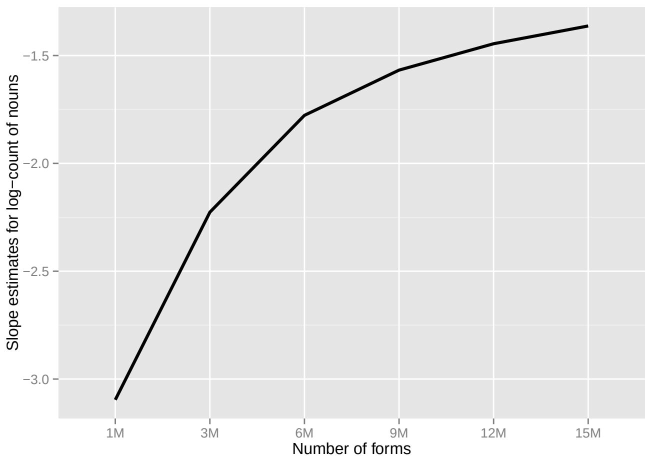


Figure 2: Asymptoting slopes

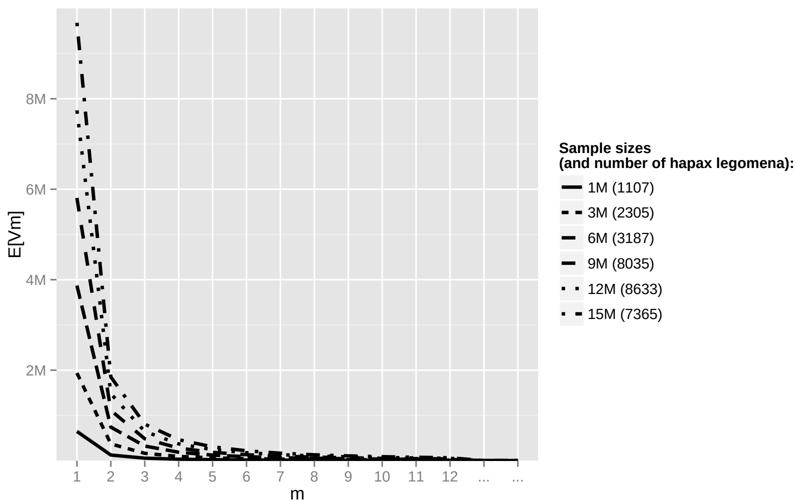


Figure 3: Zipf plot for randomly sampled words

Effects of processing complexity in perception and production. The case of English comparative alternation

Gero Kunter
English Language and Linguistics
Heinrich-Heine-Universität Düsseldorf
gero.kunter@uni-duesseldorf.de

Abstract

This paper discusses the effect of processing complexity on the English comparative alternation. The reported experiments show a processing advantage of the synthetic comparative in perception, but a preference of the analytic comparative in sentence production if the base adjective is cognitively complex. These results imply that perceptual complexity and complexity in production have diverging effects on the English comparative alternation. More generally, the paper calls for a fine-grained look at the role of processing complexity in areas of morphosyntactic variation.

1 Introduction

Most English comparatives are formed using either a synthetic form (e.g. *easier*) or an analytic form (e.g. *more important*). While most adjectives clearly prefer either the synthetic or the analytic comparative, there is a considerable number of adjectives which frequently take both forms, e.g. more *friendly* vs. *friendlier*. The decision for either form is influenced by several phonological, morphological, syntactic and semantic factors. For example, the probability of analytic comparatives increases with the number of morphemes in the adjective base. It is also higher if the comparative is in predicative than in attributive position, and it decreases with an increasing comparative/positive ratio (see Szmrecsanyi 2005, Hilpert 2008 and Mondorf 2009 for detailed discussions).

Mondorf (2009) argues that these factors are all part of a more general, audience-oriented compensatory mechanism called *more*-support: if the cognitive complexity of the adjectival base or its environment increases, speakers prefer the analytic comparatives, because they have a processing advantage over the corresponding synthetic form. For instance, an adjective that is morphologically complex is assumed to be also cognitively more complex than a simplex adjectives, and in order to compensate for this increased cognitive complexity, speakers may prefer the analytic comparative over the synthetic alternative.

Yet, there is only little psycholinguistic research that investigated this assumed processing advantage of analytic forms. A notable exception is Boyd (2007, ch. 2) who conducted a self-paced reading experiment to investigate processing differences between synthetic and analytic comparatives. Indeed, he reports shorter reaction times for the sentences containing analytic comparatives, but due to the experimental design, this evidence is only indirect and allows for alternative interpretations. As yet, then, there is only limited empirical evidence for the assumption that analytic comparatives are easier to process than synthetic comparatives. In addition, as pointed out by Mondorf (2014, 201), it is still an unresolved issue whether *more*-support is a response to increased processing loads in production or in perception.

This paper addresses these two issues. First, it presents the results from a perception experiment which tested whether analytic comparatives are indeed easier to process for listeners. Contrary to this hypothesis, the reaction times show that analytic comparatives have a processing disadvantage in perception. Then, a production experiment is discussed which elicited spoken sentences containing a comparative construction. The analysis reveals that the processing complexity is a significant predictor of the comparative alternation: with increasing complexity of the base adjective, the probability of analytic comparatives increases. Thus, the paper argues that speakers and listeners process the English comparative variants differently, and that it is the speaker who benefits from a compensatory use of *more* comparatives.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

2 Comparative variation in perception

2.1 Method

31 native speakers of Canadian English participated in an auditory decision task in which they had to decide whether the acoustic stimuli was an existing English form. The set of stimuli contained the analytic and synthetic comparative form for 60 adjective types with at least 5 attestations for both forms in the Corpus of Contemporary American English (Davies 2008-). The stimuli were produced by a male speaker of Canadian English with phonetic training. He was instructed to produce the stimuli in citation form with a single accent on the primary stressed syllable of the base adjective in both types of stimuli. Accordingly, *more* was produced stressed, but unaccented.

Alongside the 2 $\times$ 60 = 120 synthetic and analytic comparatives, the set of stimuli also included 360 distractors. Some of the distractors combined *more* with non-existing words, others combined the adjective bases with the illegal suffix *-ic*. In addition, the set of distractor items contained non-existing words ending in *-er* as well as existing words and complex words. Examples of the test stimuli are given in (1a), and distractor examples are given in (1b).

(1) a. *colder*, *happier*, *yellower*
more cold, *more wealthy*, *more yellow*

b. $*$*coldic*, *more* $*$*gorsty*, $*$*rilker*
on wire, *chasting*

2.2 Results

Figure 1 displays the density estimate for the distribution of reaction times. The solid and the dashed lines correspond to the results for synthetic and analytic stimuli, respectively.

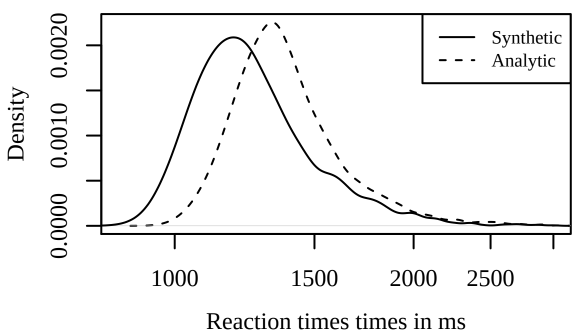


Figure 1: Density estimate of reaction times in perception experiment

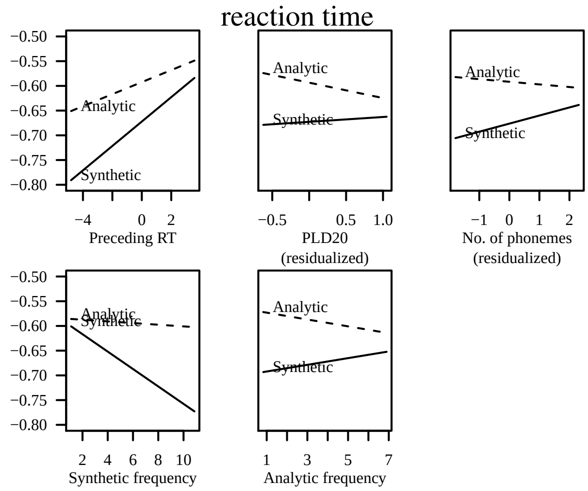


Figure 2: Partial effects of significant interactions of Class on reaction times

The density estimate suggests that reaction times are, in general, higher for analytic comparatives than for synthetic comparatives. This visual interpretation is supported by a linear mixed-effects regression model with reaction times as the dependent variable (in order to fulfill the linearity assumption of the linear model, the reaction times were power-transformed with λ = -1.52, see Box and Cox 1964). The main predictor was the factor Class (with values Synthetic and Analytic). Additional predictors addressed several influences that may be expected affect the reaction times: the subject-specific variables Handedness, Sex, and Age, the experimental variables Trial number and Reaction time in previous trial, (Preceding RT, see Baayen and Milin 2010 for a discussion), phonological variables (Metrical structure of base, residualized Number of phonems), and the lexical variables Number of phonological neighbours, Mean RT of base adjective, residualized Phonological Levenshtein distance (PLD20, all three from Balota et al. 2007), Age of acquisition (from Kuperman et al. 2012), Frequencies of base, Analytic comparative, Synthetic comparative (from COCA), Inflectional entropy (cf. Moscoso del Prado Martín et al. 2004). With the exception of the three Subject predictors, the initial model contained interactions between Class and the other predictors. Finally, random intercepts were included for the factors Subject and Adjective base.

After removal of insignificant predictors, the final model reports significant interactions between stimulus Class and Preceding RT, PLD20, Number of phonemes, Synthetic frequency, and Analytic frequency. Figure 2 displays the partial effects for these interactions. The vertical axis shows the transformed reaction times; higher values correspond to longer reaction times.

In agreement with figure 1, the partial effects reveal significantly lower estimates for the synthetic stimuli (solid lines) than for the analytic stimuli (dashed lines). This is true even in the most adverse conditions (e.g. in cases in which the synthetic comparative of a comparative is attested only very rarely in a linguistic corpus, left edge of lower right panel in figure 2).

3 Comparative variation in production

3.1 Method

41 native speakers of Canadian English participated individually in a spoken sentence completion task. The task used the same set of 60 adjectives as in the perception experiment above, but none of the participants in the production experiment had also participated in the previous task. Participants were first shown a context sentence containing the adjective in the positive. After a key press, an incomplete target sentence containing a blank and one or more target words appeared also on the screen. The participants were instructed to use the target words to fill the blank in the sentence. If necessary, they could also use additional words to complete the sentence. The sentences were constructed in such a way that a comparative construction was the most likely target for completion, but participants were not explicitly instructed to use comparatives. The structure of the incomplete sentences was the same in all trials. The subject was a simple noun phrase, followed by a copula verb. The blank to be filled followed in predicative position. This design ensured that the context-dependent factors reported in the literature such as the increased probability of analytic comparatives in predicative position were held constant for all adjectives. Example (3) shows the experimental trial for the target adjective *wealthy*.

(2) The duke is wealthy.
Yet, the king is ________.
WEALTHY

The experiment also contained 105 distractor trials that had a similar structure, but which did not contain adjectives as the target words.

3.2 Reaction times

In order to be able to investigate the effect of the processing complexity of the base adjective on the preferred comparative variant, the same 41 speakers first participated in a visual lexical decision task that gathered reaction times for the 60 target adjectives, as well as 150 other existing and non-existing distractor items. The participants were not informed about the purpose of this task, and there were at least 14 days for each participant between the lexical decision task and the production experiment. The reaction times obtained in this task were pooled for each adjective, and the median was calculated.

3.3 Results

For most of the adjectives, the completion task was successful in obtaining comparative responses from the 41 speakers. However, two participants produced hardly any comparative in the task, and were therefore excluded from the data set. 6 out of the 60 adjectives were excluded because the responses contained almost exclusively synthetic or analytic comparatives, or because the context sentence did not elicit a considerable number of comparative responses. 747 out of the remaining $39 \times 54 = 2106$ responses contained a synthetic comparative (35 %), 843 contained an analytic comparative (40 %). The remaining 516 responses (25 %) did not contain a comparative construction, and were discarded. There was notable variation between the two variants both across and within items, which indicates that English comparative variation is indeed a highly non-deterministic field that is apparently affected by both speaker-dependent and adjective-dependent factors.

Logistic general additive mixed-effects models (cf. Wood 2006) were used to investigate the relation between the median RTs and the individual responses. These models have the advantage of revealing statistically significant effects of the independent variable on the dependent even if the relation between them is not a linear one. For instance, there could a threshold in the reaction times up to which speakers strongly prefer the synthetic comparative, but beyond which they shift to analytic comparatives in a nearly categorical way. In such a case, a linear model might fail to detect this non-linear effect of RTs on the responses.

Two models were fitted: a null model which contained only a random effect for speaker, and a model with an additional smooth term for the effect of the median RTs. If processing complexity has a notable effect on speaker responses, the smooth term should turn out to be statistically significant, and the predictive accuracy of the model should improve by the addition of the term. As table 1 shows, this is indeed the case. While the null model has a total predictive accuracy of about 69 %, the addition of the smooth term for median RTs increases the accuracy by 5.6 %. There is a larger increase of predictive accuracy for analytic responses than for synthetic responses (7.1 % vs. 3.9 %).

	Synthetic	Analytic	Total
Null model	515	580	1095
	(68.9%)	(68.8%)	(68.9%)
Model with	544	640	1184
smooth term	(72.8%)	(75.9%)	(74.5%)

Table 1: Correctly predicted responses in the sentence completion task.

Figure 3 illustrates the contribution of the smooth term to the model. The vertical position of the regression line indicates the predicted probability of analytic responses for the median RTs shown on the horizontal axis. The shaded area indicates the 95 % confidence band. As the figure shows, the relation between processing complexity and comparative preference is indeed nonlinear: speakers strongly prefer the synthetic comparative for adjectives with very low RTs, but tend to favor the analytic comparative for adjectives with RTs larger than 600 ms. In sum, the production experiment shows that the processing complexity of the base adjective has an effect on the preference of analytic comparatives by speakers.

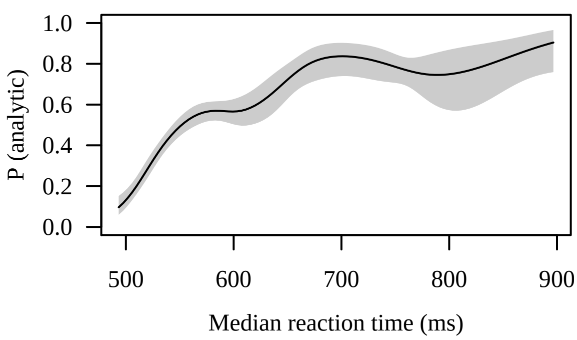


Figure 3: Effect of median reaction time on the probability of analytic responses.

4 Discussion and conclusion

The results from the first experiment show that synthetic comparatives have a clear perceptual processing advantage over the analytic correspondents. Even in conditions in which the morphological form is particularly difficult to process, the average reaction time is still faster than that for the phrasal variants. This finding makes it rather unlikely that the use of analytic comparatives in cognitively demanding environments benefits the listener. Yet, the findings from the production experiment reveal a significant relation between the selected comparative form and the processing difficulty of the adjective in question. For cognitively more complex adjectives which take longer to process, the analytic comparative is preferred, suggesting that speakers resort to the phrasal alternative if processing demands are relatively high.

One aspect to keep in mind is that lexical decision tasks like those used above to collect reaction times have a strong focus on form processing, while they are less informative about functional processing (see Yap et al. 2011 for a discussion). Even if the perception experiment has shown that the analytic form is more difficult to process for listeners, the higher explicitness of the *more* comparative may still make the comparative function more accessible for listeners than the -*er* comparative, which is also suggested by Mondorf (2009, 6). The experiments reported here do not address this issue of the comparative alternation, but looking at functional accessibility offers a promising venue of future research.

To conclude, the results imply that speakers and listeners process analytic and synthetic comparatives differently: while the morphological form is easier to process for listeners, the phrasal form has benefits for the speaker. More generally, these findings also contribute toward our understanding of morphosyntactic exponence. It is frequently argued (e.g. in McWhorter 2001) that analytic forms are less complex than synthetic forms, with consequences for fields such as the structure of contact languages or the diachronic development of a language. This paper is one of the few that explicitly address the processing efficiency of grammatical variants where one form is morphological and the other syntactic in nature. The findings suggest that the discussion of the alledged complexity of synthetic forms may also need to take into account different demands of speakers and listeners.

Acknowledgments

This work was supported by the *Deutsche Forschungsgemeinschaft* (grant KU 2896/1-1). I wish to thank Ben Tucker (University of Alberta, Edmonton) for making available to me the facilities of the Alberta Phonetics Laboratory for the experiments reported in this paper.

References

R. Harald Baayen and Petar Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research*, 3(2):12–28.

David A. Balota, Melvin J. Yap, Michael J. Cortese, Keith A. Hutchison, Brett Kessler, Bjorn Loftis, James H. Neely, Douglas L. Nelson, Greg B. Simpson, and Rebecca Treiman. 2007. The English Lexicon Project. *Behavior Research Methods*, 39(3):445–459.

Jeremy Boyd. 2007. *Comparatively speaking. A psycholinguistic study of optionality in grammar*. Ph.D. thesis, University of California, San Diego.

George E. P. Box and David R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, 26(2):211–252.

Mark Davies. 2008–. *The Corpus of Contemporary American English (COCA): 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/.

Martin Hilpert. 2008. The English comparative. language structure and language use. *English Language and Linguistics*, 12(3):395–417.

Victor Kuperman, Hans Stadthagen-Gonzalez, and Marc Brysbaert. 2012. Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, 44(4):978–990.

John H. McWhorter. 2001. The world's simplest grammars are creole grammars. *Linguistic Typology*, 5:125–166.

Britta Mondorf. 2009. *More support for* more-*support*. John Benjamins, Amsterdam.

Britta Mondorf. 2014. Apparently competing motivations in morpho-syntactic variation. In Brian MacWhinney, Andrej Malchukov, and Edith Moravcsik, editors, *Competing motivations in grammar and usage*, pages 209–228. Oxford University Press, Oxford.

Fermín Moscoso del Prado Martín, Aleksandar Kostić, and R. Harald Baayen. 2004. Putting the bits together. An information theoretical perspective on morphological processing. *Cognition*, 94(1):1–18.

Benedikt Szmrecsanyi. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory*, 1(1):113–150.

Simon N. Wood. 2006. *Generalized Additive Models. An introduction with R*. Chapman & Hall/CRC, Boca Raton, FL.

Melvin J. Yap, Sarah E. Tan, Penny M. Pexman, and Ian S. Hargreaves. 2011. Is more always better? effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, 18(4):742–750.

Lexical emergentism and the “frequency-by-regularity” interaction

Claudia Marzi **Marcello Ferro** **Vito Pirrelli**
Institute for Computational Linguistics - National Research Council - Pisa
{claudia.marzi,marcello.ferro,vito.pirrelli}@ilc.cnr.it

Abstract

In spite of considerable converging evidence of the role of inflectional paradigms in word acquisition and processing, little efforts have been put so far into providing detailed, algorithmic models of the interaction between lexical token frequency, paradigm frequency, paradigm regularity. We propose a neuro-computational account of this interaction, and discuss some theoretical implications of preliminary experimental results.

1 Introduction

Over the last fifteen years, growing evidence has accrued of the role of morphological paradigms in the developmental course of word acquisition. Children have been shown to be sensitive to sub-regularities holding among paradigm cells (see, among others, Orsolini et al., 1998; Laudanna et al., 2004 on Italian; Dabrowska, 2004, 2005 on Polish; and Labelle and Morris, 2011 on French). In line with this evidence, and contrary to both rule-based (e.g. Pinker and Ullman, 2002; Albright, 2002) and connectionist approaches to word acquisition (Rumelhart and McClelland, 1986), no unique paradigm cell can be identified as the base source of all inflected forms produced by the speaker, but the structure of the entire paradigm is understood to play a fundamental role in both word acquisition and processing.

Such evidence supports a view of the mental lexicon as an emergent integrative system, whereby words are concurrently, redundantly and competitively stored (Alegre and Gordon, 1999; Baayen et al., 2007). The view assumes that all word forms are memorised in the lexicon, thus making no distinction between regular and irregular inflected forms, or between uniquely stored bases and all other non-base forms produced by the speaker on demand (see Baayen, 2007; Marzi, 2014; for a recent overview). In addition, to capture the fact that words encountered frequently exhibit different lexical properties from words encountered relatively infrequently, any model of lexical access must assume that accessing a word in some way affects the access representation of that word (e.g. Foster, 1976; Marslen-Wilson, 1993; Sandra, 1994).

In spite of such a wealth of converging evidence, however, little efforts have been put so far into providing detailed, algorithmic models of the interaction between word frequency, paradigm frequency, paradigm regularity and lexical familiarity in word acquisition and processing. We offer here such an algorithmic account, and discuss some theoretical implications on the basis of computational simulations.

2 The computational model

In the present contribution, we use Temporal Self-organising Maps (TSOMs) to simulate dynamic effects of lexical storage, organisation and competition.

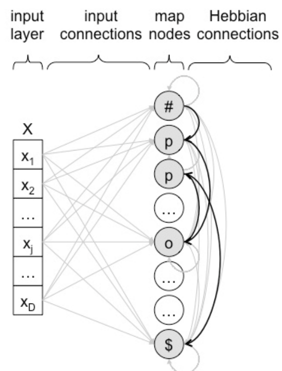


Figure 1. An integrated activation pattern for the input string “*#pop$*”. Note that two distinct, but topologically neighbouring nodes respond to the two *p*’s in *pop*, bearing witness to the process of selective sensitivity to time-bound instances of the same symbol type. For simplicity, only the nodes that are most highly activated by each input symbol are shaded and tagged with that symbol.

TSOMs, a variant of classical Kohonen’s SOMs (Kohonen, 2001), are dynamic memories that are trained to store and classify time-series of symbols through patterns of activation of fully interconnected nodes (Koutnik, 2007; Ferro et al., 2010; Pirrelli et al., 2011; Marzi et al., 2012). Map nodes mimic neural clusters, with inter-node connections representing neuron synapses whose weights determine the amount of influence that the activation of one node has on another node (Fig. 1). Each map node receives input

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

connections from an input layer where individual symbols making up a word are presented one at a time, in their order of appearance. Input connections thus convey information of the current input stimulus to map nodes. Hebbian connections, on the other hand, are strengthened each time two nodes are activated at consecutive time ticks, conveying the probabilistic expectation that one node will be activated soon after another node is activated.

When a symbol is shown on the input layer at a certain time tick, all map nodes are fired synchronously, their overall pattern of activation representing the processing response of a TSOM to the symbol at that time tick. Due to principles of topological organisation of map's responses, similar input stimuli (i.e. two instances of the same symbol in different contexts) tend to be associated with largely overlapping memory traces (e.g. the two *p* nodes activated by *pop* in Fig. 1). During training, nodes get gradually specialised to respond most strongly to specific time-bound instantiations of symbols, while remaining relatively inactive in the presence of other stimuli. A recurrent activation pattern associated with an input symbol occurring in a specific context can thus be seen as the map's memory trace for that symbol in that context.

An input word is administered to a TSOM as a time series of symbols, i.e. a sequence of letters or sounds presented on the input layer one at a time. The map's response to a word stimulus is the overall activation pattern obtained through integration of the activation patterns triggered by the individual symbols making up the word (see Fig. 1 for a simplified example with the word *pop*). Accordingly, if two input strings present some symbols in common (e.g. *pop* and *cop*, *write* and *written*), they will tend to activate largely overlapping patterns of strongly responsive nodes. Like in the case of individual symbols, the integrated activation pattern for an input word is, at the same time, the systematic processing response of the map to an input stimulus, and the word's memorised representation (or memory trace) in the map.

To investigate issues of "frequency-by-regularity" interaction (Ellis and Smith, 1998), we compared two sets of parallel experiments carried out on German verb paradigms (Marzi et al., 2014) and Italian verb paradigms. By keeping constant some input conditions, such as selection of paradigm cells and degrees of morphological redundancy within training paradigms, while varying others, such as the frequency distribution of paradigm members, we can investigate the relative contribution of input factors to the timing and pace of lexical acquisition and suggest an explanatory account of their interaction.

3 Experimental evidence

Fifty German and fifty Italian verb (sub)paradigms were selected among the most highly ranked paradigms by cumulative frequency in a reference corpus (CELEX Lexical database for German, Baayen et al., 1995; Paisà Corpus for Italian, Lyding et al., 2014). For each paradigm, an identical set of 15 cells was used for training, for an overall number of 750 inflected forms for each language. Each data set was administered to the map for 100 epochs under two different training regimes: a uniform distribution (UD: 5 tokens per word), and a function of real word frequency distributions in the reference corpus (SD: tokens are in the range of 1 to 1000). By varying frequency and comparing the inflectional complexity of training data across the two experiments, we expected to gain some insights into the interplay between morphological regularity (defined by levels of predictability in stem and ending allomorphy of training data in the two languages) and word frequency in word acquisition. After training, we monitored the behaviour of the four resulting TSOMs (namely UD Italian, SD Italian, UD German and SD German) by controlling the time of acquisition of individual words, the time of acquisition of entire paradigms, and their acquisitional time span. For our present purposes, we define the time of acquisition of a single word as the training epoch whence a TSOM can accurately recall the word in question from its memory trace. Recall is a difficult task that requires that the map has developed a clear notion of how to unfold a synchronous activation pattern (the word's memory trace) into a sequence of nodes representing the correct letters making up the word, in the appropriate order. Likewise, for each paradigm, its time of acquisition by a map is the mean acquisition epoch of all forms belonging to the paradigm.

As a general trend, TSOMs acquire word forms by token frequency, with higher-frequency words being successfully recalled at earlier learning epochs. However, when it comes to the actual timing of paradigm acquisition, things get considerably more complex, with the notion of morphological regularity interacting non-trivially with token frequency distributions. In fact, in both

German and Italian, the vast majority of paradigms are acquired earlier (p<.005) in a UD regime than in an SD regime (Fig. 2).

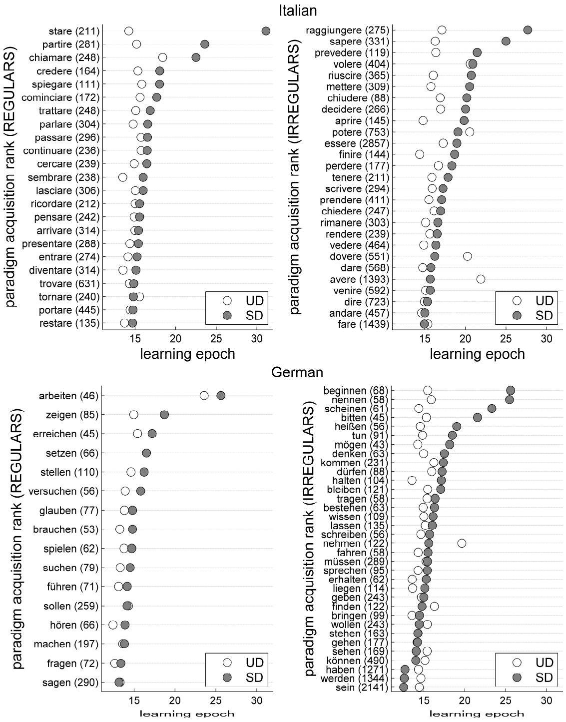


Figure 2: Time course of regular (left) and irregular (right) paradigms ranked by increasing learning epoch under SD (grey circles) and UD (white circles) regimes for both Italian (top) and German (bottom). Values are averaged across 5 map instances for each type.

4 Frequency by regularity interaction

Our simulations show that, in both languages, word forms in regular paradigms tend to be acquired earlier (significantly earlier learning epochs, p<.001), and regular paradigms are acquired more quickly (significantly shorter learning spans, i.e. lower number of epochs between the acquisition time of the first and the last member of a paradigm, p<.005) than irregular paradigms are. In German data, regular paradigms are less sensitive to token frequency effects than irregular paradigms are, as witnessed by the strong correlation (r=.95, p<.00001) between the time course of acquisition of regular paradigms in SD and UD regimes (Fig. 2, bottom left panel). Token frequency affects the acquisition of regular paradigms to a lesser extent than the acquisition of irregular ones, because regular stems can take advantage of their cumulative frequency across the whole paradigm. In fact, forms in regular paradigms exhibit a significant correlation between stem cumulative frequency and time of acquisition (r=-.40, p<.00001). Similarly, also German irregular paradigms, which exhibit a predictable stem allomorphy due to a limited number of alternants, show a correlation between stem cumulative frequency and acquisition time (r=-.24 p<.00001).

Conversely, in Italian, where verb conjugation exhibits more extensive and less predictable patterns of allomorphy than in German (Pirrelli, 2000), acquisition of irregular paradigms does not appear to benefit from stem cumulative token frequencies (r=.01, p>.5). This suggests that extensive allomorphy in a paradigm tends to minimise the influence of cumulative frequency on its acquisition, and isolated forms can only take advantage of their own token frequency, while taking no advantage of the frequency boost provided by other cells of the same paradigm. As a result, Italian irregular paradigms are acquired significantly (p<.005) later than their German homologues.

Our data cannot be explained away as a simple by-product of word-frequency effects. Experiments provide, in fact, evidence of interactive processing effects in word acquisition, whereby morphological regularity modulates frequency. Data analysis shows that recurrent patterns appear to determine global co-organisation of stored word forms and distributed, overlapping memory traces, which ultimately favour generalisation in lexical acquisition. Forms containing recurrent patterns can take advantage of the memory traces shared with other related forms, namely forms sharing the same stem, and connections between the nodes making up their memory traces are strengthened since patterns are shown more often in training, similarly to high-frequency isolated words.

This is particularly true for regular, highly entropic paradigms, i.e. those regular paradigms whose members exhibit uniform frequency distributions, and for irregular highly systematic paradigms. Conversely, where memory traces overlap less systematically, this effect is considerably reduced, as witnessed by the difference in time of acquisition between regular and irregular paradigms, particularly in Italian conjugation.

In TSOMs, the effects are the dynamic result of two interacting dimensions of memory self-organisation: (i) the syntagmatic or linear dimension, which controls the level of predictability and entrenchment of memory traces in the lexicon through the probabilistic distribution of weights over inter-node Hebbian connections; and (ii) the paradigmatic or vertical dimension, which controls for the number of similar, paradigmatically-related word forms that get co-activated when one member of a paradigm is input to the map (Pirrelli et al., 2014).

High-frequency words develop quick entrenchment of Hebbian connections, which eventually cause high levels of node activation in their memory traces and sparser co-activation of memory traces of other words. Strong connections and high activation levels mean high expectations for frequently activated memory traces, which are thus recalled more easily and are less confusable with other neighbouring words. Likewise, in regular and sub-regular paradigms, sharing memory traces can strengthen connections and raise node activation levels, since all related forms can take advantage of the memory traces shared with other members of the same paradigm.

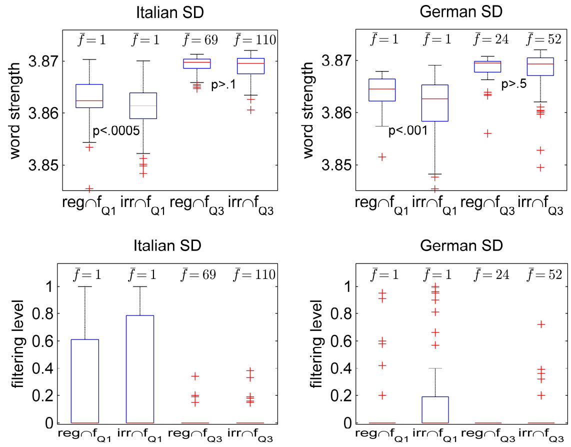


Figure 3: Levels of activation strength (top) and filtering (bottom) for Italian (left) and German (right), for four regularity-by-frequency classes. Low-frequency is set below the first quartile of frequency distributions in the two training sets, while high-frequency being set above the third quartile.

This dynamic provides an algorithmic account of the observation that regularity favours acquisition of both high- and low-frequency words, as shown in Fig. 3, where we compare average levels of activation for four classes of training word forms: low-frequency regulars, low frequency irregulars, high-frequency regulars and high-frequency irregulars.[1]

Activation levels of low-frequency words appear to be significantly stronger within regular paradigms than within irregular paradigms (Fig. 3, top). Stronger activation levels make patterns less confusable and easier to be accessed, as witnessed by the lower level of filtering[2] required for activation patterns to be recalled accurately (Fig. 3, bottom). We observe, in fact, a highly significant correlation (r=.49, p<.00001 for both datasets) between levels of filtering and words' learning epochs.

High-frequency words predictably show higher activation levels than low-frequency words, with an interesting difference of the interaction of frequency and activation levels of regulars and irregulars. High-frequency, highly irregular words (e.g. German *ist* or Italian *è*) are stored in isolation, with highly-activated memory nodes and no co-activation with other words. As a result, they require little filtering to be recalled and are acquired considerably quickly. High-frequency regular paradigms, despite in both Italian and German training sets their average frequency is nearly half the average frequency of high-frequency irregulars, show comparable levels of activation with high-frequency irregulars, due to the facilitatory effect of having more words that consistently activate the same pattern of nodes.

This evidence shows that regularity indeed modulates the interaction between frequency and activation strength, and it gives a strong indication that acquisition of regulars is typically paradigm-based, whereas acquisition of irregulars is mostly item-based.

Surely, as the notion of paradigm regularity is inherently graded, some verb systems show higher sensitivity to these effects than others. This is illustrated by German sub-regular paradigms, which present fewer and more predictable stem alternants than Italian sub-paradigms, and thus larger stem-sharing word families. Accordingly, TSOMs allocate comparatively higher levels of activation to low-frequency German sub-regulars and acquire them earlier than their Italian homologues.

The evidence reported here establishes, in our view, an important connection between aspects of morphological structure, frequency distributions of words in paradigms, and lexical acquisition in concurrent, competitive storage. Acquisition of redundant morphological patterns play an increasingly important role in an emergent lexicon, shifting acquisitional strategies from rote memorisation (typical of irregular low-entropy paradigms) to dynamic memory-based generalisation.

[1] Frequency thresholds are set below the first quartile (low frequency) and above the third quartile (high frequency) in the frequency distribution of training word forms.

[2] Filtering an integrated activation pattern refers to the process of bringing down to zero the levels of activation of nodes that do not reach a set threshold.

References

Maria Alegre and Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40: 41-61.

Harald R. Baayen, Richard Piepenbrock and Leon Gulikers. 1995. The CELEX Lexical Database (CD-ROM). Philadelphia: Linguistic Data Consortium.

Harald R. Baayen. 2007. Storage and computation in the mental lexicon. In G. Jarema and G. Libben (eds.), *The Mental Lexicon: Core Perspectives*, 81-104. Amsterdam: Elsevier.

Lucia Colombo, Alessandro Laudanna, Maria De Martino and Cristina Brivio. 2004. Regularity and/orconsistency in the production of the past participle? *Brain and Language*, 90: 128-142.

Ewa Dabrowska. 2004. Rules or schemata? Evidence from Polish. *Language and cognitive processes*, 19 (2): 225–271.

Ewa Dabrowska. 2005. Productivity and beyond: mastering the Polish genitive inflection. *Journal of child language*, 32: 191-205.

Nick C. Ellis and Richard Schmidt. 1998. Rules or Associations in the Acquisition of Morphology? The Frequency by Regularity Interaction in Human and PDP Learning of Morphosyntax. *Language and Cognitive Processes*, 13: 307-336.

Marcello Ferro, Giovanni Pezzulo and Vito Pirrelli. 2010. Morphology, Memory and the Mental Lexicon. In Pirrelli, V. (ed.), *Lingue e Linguaggio*, IX(2): 199-238.

Jan Koutnik. 2007. Inductive Modelling of Temporal Sequences by Means of Self-organization. In *Proceeding of International Workshop on Inductive Modelling*. Prague: 269-277.

Marie Labelle and Lori Morris. 2011. The acquisition of a verbal paradigm: Verb Morphology in French L1 children. *Prépublication*. (Montréal, Québec, Canada, UQAM, département de linguistique). http://www.archipel.uqam.ca/3992/1/Labelle-Morris_AcquisitionVerbalParadigm.pdf

Verena Lyding, Egon Stemle, Claudia Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell'Orletta, Henrik Dittmann, Alessandro Lenci and Vito Pirrelli. 2014. The PAISÀ Corpus of Italian Web Texts. In F. Bildhauer and R. Schäfer (eds.) *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*: 36-43. Gothenburg.

Claudia Marzi, Marcello Ferro and Vito Pirrelli. 2012. Word alignment and paradigm induction. *Lingue e Linguaggio*, XI (2): 251-274.

Claudia Marzi, Marcello Ferro and Vito Pirrelli. 2014. Morphological structure through lexical parsability. *Lingue e Linguaggio*, XIII (2): 263-290.

Claudia Marzi. 2014. Models and dynamics of the morphological lexicon in mono- and bilingual acquisition. *Unpublished PhD Dissertation.* University of Pavia. www.comphyslab.it/redirect/?id=claudia.marzi.en_phd

Margherita Orsolini, Rachele Fanari and Hugo Bowles. 1998. Acquiring regular and irregular inflections in a language with verb classes. *Language and cognitive processes*, 13(4): 425-464.

Steven Pinker and Michael Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Science*, 6: 456-463.

Vito Pirrelli, Claudia Marzi and Marcello Ferro. 2014. Two-dimensional Wordlikeness Effects in Lexical Organisation. In: Basili R., Lenci A., Magnini B. (eds.) *Proceedings of the First Italian Conference on Computational Linguistic*, December 9-11, 2014. 301-305, Pisa: Pisa University Press.

Vito Pirrelli, Marcello Ferro and Basilio Calderone. 2011. Learning paradigms in time and space. Computational evidence from Romance languages. In M. Maiden, J. C. Smith, M. Goldbach and M. O. Hinzelin (eds.), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, 135-157. Oxford: Oxford University Press.

Vito Pirrelli. 2000. *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: Istituti editoriali e poligrafici internazionali.

Kim Plunkett and Virginia Marchman. 1993. From rote learning to system building – acquiring verb morphology in children and connectionist nets. *Cognition*, 48: 21-69.

David E. Rumelhart and James L. McClelland. 1986. On learning the past tense of English verbs. In McClelland, J.L. and Rumelhart, D.E. (eds.) *Parallel distributed processing*, 217-270. Cambridge: MIT Press.

Morphological Priming in German: The Word is Not Enough (Or Is It?)

Sebastian Padó[*] **Britta D. Zeller**[*] **Jan Šnajder**[†]
[*]Stuttgart University, Institut für maschinelle Sprachverarbeitung
Pfaffenwaldring 5b, 70569 Stuttgart, Germany
[†]University of Zagreb, Faculty of Electrical Engineering and Computing
Unska 3, 10000 Zagreb, Croatia
{pado,zeller}@ims.uni-stuttgart.de jan.snajder@fer.hr

Abstract

Studies across multiple languages show that overt morphological priming leads to a speed-up only for transparent derivations but not for opaque derivations. However, in a recent experiment for German, Smolka et al. (2014) show comparable speed-ups for transparent and opaque derivations, and conclude that German behaves unlike other Indo-European languages and organizes its mental lexicon by morphemes rather than lemmas. In this paper we present a computational analysis of the German results. A distributional similarity model, extended with knowledge about morphological families and without any notion of morphemes, is able to account for all main findings of Smolka et al. We believe that this puts into question the call for German-specific mechanisms. Instead, our model suggests that cross-lingual differences between morphological systems underlie the experimentally observed differences.

1 Semantic and Morphological Priming

Priming is a general property of human language processing: it refers to the speed-up effect that a stimulus can have on subsequent processing (Meyer and Schvaneveldt, 1971). This effect is assumed to result from an activation (in a broad sense) of mental representations, and priming is a popular method to investigate properties of the mental lexicon. The original study by Meyer and Schvaneveldt established *lexical priming* (*nurse* → *doctor*), but priming effects have also been identified on other linguistic levels, such as syntactic priming (Bock, 1986) and morphological priming (Kempley and Morton, 1982).

A recent study by Smolka et al. (2014) investigated overt *morphological priming* on prefix verbs in German, where the base verb and derived verb can be semantically related (transparent derivation: *schließen – abschließen (close – lock)*) or not (opaque derivation: *führen – verführen (lead – seduce)*). Experiment 1, an overt visual priming experiment (300 ms SOA) involved 40 six-tuples that paired up a base verb with five prefix verbs of five prime types (see Figure 1). The verbs were normed carefully, e.g., for association, to exclude confounding factors. The authors reported three main findings: (a), no priming for Form and Unrelated; (b), no priming for Synonymy; (c), significant priming of the same strength for both Transparent and Opaque Derivation.

These findings suggest that morphological priming on German prefix verbs use a mechanism that is different from lexical priming, which assumes that the strength of the semantic relatedness is the main determinant of priming – i.e., lexical priming would predict finding (a), but neither (b) nor (c). The findings by Smolka et al. are also at odds with overt priming patterns found in similar experimental setups for other languages such as French (Meunier and Longtin, 2007) and Dutch (Schriefers et al., 1991), where patterns were found to be indeed consistent with lexical priming. Smolka et al. (2014) interpret this divergence as evidence for a German *Sonderweg*: the typological properties of German (separable prefixes, morphological richness, many opaque derivations) are taken to suggest a *morpheme*-based organization of the mental lexicon more similar to Semitic languages like Hebrew or Arabic than to other Indo-European languages.

Our paper investigates this claim on the computational level. We present a simple model of corpus-based word similarity, extended with a database of *morphological families*, that is able to predict the three main findings by Smolka et al. outlined above. The ability of the model to do so, even though it operates completely at the word level without any notion of morphemes, may put into question Smolka

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

Target	*binden (bind)*
1 Transparent Derivation	*zubinden (tie)*
2 Opaque Derivation	*entbinden (give birth)*
3 Synonym	*zuschnüren (tie)*
4 Form	*abbilden (depict)*
5 Unrelated	*abholzen (log)*

Figure 1: Smolka et al. (2014)'s five prime types

et al.'s call for novel morpheme-level mechanisms for German.

2 Modeling Priming

We model the priming effects shown in Smolka et al. by combining two computational information sources: A distributional semantic model, and a derivational lexicon.

Distributional Semantics and Priming. Distributional semantics builds on the distributional hypothesis (Harris, 1968), according to which the similarity of lemmas correlates with the similarity of their linguistic contexts. The meaning of a lemma is typically represented as a vector of its contexts in large text collections (Turney and Pantel, 2010; Erk, 2012), and semantic similarity is operationalized by using a vector similarity measure such as cosine similarity. Traditional models construct vectors directly from context co-occurrences, while more recent models learn distributed representations with neural networks (Mikolov et al., 2013), which can be seen as advanced forms of dimensionality reduction.

A classical test case of distributional models is exactly lexical priming, which has been modeled successfully in a number of studies (McDonald and Lowe, 1998; Lowe and McDonald, 2000). The assumption of this model family, which we call DISTSIM, is that the cosine similarity between a prime vector $\vec{p}$ and a target vector $\vec{t}$ is a direct predictor of lexical priming:

$$priming_{\text{DISTSIM}}(p,t) \;\propto\; cos\left(\vec{p},\vec{t}\right)$$

Regarding morphological priming, this model predicts the result patterns for French and Dutch but should not be able to explain the German results.

Derivational Morphology in a Distributional Model. In Padó et al. (2013), we proposed to extend distributional models with morphological knowledge in the form of *derivational families* $\mathcal{D}$, that is, sets of lemmas that are derivationally (either transparently or opaquely) related (Daille et al., 2002), such as:

> *knien$_V$ (to kneel$_V$), beknien$_V$ (to beg$_V$), Kniende$_N$ (kneeling person$_N$)*, kniend$_A$ (kneeling$_A$), Knie$_N$ (knee$_N$)

While our motivation was primarily computational (we aimed at improving similarity estimates for infrequent words by taking advantage of the shared meaning within derivational families), these families can be reinterpreted in the current context as driving *morphological generalization* in priming. More specifically, consider the following model family, which we call MORGEN and which is an asymmetrical version of the "Average Similarity" model from Padó et al. (2013):

$$priming_{\text{MORGEN}}(p,t) \;\propto\; \frac{1}{N}\sum_{p' \in \mathcal{D}(p)} cos\left(\vec{p'},\vec{t}\right)$$

This model predicts priming as the average similarity between the target t and *all* lemmas p' within the derivational family of the prime p. It operationalizes the intuition that the prime "activates" its complete derivational family, no matter if transparently or opaquely related. Each of the family members then contributes to the priming effect just like in standard lexical priming.

The MORGEN model should have a better chance of modeling Smolka et al.'s results than the DISTSIM model. Note, however, that it remains completely at the word level, with derivational families as its only source of morphological knowledge.

3 Experiment

Setup. We compute a DISTSIM model by running `word2vec` (Mikolov et al., 2013), a system to extract distributional vectors from text, with its default parameters, on the lemmatized 800M-token German web corpus SdeWaC (Faaß and Eckart, 2013). To build MORGEN, we use the derivational families from DERIVBASE v1.4, a semi-automatically induced large-coverage German lexicon of derivational families (Zeller et al., 2013).[1]

[1] DERIVBASE defines derivational families through a set of about 270 surface form transformation rules. MORGEN does not use information about rules, only family membership. Nevertheless, it is a question for future research to assess the potential criticism that the rule-based induction method implicitly introduces morpheme-level information into the families.

Prime Type	Exp. 1 (RT in ms) (Smolka et al.)	DISTSIM (cos sim) (our model)	MORGEN (cos sim) (our model)
1 Transparent Derivation	563**	**0.44*** **	**0.35*** **
2 Opaque Derivation	566**	0.28	**0.35*** **
3 Synonym	580	0.41***	**0.30* **
4 Form	600	**0.24**	**0.26**
5 Unrelated	591	0.25	0.27

Table 1: Average Reaction Times and cosines, respectively. Significance results compared to level *Unrelated*. Correct contrasts shown in boldface. Legend: *: p<0.05; **: p<0.01; ***: p<0.001

Following Smolka et al., we analyze the predictions with a series of one-way ANOVAs (factor Prime Type with reference level Unrelated). As appropriate for multiple comparisons, we adopt a more conservative significance level (p=0.01).

Results. Table 1 reports the experimental results and model predictions (average experimental reaction times, cosine model predictions, and significance of differences). Model contrasts that match experiment contrasts are marked in bold.

As expected, DISTSIM predicts the patterns of classical lexical priming: we observe significant priming effects for Transparent Derivation and Synonymy, and no priming for Opaque Derivation. This is contrary to Smolka et al.'s experimental results.

Our instance of the MORGEN model does a much better job: It predicts highly significant priming effects for both Transparent and Opaque derivations (p<0.001) while priming is not significant at p<0.01 for Synonyms (p=0.04). These predictions correspond very well to Smolka et al.'s findings (cf. Table 1). We tested for two additional contrasts analyzed by Smolka et al.: the difference in priming strength between Transparent and Opaque Derivation (not significant in either experiment or model) and the difference between Transparent Derivation and Synonym (highly significant in both experiment and model).

4 Discussion

In sum, we find a very good match between MORGEN and the experimental results, while the DISTSIM model cannot account for the experimental evidence. Recall that the main difference between the two models is that MORGEN's includes all members of the prime's derivational family into the prediction of the priming strength. This leads to the following changes compared to DISTDIM:

1. For Opaque Derivation, MORGEN typically predicts stronger priming than DISTSIM, since prime and target are typically members of the same derivational family (assuming that there are no coverage gaps in DERIVBASE), and the average similarity between the target and the words in the family is higher than the similarity to the prime itself. Taking Figure 1 as an example, the Opaque Derivation pair *entbinden (give birth) – binden (bind)* is relatively dissimilar, and the similarity increases when other pairs like *binden (bind) – zubinden (tie)* are taken into consideration.

2. For Synonymy, MORGEN typically predicts weaker priming than DISTSIM, since the average similarity between target and all members of the prime's family tends to be lower than the similarity between target and original prime. Again considering Figure 1, the Synonym pair *binden (bind) – zuschnüren (tie)* is relatively similar, while including terms derivationally related to the prime *zuschnüren (tie)* like *schnurlos (cordless)* introduces low-similarity pairs like *schnurlos (cordless) – binden (bind)*.

MORGEN is not the only model that takes a distributional stance towards morphological derivation. Marelli and Baroni (2014) propose a compositional model that computes separate distributional representations for the meanings of stems and affixes and is able to compute representations for novel, unseen derived terms. The morpheme-level approach of Marelli and Baroni's model corresponds more directly to Smolka et al.'s claims and might also be able to account for the experimental patterns.

However, our considerably simpler model, which only has knowledge about distributional families, is also able to do so. This at the very least means that morpheme-level processing is not an *indispensable* property of any model that explains Smolka et al.'s experimental results and that the evidence for a special organization of the German mental lexicon, in contrast to other languages, must be examined more carefully.

In fact, our model provides a possible alternative source of explanations for the cross-lingual differences: Since the MorGen predictions are directly influenced by the size and members of the derivational families, German opaque morphological priming may simply result from the high frequency of opaque derivations. In the future, we plan to apply the model to Dutch and French to check this alternative explanation.

Acknowledgments

We gratefully acknowledge funding by Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 732, project B9.

References

J. Kathryn Bock. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18:355–387.

Béatrice Daille, Cécile Fabre, and Pascale Sébillot. 2002. Applications of computational morphology. In Paul Boucher, editor, *Many morphologies*, pages 210–234.

Katrin Erk. 2012. Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, 6(10):635–653.

Gertrud Faaß and Kerstin Eckart. 2013. SdeWaC – a corpus of parsable sentences from the web. In Iryna Gurevych, Chris Biemann, and Torsten Zesch, editors, *Language Processing and Knowledge in the Web*, volume 8105 of *Lecture Notes in Computer Science*, pages 61–68. Springer Berlin Heidelberg.

Zellig Harris. 1968. *Mathematical Structures of Language*. Wiley.

Steve T. Kempley and John Morton. 1982. The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, pages 441–445.

Will Lowe and Scott McDonald. 2000. The direct route: Mediated priming in semantic space. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society*, pages 675–680, Philadelphia, PA.

Marco Marelli and Marco Baroni. 2014. Dissecting semantic transparency effects in derived word processing: A new perspective from distributional semantics. In *9th International Conference on the Mental Lexicon*, pages 7–8, Niagara on the Lake, Canada. Abstract.

Scott McDonald and Will Lowe. 1998. Modelling functional priming and the associative boost. In *Proceedings of the 20th Annual Conference of the Cognitive Science Society*, pages 675–680, Madison, WI.

Fanny Meunier and Catherine-Marie Longtin. 2007. Morphological decomposition and semantic integration in word processing. *Journal of Memory and Language*, 56:457–471.

David E. Meyer and Roger W. Schvaneveldt. 1971. Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2):227–234.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119.

Sebastian Padó, Jan Šnajder, and Britta Zeller. 2013. Derivational smoothing for syntactic distributional semantics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 731–735, Sofia, Bulgaria.

Herbert Schriefers, Pienie Zwitserlood, and Ardi Roelofs. 1991. The identification of morphologically complex spoken words: Continuous processing or decomposition? *Journal of Memory and Language*, 30:26–47.

Eva Smolka, Katrin H. Preller, and Carsten Eulitz. 2014. 'verstehen' ('understand') primes 'stehen' ('stand'): Morphological structure overrides semantic compositionality in the lexical representation of German complex verbs. *Journal of Memory and Language*, 72:16–36.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37(1):141–188.

Britta Zeller, Jan Šnajder, and Sebastian Padó. 2013. DErivBase: Inducing and evaluating a derivational morphology resource for German. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, pages 1201–1211, Sofia, Bulgaria.

What can distributional semantic models tell us about part-of relations?

François Morlane-Hondère
LIMSI-CNRS, Orsay, France
francois.morlane-hondere@limsi.fr

1 Introduction

The term *Distributional semantic models* (DSMs) refers to a family of unsupervised corpus-based approaches to semantic similarity computation. These models rely on the distributional hypothesis (Harris, 1954), which states that semantically related words tend to share many of their contexts. So, by collecting information about the contexts in which words are used in a corpus, DSMs are able to measure the distributional similarity of two words, which theoretically translates into a semantic one.

In recent years, these models have become very popular in a wide range of NLP tasks (Weeds, 2003; Baroni and Lenci, 2010), mainly because of the ever-increasing availability of textual data. Regardless of their use in NLP applications, distributional data provide precious information about words' *behaviour* and their tendency to appear in the same contexts. Yet, linguists have shown little interest in DSMs (Sahlgren, 2008). We believe that this kind of information can be relied on to empirically assess the validity of linguistic theories. Conversely, by shedding light on underlying linguistic factors that influence distributional behaviours, linguistic studies can contribute to improve our understanding of the results provided by DSMs.

This paper illustrates such a qualitative linguistic approach by investigating the presence of part-of relations among distributionally similar French words. We compare distributional data and a set of part-of relations provided by humans in a lexical network. In order to assess the nature of the part-of word pairs which can – or cannot – be found in DSMs, these words were sense-tagged using WordNet supersenses. Our results show considerable discrepancies between the representation of part-of *sense pairs* in distributional data.

2 Part-of relation and DSMs

As its name suggests, part-of relation – or *meronymy*[1] – holds between a part – the meronym – and its *whole* – the *holonym* –, like in *bed/pillow*, *armor/steel* or *ostrich/feather*. It is one of the central relations used in knowledge representation.

Automatic extraction of part-of relations has been addressed using many approaches, most of which are pattern-based (Berland and Charniak, 1999; Girju et al., 2006; Pantel and Pennacchiotti, 2006). However, the unsupervised nature of the distributional approach makes it an attractive alternative.

Studies were conducted to assess the nature of the semantic relations extracted by distributional models – using human judges (Kuroda et al., 2010), thesauri (Morlane-Hondère, 2013; Ferret, 2015) or *ad hoc* datasets (Baroni and Lenci, 2011). They showed that part-of relations are present in varying proportions among distributionally similar words. This very presence is interesting in that unlike synonymy, hypernymy or co-hyponymy, meronymy is not a similarity relation (Resnik, 1993; Budanitsky and Hirst, 2006): an ostrich is not *the same kind of thing* as a feather, neither an armor is the same *kind of thing* as steel. Following the distributional hypothesis, it is not expected that these kind of meronyms share a lot of contexts.

It appears, though, that a certain proportion of them tend to do so. For example, in Baroni and Lenci (2010)'s DSM, *player*, *pianist* and *musician* are among the ten most distributionally similar words of *orchestra*. In the following of this study, we compare the semantic properties of the meronyms which can be extracted using a distributional approach and the properties of the meronyms which cannot.

[1] Some authors make a distinction between part-of relation and meronymy (Cruse and Croft, 2004).

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

3 Methodology and data

3.1 The part-of dataset

The first step consists in gathering a set of meronyms. Although efforts are made to provide expert-built lexical semantic resources for French (Fišer and Sagot, 2008; Pradet et al., 2014), there is currently no freely-available equivalent – in terms of quality and coverage – to WordNet (Fellbaum, 1998) or the Moby thesaurus (Ward, 2002) for French. So, we use the JeuxDeMots (JDM) lexical network (Lafourcade, 2007), which is a GWAP (*Game With A Purpose*) in which players are asked to provide words which can be in a given relation with a given word[2].

Although collaboratively-built lexical semantic resources have shown to be valuable (Gurevych and Wolf, 2010) and although a relation in JDM must be provided by two different players to be added to the network, a certain proportion of part-of relations in JDM are actually hypernymys (*sucette/bonbon* 'lollipop/candy'), synonyms (*chef/patron* 'chief/boss') or thematic associations (*océanographie/eau* 'oceanography/water'). Two possible explanations for these confusions are the lack of linguistic expertise of the players or a misunderstanding of the instruction. Erroneous relations were manually removed from the set.

One interesting characteristic of JDM part-of relations is that a considerable number of them do not fit into traditional typologies of meronymy relations. For example, topological inclusions (*cell/prisoner*), attachment relations (*ear/earring*) or ownership (*millionaire/money*) are very common among JDM part-of pairs although they are considered to be non-meronymic relations (Winston et al., 1987).

After filtering the pairs whose members do not appear in our DSM and removing most of the erroneous relations, there were 24 089 part-of pairs left in our dataset.

3.2 Sense tagging

In a previous study (Morlane-Hondère and Fabre, 2012), we manually annotated the different meronymic sub-relations – following Winston and Chaffin (1987)'s typology – in a dataset like the one described above. The idea was to test whether there is a correlation between the nature of the relation between two words and their probability of being extracted in a DSM. However, the typology has proven to be inadequate, so we chose to annotate the words instead of their relation. This is also what we do in this study. This approach is inspired by the idea that the difference between the meronymic sub-relations is due to the semantic nature of the words involved (Murphy, 2003).

The above-mentioned lack of freely-available thesauri for French led us to use WordNet to perform this task. Words of our dataset were 1) translated to English, 2) mapped to WordNet synsets and 3) linked to their translation's *supersense(s)*. Supersenses – or *lexicographer classes* – are a set of 44 coarse semantic categories used to classify WordNet's noun, verb and adjective entries[3]. Examples of the 25 noun supersenses are GROUP, LOCATION or FOOD. Supersenses were then manually disambiguated (*drawer* can both belong to the PERSON and ARTIFACT supersenses, but only the latter fits in the pair *cabinet/drawer*).

3.3 The distributional model

We use a DSM[4] generated from the frWaC corpus (Baroni et al., 2009) – a 1.6 billion words corpus of French web pages.

Words in the DSM appear at least 20 times in the corpus and in at least 5 different contexts.

Syntactic dependencies were used as contexts using the Talismane parser (Urieli, 2013). Relations taken into account in the context vectors are the subject, object and modifier relations. Prepositions and coordinating conjunctions are also included as relations (the label of the relation being the preposition or the coordinating conjuction).

The weighting of the contexts was made using the pointwise mutual information and the cosine measure was used to compute the similarity between the context vectors. The minimum similarity threshold has been set to 0.02. The total number of word pairs in the DSM is 3 674 254.

4 Results and discussion

We then measure the proportion of semantically-annotated part-of pairs – *sense pairs* – in our set which are present in the DSM. Sense pairs which occur less than 100 times in the dataset are discarded. Table 1 provides the list of the 22 re-

[2] `http://www.jeuxdemots.org/`

[3] `http://wordnet.princeton.edu/man/lexnames.5WN.html`

[4] Provided by Franck Sajous from the CLLE-ERSS laboratory.

maining sense pairs and, for each one, the ratio of part-of pairs present in the DSM. In this section, we describe the *homogeneous* sense pairs – whose semantic classes are identical – and the *heterogeneous* ones, then we provide a detailed analysis of some of the PERSON/BODY meronyms which have been extracted by the DSM.

4.1 Homogeneous sense pairs

As expected, part-of relations composed of two words of the same class are the most represented in the DSM. 84 % of the TIME/TIME part-of pairs were extracted by the DSM. This can be explained by the fact that the members of pairs like *mois/jour* 'month/day' both appear in contexts involving temporal prepositions like *venir*_IL Y A 'to come_SINCE', *se dérouler*_DURANT 'to take place_DURING' or *scrutin*_AVANT 'election_BEFORE'.

Likewise, the spatial dimension plays a crucial role in the extraction of meronyms (78.3 % of LOCATION/LOCATION pairs are extracted). This is due to the fact that, as for time, spatial information can be conveyed by specific prepositions. Thus, LOCATION/LOCATION meronyms' shared contexts massively involve the DANS 'IN' relation.

SUBSTANCE pairs are the third best-extracted kind of pairs. The reason why 37.6 % of them has not been extracted can be illustrated by the comparison of *acier* 'steel' and two of its meronyms, namely *fer* 'iron' – which was extracted in the DSM – and *carbone* 'carbon' – which was not extracted:

1. *acier* and *fer* both appear in contexts like *grille*_EN 'grille_COMP', *forgé*_MOD 'forged_MOD' or *lame*_DE 'blade_COMP'. Thus, they appear as materials and, moreover, as materials which are used to build the same kind of things;

2. although being a material as well, *carbone* does not appear as such in the corpus. Rather, its contexts are chemical compounds like *monoxyde*_DE 'monoxide_COMP'. It is also modified by adjectives like *inorganique*_MOD 'inorganic_MOD', which describe chemical properties of *carbone*. These two kinds of contexts are not found among *acier*'s.

So, we can see that there is a discrepancy between the contexts in which *acier* appears in the corpus and the ones in which *carbone* appears: whereas *acier* – as well as *fer* – is used as a material, the representation of *carbone* that emerges from the corpus is that of a chemical element.

holonym/meronym	%	holonym/meronym	%
TIME/TIME	84	ARTIFACT/PERSON	32.6
LOC./LOC.	78.3	ARTIFACT/ARTIFACT	31.4
SUBST./SUBST.	62.4	ARTIFACT/LOC.	24.8
OBJECT/OBJECT	61	ARTIFACT/PLANT	22.8
COMM./COMM.	53.8	ARTIFACT/SUBST.	20.4
GROUP/PERSON	52.8	OBJECT/ANIMAL	19.8
LOC./ARTIFACT	46.8	PLANT/PLANT	19.7
BODY/BODY	40.5	GROUP/ANIMAL	17.1
ANIMAL/ANIMAL	41	PERSON/ARTIFACT	16.5
ARTIFACT/COMM.	39.9	ANIMAL/BODY	9.4
ACT/ARTIFACT	35.8	PERSON/BODY	5.5

Table 1: Part-of sense pairs and their presence in the DSM.

4.2 Heterogeneous sense pairs

At the other end of the scale, part-of relations composed of two words of different classes are – also logically – the less represented in the DSM.

Part-of pairs composed of words that refer to human beings or to animals and their body parts are barely present in the DSM (although being the most frequent sense pairs in our dataset). In frWaC, PERSON words appear as subjects of action (*prendre* 'to take', *dire* 'to say') or cognitive verbs (*vouloir* 'to want', *savoir* 'to know'). They are frequently modified by nationality adjectives. Body parts do not appear in such contexts. The class of body parts was actually found to be quite heterogeneous, in that body parts' distributions in the corpus differ from persons', but not in the same way:

- organ nouns mostly appear in noun compounds to indicate the location of medical interventions (*radiographie*_DE 'x-ray_MOD') or affections (*cancer_de* 'cancer_COMP' or *lésion_de* 'injury_COMP');

- limb nouns are modified by adjectives related to location and are objects of verbs like lever 'to raise' or *étendre* 'to stretch'.

All these contexts are obviously incompatible with PERSON words.

A similar distributional discrepancy can be observed with the ANIMAL/BODY sense pair, except that animal nouns tend to appear in contexts like *élevage*_DE 'farming_COMP' or *espèce*_DE 'species_COMP'. They are also modified by size adjectives. It is interesting to note that many animal body parts like *tête*_DE 'head_COMP', *peau*_DE 'skin_COMP' or *queue*_DE 'tail_COMP' do appear among the closest contexts of animal nouns. This means that the meronymic relation between nouns referring to animals and their body parts is not a paradigmatic one. Thus, it is reasonable to say that, in order to extract this particular relation, the use of syntagmatic patterns would be a better strategy than the use of a paradigmatic DSM.

The sense pair GROUP/PERSON also presents an interesting situation. Of all the heterogeneous sense pairs, meronymic relations belonging to this one are the most likely to be extracted by the distributional method. This can be explained by a tendency to use the GROUP entities in a metonymic way: although an army is not *the same kind of thing* as a soldier, both words share contexts like *tirer*_SUJ 'to shoot_SUBJ' or *tué*_PAR 'killed_BY'. Another reason is the transitivity of properties like nationality: *armée* 'army' and *soldat* 'soldier' are both modified by nationality adjectives because usually, members of the armed forces of a nation have to be citizens of this nation.

In the section 2, we mentioned the fact that three meronyms of *orchestra* were present among its ten most distributionally similar words in Baroni and Lenci (2010)'s DSM. In our data, the meronyms *orchestre/musicien* have also been extracted: as for *army* and *soldier*, these words share semantic features. They are related to the kind of music a musician and an orchestra can play (*classique*_MOD 'classical_MOD', *traditionnel*_MOD 'traditional_MOD' or *jazz*_DE 'jazz_MOD'), the kind of actions they perform (*interprété*_PAR 'performed_BY', *accompagné*_PAR 'accompanied_BY') or their nationality.

4.3 Focus on the PERSON/BODY sense pair

In the previous subsection, we saw that meronyms belonging to the PERSON/BODY are the least likely to be extracted with the distributional approach. In this subsection, we provide further insight into this result by examining the nature of the few PERSON/BODY meronymic pairs that were successfully extracted.

The examination of the 5.5 % of PERSON/BODY meronymic pairs that were successfully extracted is disappointing: the vast majority of the contexts shared by the meronym and the holonym are quite random. For example, the meronyms *homme/main* 'man/hand' share contexts like *nu*_MOD 'bare_MOD' or *dos*_DE 'back_COMP', which are not very informative about their relation. On the other hand (!) some shared contexts like *doigt*_DE 'finger_COMP' and *saisir*_SUJ 'to grab_SUBJ' are more informative. The fact that these specific features are shared by the meronyms indicates some kind of similarity between them: when a man grabs a rock, it is actually his hand that completes the action of grabbing, as well as a man's fingers are also his hand's fingers.

The meronyms *enfant/oeil* 'child/eye' also share some interesting contexts: both the meronym and the holonym are subjects of verbs of visual perception like *regarder* 'to look', *percevoir* 'to perceive' or *observer* 'to observe'. The metonymic interpretation is quite straightforward: although the eye is the child's *part* that allows him to look/perceive/observe, this ability is extended to the whole child.

This phenomenon partially explains why such meronyms share semantic – thus distributional – features and are more likely to be extracted with a DSM.

5 Conclusion

The main goal of this study is to shed light on the linguistic phenomena at work in DSMs. By comparing a set of sense-tagged part-of relations and a distributional model, we show that the semantic class of the meronyms has a dramatic influence on their probability to be extracted by a DSM. We also highlight the – positive – influence of metonymy in the extraction of heterogeneous meronyms.

These results show that the part-of relation is not a monolithic entity but a collection of different kinds of relations between different kinds of words which may or may not be distributionally similar.

Acknowledgments

This work was partially supported by the ANSM (French National Agency for Medicines and Health Products Safety) through the Vigi4MED project under grant #2013S060.

References

Marco Baroni, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. The WaCky wide web: a collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3):209–226, 2009.

Marco Baroni and Alessandro Lenci. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721, 2010.

Marco Baroni and Alessandro Lenci. How we BLESSed distributional semantic evaluation. *GEMS 2011*, pages 1–10, 2011.

Matthew Berland and Eugene Charniak. Finding parts in very large corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, ACL '99, pages 57–64, Stroudsburg, PA, USA, 1999. Association for Computational Linguistics.

Alexander Budanitsky and Graeme Hirst. Evaluating WordNet-based Measures of Lexical Semantic Relatedness. *Computational Linguistics*, 32(1):13–47, March 2006.

D. Alan Cruse and William Croft. *Cognitive linguistics*. Cambridge: Cambridge University Press, 2004.

Christiane Fellbaum, editor. *WordNet An Electronic Lexical Database*. The MIT Press, Cambridge, MA; London, May 1998.

Olivier Ferret. Typing relations in distributional thesauri. In Núria Gala, Reinhard Rapp, and Gemma Bel-Enguix, editors, *Language Production, Cognition, and the Lexicon*, volume 48 of *Text, Speech and Language Technology*, pages 113–134. Springer International Publishing, 2015.

Darja Fišer and Benoît Sagot. Combining multiple resources to build reliable wordnets. In *TSD 2008 - Text Speech and Dialogue*, Brno, Czech Republic, 2008.

Roxana Girju, Adriana Badulescu, and Dan Moldovan. Automatic discovery of part-whole relations. *Comput. Linguist.*, 32(1):83–135, March 2006.

Iryna Gurevych and Elisabeth Wolf. Expert-Built and Collaboratively Constructed Lexical Semantic Resources. *Language and Linguistics Compass*, 11(4):1074–1090, 2010.

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

Kow Kuroda, Jun'ichi Kazama, and Kentaro Torisawa. A look inside the distributionally similar terms. In *Proceedings of the Second Workshop on NLP Challenges in the Information Explosion Era (NLPIX 2010)*, pages 40–49, Beijing, China, August 2010. Coling 2010 Organizing Committee.

Mathieu Lafourcade. Making people play for Lexical Acquisition with the JeuxDeMots prototype. In *SNLP'07: 7th International Symposium on Natural Language Processing*, page 7, Pattaya, Chonburi, Thailand, December 2007.

François Morlane-Hondère. *Une approche linguistique de l'évaluation des ressources extraites par analyse distributionnelle automatique*. PhD thesis, Université de Toulouse II le Mirail, 2013.

François Morlane-Hondère and Cécile Fabre. Étude des manifestations de la relation de méronymie dans une ressource distributionnelle. In *Proceedings of TALN 2012*, Grenoble, France, June 2012.

Lynne Murphy. *Semantic Relations and the Lexicon*. Cambridge University Press, New York, 2003.

Patrick Pantel and Marco Pennacchiotti. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, ACL-44, pages 113–120, Stroudsburg, PA, USA, 2006. Association for Computational Linguistics.

Quentin Pradet, Gaël de Chalendar and Jeanne Baguenier Desormeaux. WoNeF, an improved, expanded and evaluated automatic French translation of WordNet. In *GWC 2014*, Tartu, Estonia, 2014.

Philip Resnik. *Selection and Information: a Class-Based Approach to Lexical Relationships*. PhD thesis, The Institute For Research In Cognitive Science, University of Pennsylvania, 1993.

Magnus Sahlgren. The distributional hypothesis. *Rivista di Linguistica*, 20(1):33–53, 2008.

Assaf Urieli. *Robust French syntax analysis: reconciling statistical methods and linguistic knowledge in the Talismane toolkit*. PhD thesis, Université de Toulouse II le Mirail, 2013.

Grady Ward. *Moby Thesaurus List (English),*. 2002.

Julie Weeds. *Measures and Applications of Lexical Distributional Similarity*. PhD thesis, Department of Informatics, University of Sussex, 2003.

M. E. Winston, R. Chaffin, and D. Herrmann. A taxonomy of part-whole relations. *Cognitive Science*, 11(4):417–444, December 1987.

Modeling Lexical Effects in Language Production: Where Have We Gone Wrong?

Ting Zhao
Department of Education
University of Oxford
ting.zhao@education.ox.ac.uk

Victoria A. Murphy
Department of Education
University of Oxford
victoria.murphy@education.ox.ac.uk

1 Introduction

Words have their own conceptual representations, semantic properties, and physical forms. These lexical characteristics not only set words apart as a distinct item in the lexical repertoire but also provide valuable insight into the processes and mechanisms of language production.

Over the past decades there has been a large body of research examining how word meaning, form, and usage directly affect the speed of monolingual speakers' production (e.g. Alario et al., 2004; Barry, Morrison, & Ellis, 1997; Bates et al., 2003; Bonin, Chalard, Méot, & Fayol, 2002). Of note, almost all these studies have failed to accommodate the fact that word usage, given it is a behavioral outcome (Zevin & Seidenberg, 2002, 2004), likely mediates the relationship between meaning/form and spoken production. Moreover, lexical characteristics have been predominantly examined as discrete variables in the literature, but in fact, some of them may correspond to the same layer of language production or the same aspect of lexical knowledge. Additionally, little work has been done on children's emerging bilingual lexical representations, especially those learning an L2 within input-limited contexts, possibly due to the fact that this population has only recently begun to receive focused attention in the research field.

In order to delineate the exact manner in which lexical effects come into play, the present study used structural equation modeling to perform a simultaneous test of the complex relationships among a variety of lexical variables and to assess their direct, indirect and total effects on L2 lexical processing efficiency. Furthermore, attempts were also made to estimate and then to compare three types of hypothesized models, in which the lexical relationships were specified differently with respect to spoken production in L2 learners.

2 Lexical characteristics that contribute to the speed of spoken production

This study considers three lexical layers (i.e. Meaning, Form, and Usage), each of which is underpinned by its own manifest indicators. The lexical variables under examination all have been found to significantly influence the speed of lexical processing, as will be briefly reviewed below.

Meaning. (1) Word concreteness (WC): A main difference between concrete and abstract words lies in the existence of sensorimotor attributes of the former. A number of studies have revealed that concrete words exhibit preferential processing relative to abstract words (e.g. De Groot, 1992; Jin, 1990; Schwanenflugel & Akin, 1994). (2) Word typicality (WT): The degree of a lexical item's typicality depends upon how many attributes that it shares with other members of the same category. Typical items are usually processed more accurately and faster relative to atypical items in a range of tasks (e.g. Bjorklund & Thompson, 1983; Jerger & Damian, 2005; Southgate & Meints, 2000). (3) Semantic neighborhood density (SND): Words with high SND are characterized by having a great deal of semantic neighbors and low semantic distance, whereas low-SND words typically have few semantic neighbors and high semantic distance. The superiority of high SND over low SND words for processing has been observed in lexical decision, word naming, and semantic categorization (e.g. Buchanan, Westbury, & Burgess, 2001; Siakaluk, Buchanan, & Westbury, 2003; Yates, Locker, & Simpson, 2003). (4) Number of related senses (NoRS): Many words are polysemous in terms of having several different but related senses. Compared to monosemous words, polysemous words exhibit preferential processing in a variety of tasks (e.g. Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007; Lichacz, Herdman, Lefevre, & Baird, 1999).

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

Form. Word length can be measured orthographically (i.e. NoL: number of letters) or phonologically (i.e. NoP: number of phonemes and NoS: number of syllables). The presence of length effects has been reported in several previous studies (e.g. Alario et al., 2004; Cuetos, Ellis, & Alvarez, 1999; De Groot, Borgwaldt, Bos, & Van den Eijnden, 2002) although the predictive power of each specific measure varies across research contexts possibly due to their examination of different languages (Bates et al., 2003).

Usage. Usage is represented by subjective word frequency (SWF) and /or age of acquisition (AoA), both of which have been observed to significantly affect the speed of spoken production in such a way that individuals take less effort to access high-frequency and early-acquired words relative to low-frequency and late-acquired ones (e.g. Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Barry et al., 1997; Morrison, Ellis, & Quinlan, 1992). AoA effects interact with frequency effects in such a way that the former is partly dependent on the latter (Brysbaert & Ghyselinck, 2006).

3 Methodology and analytical strategies

3.1 Methodology

Participants. Thirty-nine 5th grade children (aged 10-11 years) and 94 undergraduates (aged 17-20 years) were recruited. All had Chinese as their native language and English as their second. The child sample had been learning English as a foreign language for about 2.5 years, and the adult sample for approximately 10 years.

Stimuli. The experiment consisted of two blocks of stimulus words and one block of filler words. Each block had 35 (in the child group) / 66 (in the adult group) valid trials. The stimuli were selected from ten semantic categories in almost equal numbers. They were all presented in the same format over the course of the experiments.

Procedures. The participants were tested individually in a quiet room. They performed picture naming in L2 (English) and then L1 (Chinese)-to-L2 (English) translation. As a stimulus appeared on the screen, the participants were asked to produce the L2 word as rapidly and accurately as possible. The SuperLab software (Cedrus Corporation, 2007) generated stimulus presentations. Response latencies (RLs), defined as the duration between the presentation of a stimulus and the initiation of a vocal response, were recorded using the Audacity software, and then manually calculated for analysis.

Norms of lexical variables. The values of WC, WT, and SWF were rated by the participants on Likert scales. The values of other lexical variables were obtained from psycholinguistics databases such as the Irvine Phonotactic Online Dictionary (Vaden, Halpin, & Hickok, 2009) and the Wordmine2 (Durda & Buchanan, 2006).

3.2 Analytical strategies

Structural equation modeling (SEM), which combines path analysis, confirmatory factor analysis, and analysis of structural models, was used to estimate the goodness-of-fit of three types of hypothesized models. This analytical strategy, as an extension of multiple regression, enables researchers to estimate not only the direct effects but also indirect effects that one variable has upon another. Moreover, SEM can be used to measure the proportion of variance explained by the models proposed in the present study so as to hold general implications for the lexical processing system as a whole, although it should be acknowledged that this type of analysis might lack a specific focus on certain variables through purposeful manipulation of experimental materials. Additionally, latent variables are formed to manifest different dimensions that are underpinned by their own indicators. In so doing, the present study moves away from the examination of each lexical variable to that of specified constructs and structural relations between constructs, thus a better understanding of the nature of lexical characteristics can be gained at a more macro level.

Conducting SEM typically involves six steps (Kline, 2011): model specification, model identification, select good measures, model estimation, model evaluation and modification, and interpreting and reporting results. Moreover, as recommended by Kline (2011), SEM was conducted in two steps in the present study, that is, the measurement models were validated in terms of convergent validity, discriminant validity, and reliability before the structural models proceeded to be estimated. One last thing to note is that the data entered for analysis were lexical items. The stimulus size in the adult group was considered sufficiently large for performing SEM analysis. In order to reduce the complexity of the hypothesized model specifying children's L2 lexical processing, composite variables rather than latent variables were constructed to decrease the number of stimulus words required for this type of analysis.

Three competing models were hypothesized and estimated to determine which one best fitted the data. The first model concerns only the direct relationship between the lexical variables, and picture naming and translation latencies. The second model identifies word usage as a mediator and examines the indirect effects of meaning and form variables on the recorded RLs. The third model considers both direct and indirect effects of word meaning and form on the outcome variable. To illustrate, an example of these three types of hypothesized models that specify the possible relationships between lexical variables and the speed of adults' picture naming is presented in Appendix A.

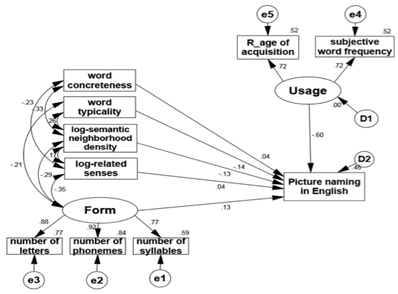
The goodness of model fit was estimated according to six types of indices, including model χ^2, CFI, RMSEA, AGFI, GFI, and NFI. A rule of thumb is that an RMSEA below .08 indicates reasonable fit, and values greater than .90 for the CFI, AGFI, GFI, and NFI suggest close approximate fit. SEM was run using IBM SPSS AMOS v.20.

It should be noted that, before performing SEM analysis, the whole RL data set was screened for incorrect and omitted responses, outliers (low cut-off: below 350ms, high cut-off: 3 SDs), and those participants and stimulus items with an exceptionally high error rate. As conventionally done, RLs were then averaged to generate a summary score for each lexical item, and these values were entered into final SEM analysis.

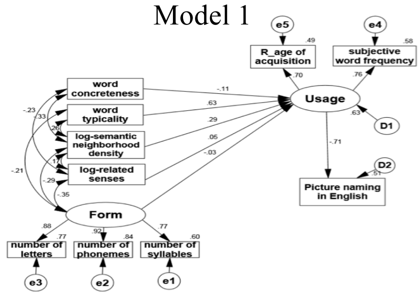
4 Results

The model-fit indices of the three models under examination across two types of productive tasks in both populations are presented in Appendix B. Comparatively, the child and adult data could best be modeled by the third model where word meaning and form not only make direct but also indirect contribution to the RLs.

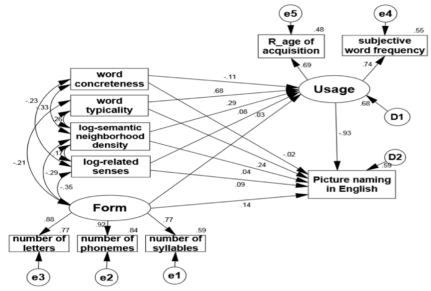
Take picture naming in adults as an example (see SEM results in Appendix B and Figure 1), it is clear that Model 3 achieved a much better model fit than Model 1, and Model 3 explained more variance in naming latencies (59%) than Model 1 (45%) and Model 2 (51%). Additionally, among all the lexical variables included in Model 3, only word usage was found to make a significant and direct contribution to the naming latencies. Similar results held for adults' L1-to-L2 translation (see Appendix C for details).



Model 1



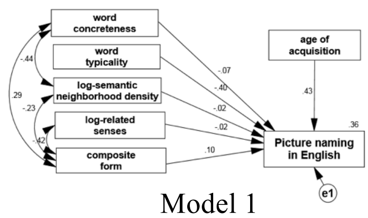
Model 2



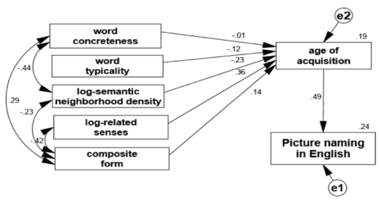
Model 3

Figure 1: SEM results: Picture naming in adults

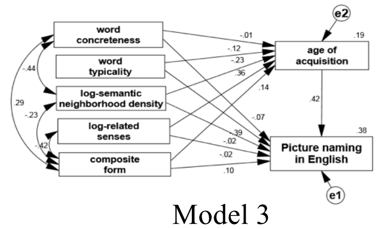
As regards children's picture naming, the results presented in Appendix B shows that Model 3 reached a better model fit than Models 1 and 2. Moreover, Figure 2 indicates that Model 3 (38%) explained more variance in naming speed than Model 1 (36%) and Model 2 (24%). In addition, word usage, as represented by age of acquisition, together with word typicality were found to significantly and directly predict the naming speed in Model 3. Similar results were observed with children's L1-to-L2 translation (see Appendix C for details).



Model 1



Model 2



Model 3

Figure 2: Path analysis results: Picture naming in children

Taken together, these results indicate that word usage does not exist independently of other lexical variables but rather mediates the impact of meaning and form on L2 children's and adults' productive performance. In comparison, the indirect effects of meaning and form on L2 lexical processing efficiency were found to be more noticeable with adults relative to with children.

5 Discussion and conclusions

The present study uses SEM as a methodological improvement to investigate the relationships between a range of lexical variables and L2 lexical processing efficiency in both children and adults. A comparison of the three different types of models indicates that word meaning and form makes not only direct but also indirect contribution to the speed of L2 lexical processing, and word usage likely mediates the extent to which meaning and form influence the processing outcome. Furthermore, a comparison between children and adults suggests that the importance of word usage tends to increase with age.

A note of caution thus should be raised when interpreting the results of previous studies where the mediating effects of word usage have not been adequately addressed. Accordingly, future research modeling lexical effects would be well advised to consider the indirect effect that word meaning and form have on L2 learners' productive performance via usage.

Although this study provides new insights into how lexical variables are related to each other, there are several limitations that should be acknowledged. First, since this research focuses only on L2 learners within input-limited contexts, whether or not the same results still hold for other L2 learner types, particularly those whose L1s are not Sino-Tibetan languages, as well as for monolingual speakers needs to be further investigated. Importantly, examining these issues would allow us to gain a better understanding of the nature of lexical characteristics by addressing the issue of whether lexical effects are language-dependent or universal across languages. Second, not all the variance can be explained the included lexical variables, partly due to the fact that it seems implausible to cover every possible feature of a lexical item because of theoretical and practical considerations. Third, given the use of a non-experimental design, it would be difficult to make unequivocal explanations of causality among the variables of interest.

To conclude, the model that considers both direct and indirect effects of meaning and form on L2 lexical processing efficiency may be superior to those that do not. As also observed in our study, word usage does play a mediating role in lexical processing, in part reflecting that 'only in the stream of thought and life do words have meanings' (Wittgenstein, 1967, p.31).

References

Alario, F-Xavier, Ferrand, L., Laganaro, M., New, B., Frauenfelder, U.H., & Segui, J. (2004). Predictors of picture naming speed. *Behavior Research Methods, 36*(1): 140-155.

Balota, D.A., Cortese, Michael J, Sergent-Marshall, Susan D, Spieler, Daniel H, & Yap, MelvinJ. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133*: 283-316.

Barry, C., Morrison, C.M., & Ellis, A.W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency, and name agreement. *The Quarterly Journal of Experimental Psychology, 50*(3): 560-585.

Bates, Elizabeth, D'Amico, Simona, Jacobsen, Thomas, Székely, Anna, Andonova, Elena, Devescovi, Antonella, . . . Pléh, Csaba. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review, 10*(2): 344-380.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research, 24*(1): 57-65.

Bjorklund, David F, & Thompson, B.E. (1983). Category typicality effects in children's memory performance: Qualitative and quantitative differences in the processing of category information. *Journal of Experimental Child Psychology, 35*(2): 329-344.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology, 93*(1): 89-114.

Brysbaert, Marc, & Ghyselinck, Mandy. (2006). The effect of age of acquisition: Partly frequency related, partly frequency independent. *Visual Cognition, 13*(7-8): 992-1011.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8 (3)*: 531-544.

Cedrus Corporation. (2007). SuperLab 4.5. San Pedro, CA.

Cuetos, Fernando, Ellis, A.W., & Alvarez, B. (1999). Naming times for the Snodgrass and Vanderwart pictures in Spanish. *Behavior Research Methods, 31*(4): 650-658.

De Groot, A.M.B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5): 1001-1018.

De Groot, A.M.B., Borgwaldt, Susanne, Bos, Mieke, & Van den Eijnden, Ellen. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language, 47*(1): 91-124.

Durda, K., & Buchanan, L. . (2006). WordMine2. Retrieved December 1, 2012. from http://www.wordmine2.org

Jerger, S., & Damian, M.F. (2005). What's in a name? Typicality and relatedness effects in children. *Journal of Experimental Child Psychology, 92*(1): 46-75.

Jin, Y.S. (1990). Effects of concreteness on cross-language priming in lexical decisions. *Perceptual and Motor Skills, 70*(3): 1139-1154.

Klepousniotou, E., & Baum, S.R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics, 20*(1): 1-24.

Kline, Rex B. (2011). *Principles and Practice of Structural Equation Modeling* (3rd ed.). New York: Guilford Press.

Lichacz, F.M., Herdman, C.M., Lefevre, J.A., & Baird, B. (1999). Polysemy effects in word naming. *Canadian Journal of Experimental Psychology, 53*(2): 189-193.

Morrison, C.M., Ellis, A.W., & Quinlan, P.T. (1992). Age of acquisition, not word frequency, affects object naming, not object recognition. *Memory & Cognition, 20*(6): 705-714.

Schwanenflugel, P.J., & Akin, Carolyn E. (1994). Developmental trends in lexical decisions for abstract and concrete words. *Reading Research Quarterly, 29*(3): 251-264.

Siakaluk, P.D., Buchanan, L., & Westbury, C. (2003). The effect of semantic distance in yes/no and go/no-go semantic categorization tasks. *Memory & Cognition, 31*(1): 100-113.

Southgate, V., & Meints, K. (2000). Typicality, naming, and category membership in young children. *Cognitive Linguistics, 11*(1/2): 5-16.

Vaden, KI, Halpin, HR, & Hickok, GS. (2009). Irvine Phonotactic Online Dictionary, Version 2.0. Retrieved January 30, 2013 from http://www.iphod.com/search/V2ListWords.html.

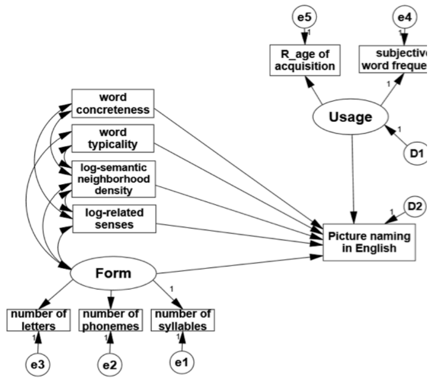
Yates, Mark, Locker, Lawrence, & Simpson, Greg B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Memory & Cognition, 31*(6): 856-866.

Zevin, J.D., & Seidenberg, M.S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language, 47*(1): 1-29.

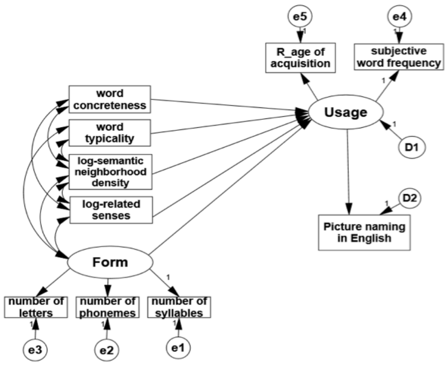
Zevin, J.D., & Seidenberg, M.S. (2004). Age-of-acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory & Cognition, 32*(1): 31-38.

Appendix A. An example of the hypothesized models

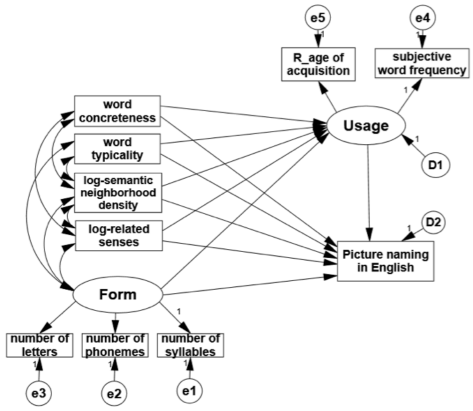
Adult picture naming



Model 1



Model 2



Model 3

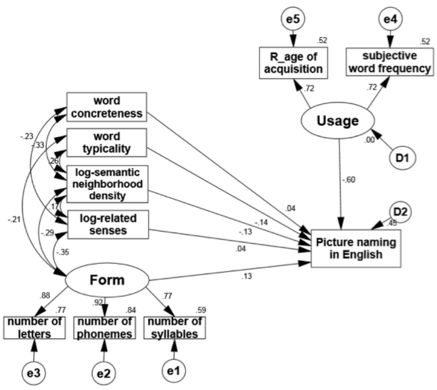
Appendix B. Fit indices for the hypothesized models

			χ^2 (*p*)	*df*	*CFI*	*RMSEA*	*AGFI*	*GFI*	*NFI*
Adults	Picture naming	Model 1	173.83 (.00)	27	.81	.17	.74	.87	.79
		Model 2	52.16 (.00)	27	.97	.07	.90	.95	.94
		Model 3	45.96 (.00)	22	.97	.08	.90	.96	.94
	Chinese-English translation	Model 1	169.46 (.00)	27	.81	.16	.75	.88	.79
		Model 2	45.69 (.01)	27	.98	.06	.91	.96	.94
		Model 3	41.87 (.01)	22	.97	.07	.90	.96	.95
Children	Picture naming	Model 1	28.17 (.00)	11	.85	.27	.82	.93	.79
		Model 2	28.52 (.00)	11	.88	.12	.83	.94	.79
		Model 3	5.67 (.46)	6	1.00	.00	.93	.99	.96
	Chinese-English translation	Model 1	28.17 (.00)	11	.81	.12	.82	.93	.74
		Model 2	23.09 (.02)	11	.86	.10	.87	.95	.79
		Model 3	5.67 (.46)	6	1.00	.00	.93	.99	.95

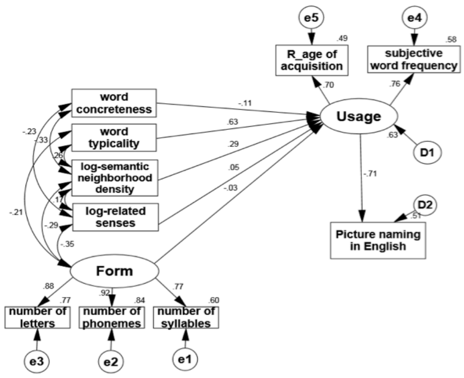
Appendix C. SEM results of the hypothesized models

Adults:

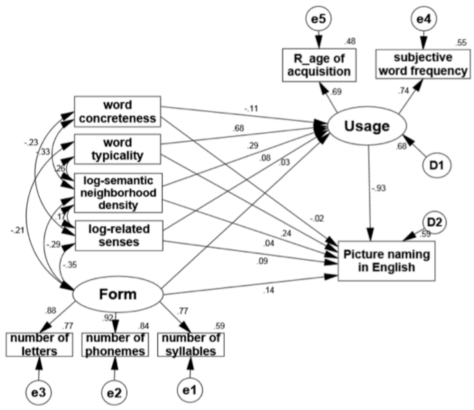
Picture naming



Model 1

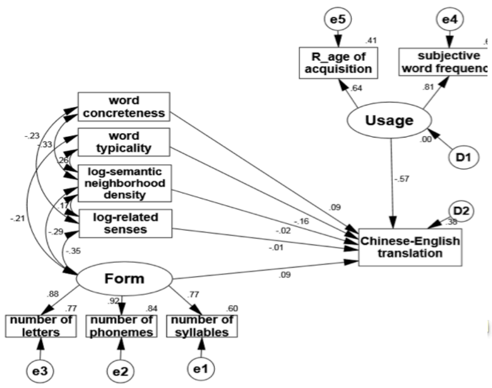


Model 2

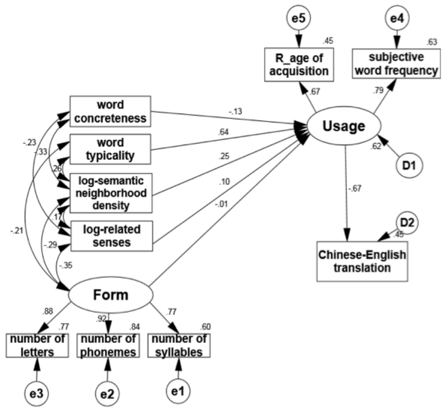


Model 3

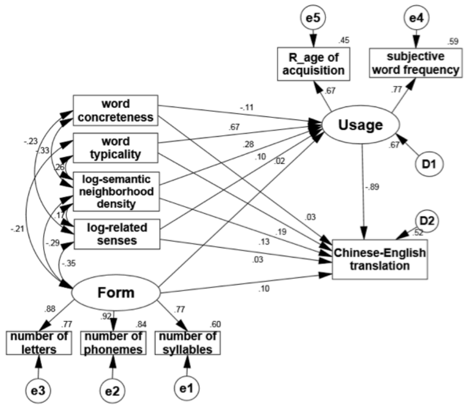
L1-to-L2 translation



Model 1



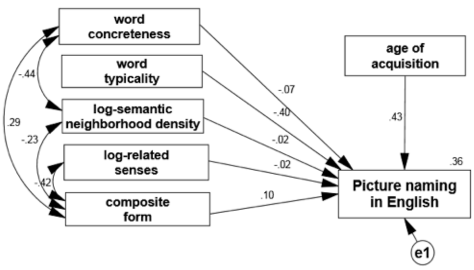
Model 2



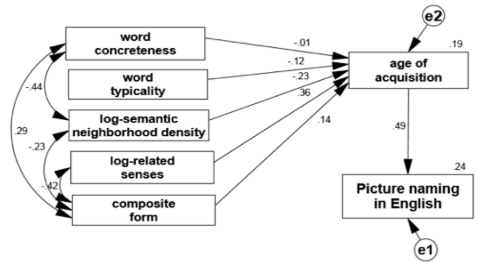
Model 3

Children:

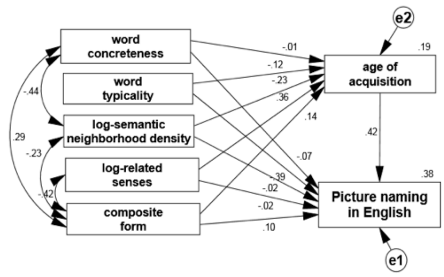
Picture naming



Model 1

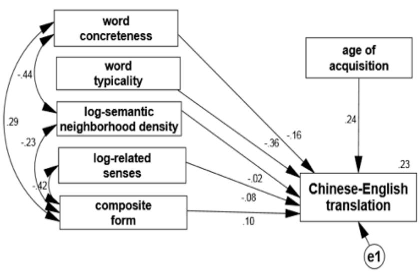


Model 2

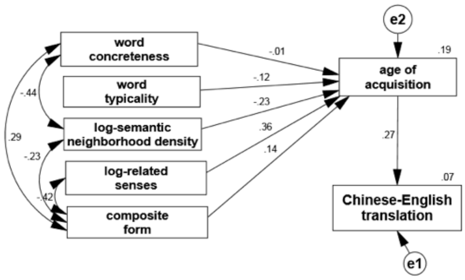


Model 3

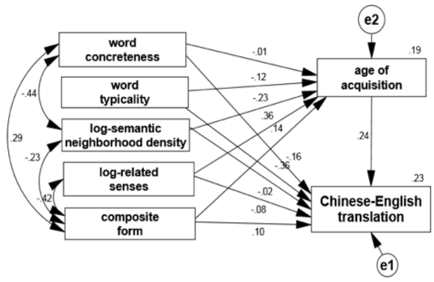
L1-to-L2 translation



Model 1



Model 2



Model 3

Activating Attributes in Frames

Jens Fleischhauer
Department of General Linguistics
Heinrich-Heine-Universität Düsseldorf
Universitätsstrasse 1
40225 Düsseldorf, Germany
fleischhauer@phil.uni-duesseldorf.de

1 Introduction

The general topic addressed in this paper is the activation of scalar attributes in the context of degree gradation of non-scalar verbs. Non-scalar verbs such as German *stinken* 'stink' do not lexically encode a scale, meaning there is no scalar attribute in their lexical representation. Nevertheless such verbs can be used in a degree context as in (1). In the sentence, the intensifier *sehr* 'very' specifies the intensity of the dog's smell.

(1) *Der Hund stinkt sehr.*
the dog stinks very
The dog stinks very much.

If the verb does not lexicalize a scale, a scalar attribute has to be activated in the degree context; otherwise the degree construction could not be interpreted. Therefore, I will argue (i) that the scalar attribute is retrieved from the conceptual knowledge associated with a meaning component specified in the verb, and (ii) that frames provide a suitable means of representing the process of (scalar) attribute activation. The aim of the paper is to illustrate how this process is constrained.

2 Verb gradation

Following Bierwisch (1989), gradation is a linguistic process of comparing two degrees on a scale. Gradation is usually associated with adjectives, and languages like English and German have special adjectival degree morphology such as comparative *-er* and superlative *-est* in English. However, gradation is not restricted to adjectives (Sapir , 1944; Bolinger , 1972); verbs and nouns can also be graded (see e.g. Morzycki (2009) on the gradation of nouns). Verbs and nouns differ from adjectives in not having special degree morphemes (at least in English and German). A further difference between the gradation of adjectives and verbs is that verb gradation is more complex than its adjectival equivalent. It is either possible to specify the temporal extent (duration or frequency) of an eventuality or to specify the degree of a gradable property associated with the verb. The first type is called 'extent gradation', the second is called 'degree gradation' (Bolinger , 1972; Löbner , 2012; Fleischhauer , 2014). Two German examples of verbal degree gradation are shown in (2).

(2) a. *Peter ist sehr gewachsen.*
Peter is very grown
'Peter has grown a lot.'
b. *Peter hat sehr geblutet.*
Peter has very bled
'Peter bled a lot.'

In (2-a), the intensifier *sehr* specifies the degree to which Peter increased in size; it is a vague, context-dependent high degree (see Fleischhauer (2013) for a deepter discussion of degree gradation of change of state verbs). In (2-b) the intensifier indicates the quantity of emitted blood.

There is a crucial difference between the verbs *wachsen* 'grow' and *bluten* 'bleed' in (2); the former is lexically scalar, whereas the latter is not. A verb is lexically scalar iff it expresses a scalar predication in every context of use (see, among others, Levin and Rappaport Hovav (2010) and Fleischhauer and Gamerschlag (2014) on scalar verbs). In (3-a) *wachsen* expresses a comparison between the size of the child at the beginning of the event and its size at the end of the event. Hence, it expresses a scalar predication although it is not modified by an intensifier.

(3) a. *Peter ist gewachsen.*
Peter is grown
'Peter has grown.'
b. *Peter hat geblutet.*
Peter has bled
'Peter bled.'

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

The sentence in (3-b) does not compare the quantity of blood emitted by the boy to some other quantity; hence, the verb is lexically non-scalar. This means that only *wachsen* but not *bluten* lexically encodes a scale.

Although the verb *bluten* is gradable (2-b), it does not lexicalize a scale. The gradation scale varies for different verbs: it is an intensity scale in (1) and a quantity scale in (2-b). Since the scale varies for different verbs, it is not contributed by the intensifier. Rather, a suitable gradation scale is rather from conceptual knowledge.

3 Frames

Frames, in the sense of Barsalou (1992a; 1992b), are recursive attribute-value structures. A frame is a representation of a concept and represents the referent of the concept in terms of its attributes, the values of the attributes, the attributes of the values and so on. One way of representing frames is by using attribute-value matrixes (AVMs) like in figure 1. The AVM in figure 1 shows a partial frame for the concept 'tree' (based on Petersen and Oswald (2012)). A tree consists of a crown and a trunk, hence CROWN[1] and TRUNK are attributes in the frame of 'tree'. The value of the attribute CROWN is the underspecified value or, in different terms, the uninstantiated type 'crown'. The value of trunk is the uninstantiated type 'trunk' which can be further characterized as having an attribute BARK. The bark of the tree is characterized as having a certain color.

$$\begin{bmatrix} \textit{tree} & \\ \textsc{crown} & \text{crown} \\ \textsc{trunk} & \text{trunk} \begin{bmatrix} \textsc{bark} & \text{bark} \begin{bmatrix} \textsc{color} & \text{color} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Figure 1: Partial frame for the concept 'tree'.

Following Löbner (1998; 2014) and Petersen (2007), attributes are partial functions; they assign a unique value to their possessor argument. The requirement of functionality provides a formal constraint on possible attributes. As attributes are functions, it is possible to distinguish scalar and non-scalar attributes by looking at their domains. If the values in the domain are linearly ordered, the attribute is a scalar one (e.g. SIZE). If there is no linear order of the domain's values, it is

[1]Attributes are written in small capitals.

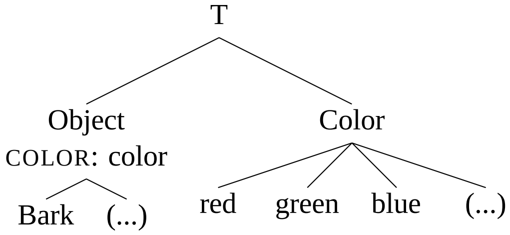


Figure 2: Partial type signature.

a non-scalar attribute, such as COLOR.

To restrict the admissible attributes for a frame and the admissible values for an attribute, types can be assigned to frames. Types are ordered with regard to their specificity in a type signature (Carpenter , 1992), as shown in figure 2. The type signature defines 'bark' as a subtype of the type 'object'; 'red', 'green' and 'blue' are defined as subtypes of 'color'. The type signature is enriched with appropriateness conditions (ACs) which serve two tasks: first, they restrict the set of appropriate attributes for frames to a certain type. Second, ACs specify the appropriate values for an attribute; it is required that all values of an attribute are of a certain type (see Petersen (2007), Petersen et al. (2008), Petersen and Gamerschlag (2014)). COLOR restricts its values to be of the type 'color' or one of its subtypes. Furthermore, the attribute COLOR is an appropriate attribute for 'object'. Since 'bark' is a subtype of 'object', it inherits this AC. Thus, objects of the type 'bark' have a color but do not have, for example, a price, since the type signature does not define PRICE as an appropriate attribute for 'bark'.

4 Frame analysis of degree gradation

In section 2, I suggested that the degree context activates the relevant gradation scale in the case of lexically non-scalar verbs. This process is not arbitrary but restricted by the lexical semantics of the verb. There are two reasons for this assumption: First, each semantic class of gradable verbs is only related to a single gradation scale. Second, different semantic classes of verbs are related to different gradation scales. As discussed above, verbs of substance emission such as *bluten* 'bleed' are related to a quantity scale (2-b), but verbs of smell emission, like *stinken* 'stink' in (1), are related to an intensity scale.

In the following, the analysis concentrates on the verb *bluten*. The verb denotes a process of substance emission. Its single argument is the emitter, the one who is emitting blood. The emittee, which is the emitted substance, is an implicit semantic argument of the verb (Goldberg (2005) speaks of an incorporated theme argument). A frame representation for *bluten*, capturing the mentioned aspects, is given in figure 3. The boxed numeral in the frame indicates structure sharing (Pollard and Sag , 1994) and indicates that the value of EMITTER is coextensive with a some other structure, the externally specified subject.

$$\begin{bmatrix} \textit{substance emission} & \\ \text{EMITTER} & \boxed{1} \\ \text{EMITTEE} & \text{blood} \end{bmatrix}$$

Figure 3: Frame for the verbal concept *bluten* ‘bleed’.

Degree gradation affects the quantity of the emitted blood; hence QUANTITY is an attribute of the emittee. The frame representation for *sehr bluten* ‘bleed a lot’ is shown in fig 4. The intensifier *sehr* activates the scalar attribute QUANTITY in the frame of *bluten* and specifies the value of QUANTITY as ‘high’.

$$\begin{bmatrix} \textit{substance emission} & \\ \text{EMITTER} & \boxed{1} \\ \text{EMITTEE} & \text{blood} \begin{bmatrix} \text{QUANTITY high} \end{bmatrix} \end{bmatrix}$$

Figure 4: Frame for *sehr bluten* ‘bleed a lot’.

As QUANTITY is an attribute of ‘blood’, it is the object knowledge associated with ‘blood’ that licenses its activation. A partial frame for ‘blood’ is given in figure 5.

$$\begin{bmatrix} \textit{blood} & \\ \text{CONSISTENCY} & \text{liquid} \\ \text{COLOR} & \text{red} \\ \text{QUANTITY} & \text{quantity} \end{bmatrix}$$

Figure 5: Partial frame for the concept *Blut* ‘blood’.

It is part of our knowledge of ‘blood’ that it has a certain consistency (‘liquid’), has a certain color (‘red’) and is of a certain quantity. While the attributes CONSISTENCY and COLOR have fixed values for blood, the value of QUANTITY is dependent on the possessor of the blood. In figure 5 the only scalar attribute is QUANTITY, hence it is the only attribute that can be activated in a degree context to provide a suitable gradation scale.

I propose the constraint in (4) as a restriction for the activation of scalar attributes in the frames of lexically non-scalar verbs:

(4) Only meaning components that are lexically specified in the verb license the activation of scalar attributes.

In the frame for *bluten* (figure 3) only the emittee is lexically specified as being blood. The emitter is not specified in the verb, rather it is introduced by the subject argument and therefore does not give access to specific conceptual knowledge.

5 Restricting the scalar attribute

An apparent problem is the claim that the frame for *bluten* only contains one scalar attribute, namely QUANTITY. It is clearly the case that we cannot only speak of the quantity of blood but also of its temperature or pressure. TEMPERATURE as well as PRESSURE are scalar attributes too, so the question emerges why it is only QUANTITY but not TEMPERATURE or PRESSURE that is activated in a degree context?

To tackle this problem one has to realize that the gradable verbs of substance emission are not restricted to those that express an emission of a liquid like blood. Other verbs of this class express the emission of a solid like hair in (5).

(5) *Die Katze hat sehr gehaart.*
the cat has very shed
‘The cat lost many hairs.’

The type signature in figure 6 defines ‘liquid’ to be a supertype of ‘blood’ and ‘water’ and to be a subtype of ‘substance’. ‘Solids’ are also a subtype of ‘substance’ and form the supertype of, for example, ‘hair’ and ‘scall’. The attributes shared by liquids and solids are inherited from their common supertype, for example CONSISTENCY and QUANTITY. But there are attributes which ‘hair’ and ‘blood’ do not share and these are inherited from the more specific supertypes ‘liquids’ and ‘solids’ respectively. For example, liquids do have a temperature and a pressure but we do not think of solids in terms of the attributes PRESSURE and TEMPERATURE. This does not result in the claim that solids do not have a temperature but I do not think that TEMPERATURE is an attribute in our object knowledge of ‘hair’ or ‘scall’; so we do not

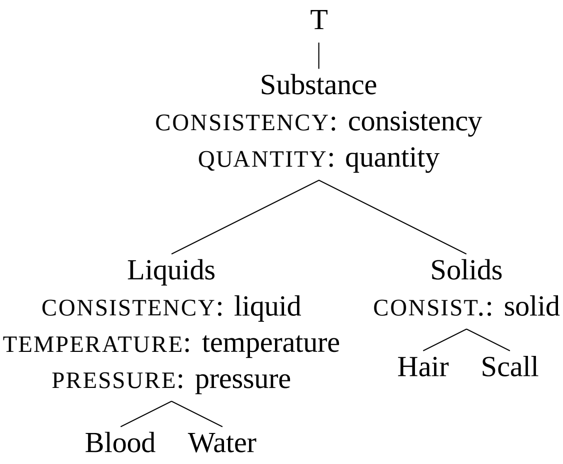


Figure 6: Partial type signature.

represents these concepts by using the attribute TEMPERATURE.

As verbs of substance emission do not only express the emission of liquids but of solids too, the admissible scalar attributes that can be activated in a degree context are restricted to those inherited from the common supertype of liquids and solids, which is 'substance'. Since QUANTITY but not TEMPERATURE or PRESSURE is inherited from 'substance', it is only QUANTITY that can be activated in the context of degree gradation. Beside the constraint in (4) a further constraint restricting the activation of scalar attributes is required:

(6) The activation of scalar attributes is restricted to those attributes which are inherited form the most specific common supertype.

The most specific common supertype for emittable substances like 'blood' and 'hair' is 'substance'. Hence, (6) restricts the activation of scalar attributes to those which are inherited from 'substance'; those attributes inherited from a more specific supertype like 'liquids' cannot activated in a degree context.

6 Conclusion

In this paper, I have shown that lexically nonscalar verbs can be graded by intensifiers like *sehr*. But this requires the activation of a suitable scalar attribute, otherwise the degree construction could not be interpreted. The process of attribute activation is not unconstrained, rather the lexical meaning of the verb as well as conceptual knowledge provide constraints on this process. The scalar attribute is activated from the conceptual knowledge associated with a meaning component lexically specified in the verb. Furthermore, the gradable attributes that can be activated are restricted to those inherited from the most specific common supertype. This ensures a homogeneous interpretation of degree gradation of verbs of substance emission, otherwise degree gradation of verbs (of substance emission) would be totally idiosyncratic.

Frames provide a suitable framework for the analysis of the sketched phenomenon as they allow representing lexical knowledge and conceptual knowledge in the same representational format. The frame analysis in this paper concentrates on a single semantic verb class but it can easily be extended to cover other classes of gradable verbs, for example verbs of smell/light/sound emission or experiencer verbs, too. I propose that the general constraints formulated in (4) and (6) hold for these classes of verbs as well, the only difference consists in the associated conceptual knowledge.

The process of attribute activation is not restricted to scalar attributs in the context of verbal degree gradation. A similar process occurs if verbs of sound emission are used for denoting motion events like in (7) (based on Kaufmann (1995, 93)). In this construction, a motion frame is activated which is licensed by the fact that the motion of a motorbike produces a yowling sound. In this case and in opposition to verbal degree gradation, knowledge of the subject referent is relevant too.

(7) *Das Motorrad jaulte über die*
the motorbike yowled over the
Kreuzung.
crossing
'The motorbike yowled over the crossing.'

It is a promising task for the future to explore the process of attribute activation in more details and to see how the activation of attributes from the conceptual knowledge is constrained by lexical semantics and other factors.

Acknowledgments

The paper is a result of my work in the Collaborative Research Center "The Structure of Representations in Language, Cognition, and Science" supported by the German Science Foundation (DFG).

References

[Barsalou 1992a] Lawrence Barsalou. 1992a. *Cognitive Psychology. An overview for cognitive scientists.* Hillsdale/ NJ, Lawrence Erlbaum Association.

[Barsalou 1992b] Lawrence Barsalou. 1992b. Frames, concepts, and fields. In A. Lehrer and E. F. Kittay (eds.), *Frames, fields, and contrasts*, 21–74. Hillsdale/ NJ, Lawrence Erlbaum Association.

[Bierwisch 1989] Manfred Bierwisch. 1989. The Semantics of Gradation. In M. Bierwisch and E. Lang (eds.), *Dimensional Adjectives*, 71–261. Berlin, Springer.

[Bolinger 1972] Dwight Bolinger. 1972. *Degree Words*. Mouton, The Hague.

[Carpenter 1992] Bob Carpenter. 1992. *The Logic of Typed Feature Structures.* Cambridge, Cambridge University Press.

[Fleischhauer 2013] Jens Fleischhauer. 2013. Interaction of telicity and degree gradation in change of state verbs. In B. Asrenijevic, B. Gehrke and R. Marin (eds.), *Studies in Composition and Decomposition of Event Predicates*, 125–152. Dordrecht, Springer.

[Fleischhauer 2014] Jens Fleischhauer. 2014. *Degree Gradation of Verbs.* Doctoral dissertation, Heinrich-Heine-Universität Düsseldorf.

[Fleischhauer and Gamerschlag 2014] Jens Fleischhauer and Thomas Gamerschlag. 2014. We're going through changes: How change of state verbs an arguments combine in scale composition. *Lingua* 141:30–47.

[Goldberg 2005] Adele Goldberg. 2005. Argument realization: The Role of constructions, lexical semantics and discourse factors. In J.-O. Ostman and M. Fried (eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, 17–43. Amsterdam, John Benjamins.

[Kaufmann 1995] Ingrid Kaufmann. 1995. What is an (im)possible verb? Restrictions on Semantic Form and their consequences for argument structure. *Folia Linguistica* XXIX/1-2:67–103.

[Löbner 1998] Sebastian Löbner. 1998. Definite Associative Anaphora. In S. Botley (ed.), *Proceedings of DAARC96 - Discourse Anaphora and Resolution Colloquium. Lancaster University, July 17th-18th*. Lancaster.

[Löbner 2012] Sebastian Löbner. 2012. Subcompositionality. In M. Werning, W. Hinzen and E. Machery (eds.), *The Oxford Handbook of Compositionality*, 220–241. Oxford, Oxford University Press.

[Löbner 2014] Sebastian Löbner. 2014. Evidence for frames from natural language. In T. Gamerschlag, D. Gerland, R. Osswald, and W. Petersen (eds.), *Frames and Concept Types: Applications in Language and Philosophy*, 23–68. Dordrecht/ Heidelberg/ New York, Springer.

[Morzycki 2009] Marcin Morzycki. 2009. Degree modification of gradable nouns: size adjectives and adnominal degree morphemes. *Natural Language Semantics* 17:175–203.

[Petersen 2007] Wiebke Petersen. 2007. Representation of Concepts as Frames. In J. Skilters, F. Toccafondi, and G. Stemberger (eds.), *Complex Cognition and Qualitative Science. The Baltic International Yearbook of Cognition, Logic and Communication.* Vol 2, 151–170. Riga, University of Latvia.

[Petersen et al.2008] Wiebke Petersen, Jens Fleischhauer, Peter Bücker, and Hakan Beseoglu. 2008. A Frame-based Analysis of Synaesthetic Metaphors. *The Baltic International Yearbook of Cognition, Logic and Communication.* Vol 3, 1–22.

[Petersen and Gamerschlag 2014] Wiebke Petersen and Thomas Gamerschlag. 2014. Why chocolate eggs can taste old but not oval: A frame-theoretic analysis of inferential evidentials. In T. Gamerschlag, D. Gerland, R. Osswald, and W. Petersen (eds.), *Frames and Concept Types: Applications in Language and Philosophy*, 199–220. Dordrecht/ Heidelberg/ New York, Springer.

[Petersen and Osswald 2012] Wiebke Petersen and Tanja Osswald. 2012. A Formal Interpretation of Concept Types and Type Shifts. In K. Kosecki and J. Badio (eds.), *Cognitive Processes in Language*, 183–191. Frankfurt, Peter Lang.

[Pollard and Sag 1994] Carl Pollard and Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar.* Chicago, The University of Chicago Press.

[Rappaport Hovav and Levin 2010] Malka Rappaport Hovav and Beth Levin. 2010. Reflections on manner/result complementarity. In M. Rappaport Hovav, E. Doron and I. Sichel (eds.), 21–38. Oxford, Oxford University Press. *Syntax, Lexical Semantics, and Event Structure*

[Sapir 1944] Edward Sapir. 1944. Grading: A Study in Semantics. *Philosophy of Science* 11(2):93–116.

Modelling semantic transparency in English compound nouns

Melanie J. Bell
Anglia Ruskin University
Cambridge
U.K.
melanie.bell@anglia.ac.uk

Martin Schäfer
Friedrich Schiller University
Jena
Germany
post@martinschaefer.info

1 Introduction

Semantic transparency is known to play an important role in the storage and processing of complex words (e.g. Marslen-Wilson et al. 1994), and human raters of transparency achieve high levels of agreement (e.g. Frisson et al. 2008, Munro et al. 2010). In the case of noun-noun compounds, overall transparency is largely determined by the transparency of the individual constituents. For example, Reddy et al. (2011) showed that the perceived transparency of a compound is highly correlated with both the sum and the product of the perceived transparencies of its constituents. Furthermore, many psycholinguistic studies find significant effects for semantic transparency using a four-way distinction based on perceived constituent transparency: transparent-transparent (e.g. *carwash*), transparent-opaque (e.g. *jailbird*), opaque-transparent (e.g. *strawberry*) and opaque-opaque (e.g. *hogwash*) (Libben et al. 2003). Bell and Schäfer (2013) modelled the transparency of individual compound constituents and showed that shifted word senses reduce perceived transparency, while certain semantic relations between constituents increase it. However, this finding is problematic in at least two ways. Firstly, it is not clear whether there is a solid basis for establishing whether a specific word sense is shifted or not. For example, *card* in *credit card* is clearly shifted if viewed etymologically, but may not synchronically be perceived as shifted due to its frequent use. Secondly, work on conceptual combination by Gagné and collaborators has shown that relational information in compounds is accessed via the concepts associated with individual modifiers and heads, rather than independently of them (e.g. Spalding et al. 2010 for an overview). This leads to the hypothesis that it is not whether a specific word sense is etymologically shifted, nor whether a specific semantic relation is used *per se*, that makes a compound constituent more or less transparent; rather, it is the degree of expectedness of a particular word sense and a particular relation for a given constituent. In this paper, we provide evidence in support of this hypothesis: the more expected the word sense and relation for a constituent, the more transparent it is perceived to be.

2 Method

We used the publicly available dataset described in Reddy et al. (2011), which gives human transparency ratings for a set of 90 compound types and their constituents (N1 and N2), and comprises a total of 7717 ratings. To model the expectedness of word senses and semantic relations for a given compound constituent, we used the constituent families of the compounds, which we extracted in a two step process. We took all strings of exactly two nouns that follow an article in the British National Corpus and which also occur four times or more in the USENET corpus (Shaoul and Westbury 2010). From this set, we extracted the positional constituent families for all constituent nouns in the Reddy et al. dataset, giving a total of 4553 compounds for the N1 families and 9226 for the N2 families. Each of these compound types was coded for the semantic relation between the constituents (after Levi 1978), and for the WordNet sense of the constituent under consideration (Princeton 2010). We then calculated the proportion of compound types in each constituent family with each semantic relation (relation proportion), and each WordNet sense of the constituent in question (synset proportion). We take these two measures to reflect the expectedness of the respective relations and WordNet senses of the constituents: if a relation or sense occurs in a high proportion of the constituent family, it is more expected. These variables were used, along with other quantitative measures, as predictors in ordinary least squares regression models of perceived constituent transparency. The final model for the transparency of N1 is given in Table 1:

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

	Coef	S.E.	t	Pr(>|t|)
Intercept	-4.6413	0.6593	-7.04	<0.0001
relation proportion in N1family	-0.2187	0.6013	-0.36	0.7161
log family size of N1	-0.0189	0.0931	-0.20	0.8395
synset proportion in N1family	-0.2426	0.6152	-0.39	0.6934
log synset count of N1	-0.7939	0.2469	-3.22	0.0013
compound proportion in N1 family (token-based)	3.0130	0.6788	4.44	<0.0001
log frequency of N1	0.8728	0.0569	15.34	<0.0001
relation proportion * log family size	0.3311	0.1305	2.54	0.0113
synset proportion * log synset count	0.6855	0.3161	2.17	0.0303
compound proportion * log frequency N1	-0.2804	0.0816	-3.44	0.0006

Table 1: Final model for the transparency of N1, R^2 adjusted = 0.334

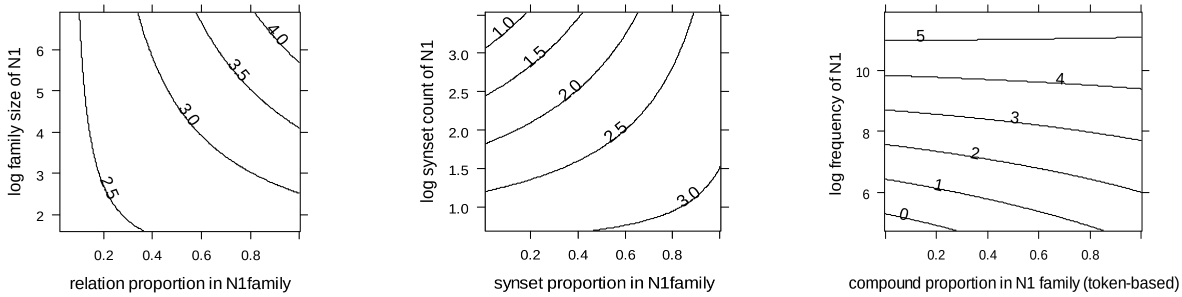


Figure 1. Interaction plots for N1 transparency

3 Results

All predictors in our model enter into significant interactions, and these are shown graphically in Figure 1, where the contour lines on the plots represent perceived transparency of the first constituent (N1). The first plot shows an interaction between relation proportion and overall (log) family size: for small families, relation proportion plays little role, whereas for larger families, in accordance with our hypothesis, the transparency of N1 increases with the proportion of the corresponding relation in the family. The second plot shows the interaction between the synset proportion and the total number of a constituent's senses (as listed in WordNet): only if there is a sufficient number of different senses in the family is their proportion a reliable predictor of semantic transparency. There is also a small but significant interaction between the log frequency of a constituent and the proportion of the constituent family (in terms of tokens) represented by the compound in question: this shows that transparency increases with frequency, but only in the lower frequently ranges does the proportion in the family play a role.

4 Conclusion

Overall, the model provides clear evidence for our hypothesis. N1 is rated as most transparent when it is a frequent word, with a large family, occurring with its preferred semantic relation and most frequent sense, and with few other senses to compete. We interpret the results as indicating that compound constituents are perceived as more transparent when they are more expected (both generally and with a specific sense) and when they occur in their most expected semantic environments. In information theory, the less expected an event, the greater its information content: in so far as perceived transparency is a reflection of expectedness, it can therefore also be seen as the inverse of informativity.

Acknowledgements

This work was made possible by three short visit grants from the European Science Foundation through NETWORDS - The European Network on Word Structure (grants 4677, 6520 and 7027), for which the authors are extremely grateful.

References

Bell, Melanie J. and Martin Schäfer. 2013. Semantic transparency: challenges for distributional semantics. In Aurelie Herbelot, Roberto Zamparelli and Gemma Boleda eds., *Proceedings of the IWCS 2013 workshop: Towards a formal distributional semantics*, 1–10. Potsdam: Association for Computational Linguistics.

Frisson, Steven, Elizabeth Niswander-Klement and Alexander Pollatsek. 2008. The role of semantic transparency in the processing of English compound words. *British Journal of Psychology* 991, 87–107.

Levi, Judith N. 1978. *The syntax and semantics of complex nominals*. New York: Academic Press.

Marslen-Wilson, William, Lorraine K. Tyler, Rachelle Waksler and Lianne Older. 1994. Morphology and meaning in the English mental lexicon. *Psychological Review* 101, 1: 3-33.

Munro, Robert, Steven Bethard, Victor Kuperman, Vicky Tzuyin Lai , Robin Melnick, Christopher Potts, Tyler Schnoebelen and Harry Tily. 2010. Crowdsourcing and language studies: the new generation of linguistic data. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, pp. 122-130. Association for Computational Linguistics.

Princeton University. 2010. *WordNet*. <http://wordnet.princeton.edu>

Reddy, Siva, Diana McCarthy and Suresh Manandhar. 2011. An empirical study on compositionality in compound nouns. In *Proceedings of The 5th International Joint Conference on Natural Language Processing 2011 IJCNLP 2011*, Chiang Mai, Thailand

Shaoul, Cyrus and Chris Westbury. 2010. An anonymized multi-billion word USENET corpus 2005-2010 http://www.psych.ualberta.ca/˜westburylab/downloads/usenet.download.html

Spalding, Thomas L., Christina L. Gagné, Allison C. Mullaly and Hongbo Ji. 2010. Relation-based interpretation of noun-noun phrases: A new theoretical approach. *Linguistische Berichte Sonderheft* 17, 283-315

Wurm, Lee H. 1997. Auditory processing of prefixed English words is both continuous and decompositional. *Journal of Memory and Language*, 37, 438–461.

A bottom up approach to category mapping and meaning change

Haim Dubossarsky
The Edmond and Lily Safra Center for Brain Sciences
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
haim.dub@gmail.com

Yulia Tsvetkov
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
ytsvetko@cs.cmu.edu

Chris Dyer
Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213 USA
cdyer@cs.cmu.edu

Eitan Grossman
Linguistics Department and the Language, Logic and Cognition Center
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
eitan.grossman@mail.huji.ac.il

Abstract

In this article, we use an automated bottom-up approach to identify semantic categories in an entire corpus. We conduct an experiment using a word vector model to represent the meaning of words. The word vectors are then clustered, giving a bottom-up representation of semantic categories. Our main finding is that the likelihood of changes in a word's meaning correlates with its position within its cluster.

1 Introduction

Modern theories of semantic categories, especially those influenced by Cognitive Linguistics (Geeraerts and Cuyckens, 2007), generally consider semantic categories to have an internal structure that is organized around prototypical exemplars (Geeraerts, 1997; Rosch, 1973).

Historical linguistics uses this conception of semantic categories extensively, both to describe changes in word meanings over the years and to explain them. Such approaches tend to describe changes in the meaning of lexical items as changes in the internal structure of semantic categories. For example, (Geeraerts, 1999) hypothesizes that changes in the meaning of a lexical item are likely to be changes with respect to the prototypical 'center' of the category. Furthermore, he proposes that more salient (i.e., more prototypical) meanings will probably be more resistant to change over time than less salient (i.e., less prototypical) meanings.

Despite the wealth of data and theories about changes in the meaning of words, the conclusions of most historical linguistic studies have been based on isolated case studies, ranging from few single words to few dozen words. Only recently though, have usage-based approaches (Bybee, 2010) become prominent, in part due to their compatibility with quantitative research on large-scale corpora (Geeraerts et al., 2011; Hilpert, 2006; Sagi et al., 2011). Such approaches argue that meaning change, like other linguistic changes, are to a large extent governed by and reflected in the statistical properties of lexical items and grammatical constructions in corpora.

In this paper, we follow such usage-based approaches in adopting Firth's famous maxim "You shall know a word by the company it keeps," an axiom that is built into nearly all diachronic corpus linguistics (see Hilpert and Gries, 2014 for a state-of-the-art survey). However, it is unclear how such 'semantic fields' are to be identified. Usually, linguists' intuitions are the primary evidence. In contrast to an intuition-based approach, we set out from the idea that categories can be extracted from a corpus, using a 'bottom up' methodology. We demonstrate this by automatically categorizing the entire lexicon of a corpus, using clustering on the output of a word embedding model.

We analyze the resulting categories in light of the predictions proposed in historical linguistics regarding changes in word meanings, thus providing a full-scale quantitative analysis of changes in the meaning of words over an entire corpus. This approach is distinguished from previous research by two main characteristics: first, it provides an exhaustive analysis of an entire corpus; second, it is fully bottom-up, i.e., the categories obtained emerge from the data, and are not in any way based on linguists' intuitions. As such, it provides an independent way of evaluating linguists' intuitions, and has the potential to turn up new, unintuitive or even counterintuitive

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

facts about language usage, and hence, by hypothesis, about knowledge of language.

2 Literature review

Some recent work has examined meaning change in large corpora using a similar bottom-up approach and word embedding method (Kim et al., 2014). These works analyzed trajectories of meaning change for an entire lexicon, which enabled them to detect if and when each word changed, and to measure the degree of such changes. Although these works are highly useful for our purposes, they do not attempt to explain why words differ in their trajectories of change by relating observed changes to linguistic parameters.

Wijaya and Yeniterzi (2011) used clustering to characterize the nature of meaning change. They were able to measure changes in meaning over time, and to identify which aspect of meaning had changed and how (e.g., the classical semantic changes known as ‘broadening,’ ‘narrowing,’ and ‘bleaching’). Although innovative, only 20 clusters were used. Moreover, clustering was only used to describe patterns of change, rather than as a possible explanatory factor.

3 Method

A distributed word vector model was used to learn the context in which the words-of-interest are embedded. Each of these words is represented by a vector of fixed length. The model changes the vectors’ values to maximize the probability in which, on average, these words could predict their context. As a result, words that predict similar contexts would be represented with similar vectors. This is much like linguistic items in a classical structuralist paradigm, whose interchangeability at a given point or ‘slot’ in the syntagmatic chain implies they share certain aspects of function or meaning.

The vectors’ dimensions are opaque from a linguistic point of view, as it is still not clear how to interpret them individually. Only when the full range of the vectors’ dimensions is taken together does meaning emerges in the semantic hyperspace they occupy. The similarity of words is computed using the cosine distance between two word vectors, with 0 being identical vectors, and 2 being maximally different:

$$(1) \qquad 1-\frac{\sum_{i=1}^{d} W_i \times W'_i}{\sqrt{\sum_{i=1}^{d}(W_i)^2} \times \sqrt{\sum_{i=1}^{d}(W'_i)^2}}$$

Where *d* is the vector’s dimension length, and W_i and W_i’ represent two specific values at the same vector point for the first and second words, respectively.

Since words with similar meaning have similar vectors, related words are closer to each other in the semantic space. This makes them ideal for clustering, as word clusters represent semantic ‘areas,’ and the position of a word relative to a cluster centroid represents its saliency with respect to the semantic concept captured by the cluster. This saliency is higher for words that are closer to their cluster centroid. In other words, a word’s closeness to its cluster centroid is a measure of its prototypicality. To test for the optimal size of the ‘semantic areas,’ different numbers of clusters were tested. For each the clustering procedure was done independently.

To quantify diachronic word change, we train a word vector model on a historical corpus in an orderly incremental manner. The corpus was sorted by year, and set to create word vectors for each year such that the words’ representations at the end of training of one year are used to initialize the model of the following year. This allows a yearly resolution of the word vector representations, which are in turn the basis for later analyses. To detect and quantify meaning change for each word-of-interest, the distance between a word’s vector in two consecutive decades was computed, serving as the degree of meaning change a word underwent in that time period (with 2 being maximal change and 0 no change).

Having two representational perspectives – synchronic and diachronic – we test the hypothesis that words that exhibit stronger cluster saliency in the synchronic model – i.e., are closer to the cluster centroid – are less likely to change over time in the diachronic model. We thus measure the correlation between the distance of a word to its cluster centroid at a specific point in time and the degree of change the word underwent over the next decade.

4 Experiment

We used the 2nd version of Google Ngram of fiction English, from which 10 millions 5-grams were sampled for each year from 1850-2009 to serve as our corpus. All words were lower cased.

Word2vec (Mikolov et al., 2013) was used as the distributed word vector model. The model was initiated to 50 dimensions for the word vectors’ representations, and the window size for context set to 4, which is the maximum size given the constraints of the corpus. Words that appeared less than 10 times in the entire corpus were discarded from the model vocabulary. Training the model was done year by year, and versions of the model were saved in 10 year intervals from 1900 to 2000.

The 7000 most frequent words in the corpus were chosen as words-of-interest, representing the entire lexicon. For each of these words, the cosine distance between its two vectors, at a specific year and 10 years later, was computed using (1) above to represent the degree of meaning change. A standard K-means clustering procedure was conducted on the vector representations of the words for the beginning of each decade from 1900 to 2000 and for different number of clusters from 500 until 5000 in increments of 500. The distances of words from their cluster centroids were computed for each cluster, using (1) above. These distances were correlated with the degree of change the words underwent in the following ten-year period. The correlation between the distance of words from random centroids of different clusters, on the one hand, and the degree of change, on the other hand, served as a control condition.

4.1 Results

Table 1 shows six examples of clusters of words. The clusters contain words that are semantically similar, as well as their distances from their cluster centroids. It is important to stress that a centroid is a mathematical entity, and is not necessarily identical to any particular exemplar. We suggest interpreting a word's distance from its cluster's centroid as the degree of its proximity to a category's prototype, or, more generally, as a measure of prototypicality. Defined in this way, *sword* is a more prototypical exemplar than *spear* or *dagger*, and *windows*, *shutters* or *doors* may be more prototypical exemplars of a *cover of an entrance* than *blinds* or *gates*. In addition, the clusters capture near-synonyms, like *gallop* and *trot*, and level-of-category relations, e.g., the modal predicates *allowed*, *permitted*, *able*. The very fact that the model captures clusters and distances of words which are intuitively felt to be semantically closer to or farther away from a category prototype is already an indication that the model is on the right track.

sword, 0.06	*allowed, 0.02*
spear, 0.07	*permitted, 0.04*
dagger, 0.09	*able, 0.06*

shutters, 0.04	*hat, 0.03*
windows, 0.05	*cap, 0.04*
doors, 0.08	*napkin, 0.09*
curtains, 0.1	*spectacles, 0.09*
blinds, 0.11	*helmet, 0.13*
gates, 0.13	*cloak, 0.14*
gallop, 0.02	*handkerchief, 0.14*
trot, 0.02	*cane, 0.15*

Table 1: Example for clusters of words using 2000 clusters and their distance from their centroids.

Figure 1 shows the analysis of changes in word meanings for the years 1950-1960. We chose this decade at random, but the general trend observed here obtains over the entire period (1900-2000). There is a correlation between the words' distances from their centroids and the degree of meaning change they underwent in the following decade, and this correlation is observable for different number of clusters (e.g., for 500 clusters, 1000 clusters, and so on). The positive correlations ($r>.3$) mean that the more distal a word is from its cluster's centroid, the greater the change its word vectors exhibit the following decade, and vice versa.

Crucially, the correlations of the distances from the centroid outperform the correlations of the distances from the prototypical exemplar, which was defined as the exemplar that is the closest to the centroid. Both the correlations of the distance from the cluster centroid and of the distance from the prototypical exemplar were significantly better than the correlations of the control condition (all *p's < .001 under permutations tests*).

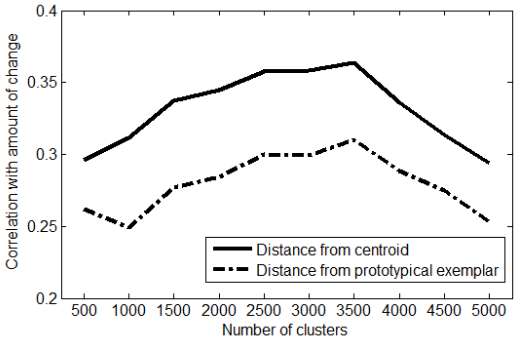


Figure 1. Change in the meanings of words correlated with distance from centroid for different numbers of clusters, for the years 1950-1960.

In other words, the likelihood of a word changing its meaning is better correlated with the distance from an abstract measure than with the distance from an actual word. For example, the likelihood of change in the *sword-spear-dagger* cluster is better predicted by a word's closeness to the centroid, which perhaps could be conceptualized as a non-lexicalized ‘elongated weapon with a sharp point,’ than its closeness to an actual word, e.g., *sword*. This is a curious finding, which seems counter-intuitive for nearly all theories of lexical meaning and meaning change.

The magnitude of correlations is not fixed or randomly fluctuating, but rather depends on the number of clusters used. It peaks for about 3500 clusters, after which it drops sharply. Since a larger number of clusters necessarily means smaller ‘semantic areas’ that are shared by fewer words, this suggests that there is an optimal range for the size of clusters, which should not be too small or too large.

4.2 Theoretical implications

One of our findings matches what might be expected, based on Geeraert’s hypothesis, mentioned in Section 1: a word’s distance from its cluster’s most prototypical exemplar is quite informative with respect to how well it fits the cluster (Fig. 1). This could be taken to corroborate Roschian prototype-based views. However, another finding is more surprising, namely, that a word’s distance from its real centroid, an abstract average of the members of a category by definition, is even better than the word’s distance from the cluster’s most prototypical exemplar.

In fact, our findings are consonant with recent work in usage-based linguistics on attractors, ‘the state(s) or patterns toward which a system is drawn’ (Bybee and Beckner, 2015). Importantly, attractors are ‘mathematical abstractions (potentially involving many variables in a multidimensional state space)’. We do not claim that the centroids of the categories identified in our work are attractors – although this may be the case – but rather make the more general point that an abstract mathematical entity might be relevant for knowledge of language and for language change.

In the domain of meaning change, the fact that words farther from their cluster’s centroid are more prone to change is in itself an innovative result, for at least two reasons. First, it shows on unbiased quantitative grounds that the internal structure of semantic categories or clusters is a factor in the relative stability over time of a word’s meaning. Second, it demonstrates this on the basis of an entire corpus, rather than an individual word. Ideas in this vein have been proposed in the linguistics literature (Geeraerts, 1997), but on the basis of isolated case studies which were then generalized.

5 Conclusion

We have shown an automated bottom-up approach for category formation, which was done on an entire corpus using the entire lexicon.

We have used this approach to supply historical linguistics with a new quantitative tool to test hypotheses about change in word meanings. Our main findings are that the likelihood of a word’s meaning changing over time correlates with its closeness to its semantic cluster’s most prototypical exemplar, defined as the word closest to the cluster’s centroid. Crucially, even better than the correlation between distance from the prototypical exemplar and the likelihood of change is the correlation between the likelihood of change and the closeness of a word to its cluster’s actual centroid, which is a mathematical abstraction. This finding is surprising, but is comparable to the idea that attractors, which are also mathematical abstractions, may be relevant for language change.

Acknowledgements

We thank Daphna Weinshall (Hebrew University of Jerusalem) and Stéphane Polis (University of Liège) for their helpful and insightful comments. All errors are, of course, our own.

Reference

Joan Bybee. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.

Joan Bybee and Clay Beckner. 2015. Emergence at the cross linguistic level. In B. MacWhinney and W. O'Grady (eds.), *The handbook of language emergence*, 181-200. Wiley Blackwell.

Dirk Geeraerts. 1997. *Diachronic prototype semantics. A contribution to historical lexicology*. Oxford: Clarendon Press.

Dirk Geeraerts. 1999. Diachronic Prototype Semantics. A Digest. In: A. Blank and P. Koch (eds.), *Historical semantics and cognition*. Berlin & New York: Mouton de Gruyter.

Dirk Geeraerts, and Hubert Cuyckens (eds.). 2007. *The Oxford handbook of cognitive linguistics*. Oxford: Oxford University Press.

Dirk Geeraerts, Caroline Gevaerts, and Dirk Speelman. 2011. How Anger Rose: Hypothesis

Testing in Diachronic Semantics. In J. Robynson and K. Allan (eds.), *Current methods in historical semantics*, 109-132. Berlin & New York: Mouton de Gruyter.

Martin Hilpert. 2006. Distinctive Collexeme Analysis and Diachrony. *Corpus Linguistics and Linguistic Theory*, 2 (2): 243–256.

Martin Hilpert and Stefan Th. Gries. 2014. Quantitative Approaches to Diachronic Corpus Linguistics. In M. Kytö and P. Pahta (eds.), *The Cambridge Handbook of English Historical Linguistics*. Cambridge: Cambridge University Press, 2014.

Yoon Kim, Yi-I Chiu, Kentaro Haraki, Darshan Hegde, and Slav Petrov. 2014. Temporal Analysis of Language through Neural Language Models. *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, 61-65. Baltimore, USA.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. *Proceedings of NAACL-HLT 2013:* 746–751. Atlanta, Georgia.

Eleanor H. Rosch. 1973. Natural Categories. *Cognitive Psychology* 4 (3): 328–350.

Eyal Sagi, Stefan Kaufmann, and Brady Clark. 2011. Tracing semantic change with latent semantic analysis. In K. Allan and J.A. Robinson (eds.), *Current methods in historical semantics*, 161-183. Berlin & New York: Mouton de Gruyter.

Derry T. Wijaya and Reyyan Yeniterzi. 2011. Understanding semantic change of words over centuries. In *Proceedings of the 2011 international workshop on DETecting and Exploiting Cultural diversiTy on the social web (DETECT ’11*) 35-40. Glasgow, United Kingdom.

What NN compounding in child language tells us about categorization

Maria Rosenberg
Dept. of Language Studies
Umeå university
SE-901 87 Umeå
SWEDEN
maria.rosenberg@umu.se

Ingmarie Mellenius
Dept. of Language Studies
Umeå university
SE-901 87 Umeå
SWEDEN
ingmarie.mellenius@umu.se

1 Introduction

The present study examines novel NN compounds, produced on line, in Swedish child language, with focus on categorization. Given that NN compounds denote objects, we concentrate on the categories those objects belong to. In that way, our study aims to provide evidence of object categorization in preschool children. Two questions are put forward:

(i) Does perception play a crucial role for the children's coinages?

(ii) In what way do structural and processing views on categorization apply to the data?

Swedish children produce compounds already at age two, reflecting the fact that compounding is a productive word formation device. In short, Swedish compounds are right-headed, written as one word, pronounced with a two-peak-intonation, and can exhibit liaison forms.

2 Theoretical background

Clark (2004) argues that language acquisition builds upon already established conceptual information, which enables the child to categorize objects, relations and events. Children rely mainly on shape as they embark on the mapping of words for objects onto their conceptual categories of objects, but also pay attention to texture, size, sound, motion and function. Even into adulthood, children continue the mapping of unknown linguistic items onto conceptual representations. Young children occasionally form emergent categories, based on non-conventional distinctions (e.g. *ball* for round things). Clark (2004) notes that the pairing of word to object enables the child to perceive similarities between cognitive categories, and allows for alternate perspectives on objects (cf. Waxman and Markow, 1995). "Do their [children's] categories reflect only what their language offers, or do they – must they– make use of other representations too?" (Clark, 2004:472).

Berman (2009) emphasizes that there is a substantial difference in adults' vs. children's lexicons of established compounds, and that children have to grasp inter alia the idea of subcategorization. Clark and Berman claim that "knowledge of the pertinent lexical items, and not the constructions they appear in, is more important for [children's] compounding" (1987:560).

In conceptual development, category structures change with age (Keil and Kelly, 1987). Object categorization allows generalization over properties of objects and of novel category members (Mandler, 2000).

Bornstein and Arterberry (2010) mention two complementing views of categorization: processing and structural. On the processing view, categories are flexible and category membership of objects can vary in different situations (cf. e.g. Jones and Smith, 1993). On the structure view, categories are hierarchically organized taxonomies (cf. e.g. Murphy, 2002). Instead of Rosch's (1978) superordinate-basic-subordinate, levels of category inclusiveness can be ordered in a neutral way, such as L_1 (animal), L_2 (cat, dogs), L_3 (collies, shepherds), L_4 (scotch collies, border collies) (Bornstein and Arterberry, 2010:3).

Whether categorization proceeds from concrete to abstract or the other way around is still under debate. Differentiation theory (e.g. Gibson, 1969) stipulates that the ability to make finer differentiations emerges after broad conceptions are acquired. Likewise, Bornstein and Arterberry (2010) indicate that more inclusive levels of categorization appear before less inclusive ones,

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

and that high perceptual contrasts have precedence over low. Fisher (2011) suggests that at age 3-5, perceptual information is anchored more strongly than conceptual information; cognitive flexibility develops with age.

Yet, according to Smith (1984), preschool children show the ability of both concrete categorization, due to perceptual characteristics, and abstract categorization, leaning on conceptual relationships. Nguyen and Murphy (2003) posit three categorization forms: taxonomic (see above), script and thematic. Script-based categories include objects (e.g. egg, cereal) with the same functional role in a routine event (e.g. eating breakfast). Thematic categories involve objects that usually appear together (e.g. bowl-cereal). Nguyen and Murphy (2003) show that children, aged 3 to7, use taxonomic and script categorization in a flexible way.

3 Data and method

The data consists of 383 spontaneously produced NN compounds from three monolingual Swedish children, aged 1 to 6, collected longitudinally and including contextual information. The children often give an explanation of the intended meaning, e.g. *hundstall* 'dog-stable', 'where dogs live, outside'. Hence, they seem to understand the semantics of their novel compound. We use a strict selection criterion: only non-established compounds in contemporary Swedish are considered.

As a first step to analyze our data, we sort the compounds in two ways: (i) based on N_1; (ii) based on N_2. This is a way of locating items belonging to a same morphological family (cf. Schreuder and Baayen, 1997). As a second step, the data is analyzed according to: (iii) level of inclusiveness; (iv) script; (v) thematicity; (vi) perception (real-world referent or not, high contrasts vs. low). As a third step, other characteristics appearing from the children's compounds are analyzed.

4 Analysis

In the analysis we provide evidence of categorization concerning larger groups of compounds. Below follows some preliminary findings. Note that the compounds can be analyzed according to different parameters and, thus, some of them go into several labels, depending on the parameter taken under account.

4.1 N_1 and N_2 sorting

The sorting of N1 and N2 shows that several nouns reoccur in the children's compounds. 126 N1 of the 383 compounds were either identical or belonging to the same morphological family, such as *morotsvatten* 'carrot-*s*-water and *moröttermacka* 'carrots-sandwich'. With respect to N_2, this number was as high as 143.

The largest morphological family found in our data contains *vatten* 'water'. 12 compounds are attested (7 compounds from one child, whereof 4 have *vatten* as N_1, and 3 as N_2). The two other children used *vatten* in 4 and 1 instances respectively, such as *vattenkaffe* 'water-coffee' or the aforementioned *morotsvatten* 'carrot-s-water'. Other nouns that reoccurred nearly ten times among the innovations of all three children were *bil* 'car' *kläder* 'clothes', *mamma* 'mommy' and *väg* 'road' (cf. 4.6).

It is worth noting that although the same nouns were used in several compounds, they did not always uphold the same relation to the other constituent: *pizzabil* 'pizza-car' was used for a car with a pizza print on it (viz. perceptually), whereas *dimbil* 'fog-car' referred to an imaginative car spraying fog (viz. abstractly).

Overall, the overlap between the same nouns being used in several compounds and as first or second constituent, can be taken as support for Clark's and Berman's (1987) claim (cf. 2) that children use lexical items that they are familiar with in their compounding.

4.2 Level of inclusiveness

As for the level of inclusiveness, the compounds in our data are situated on L_1 (*björkgrej* 'birch-thing'), L_2 (*brödrosta* 'toaster'), L_3 (*äppelsvans* 'apple-tail') or L_4 (*hjärtklackskorna* 'heart-heel-shoes), with L_3 as the predominant level. If we look only at N_1 or N_2 in isolation, they can also correspond to items located at L_1 (*djur* 'animal', L_2 (*björn* 'bear'), or L3 (*äppeljuice* 'apple-juice') in three-psart compounds.

Moreover, there are some compounds in our data containing a taxonomic relation between the constituents: two examples are *ugglafågel* 'owl-bird' and *skinndjur* 'skin-animals'.

4.3 Script-based categories

Entire sets of the compounds can be analyzed as having the same role with respect to a script, in which the compounds fulfill the same part. All three children categorize clothes according to season or weather, as indicated by N_1: *sommar-*

vantar ‘summer-gloves’, *snöstrumpor* ‘snow-stockings’ or *vinterficka* ‘winter-pocket’.

There are also compounds in our data where N_1 and N_2 participate in the same scripts that concern different types of edibles: *grötmjölk* ‘porridge-milk’ (eating breakfast”) or *pizzahamburgare* ‘pizza-hamburger’ (eating dinner”) or *saftglass* ‘syrup-ice cream’ (eating dessert).

4.4 Thematic categories

Thematic categories, items with close semantic association based on, e.g., contiguity, are numerous within the compounds. An example is *häxafiskspö* ‘witch-fish-wand’, where the child aims at a wand used by a witch, but confuses *trollspö* ‘magic-wand’ with *fiskespö* ‘fishing-pole’, and than adds the user of the item in question (actually a case of “overcategorization”, cf. 4.6).

Several themes are found. One is “sweets”, giving rise to numerous compounds, semantically associated or not, such as *silvergodis* ‘silver-candy’ and *godisstrumpor* ‘candy-stockings’.

Most of the thematic categorization found in the children’s innovation is abstract and grounded in conceptual information. Furthermore, the thematic relations are mostly of an inherent nature, such as manifested by *djungelträd* ‘jungle-tree’, rather than temporal, such as *fotbollsplanet* ‘football-planet’.

4.5 Perception

Compounds categorized according to Shape are attested, such as *R-paprika* ‘a piece of paprika that looks like a R’, or *mössaboll* ‘hat crumpled into the shape of a ball’. Shape may concern either the head or the non-head of the compound. Texture is involved in many of the children’s compounds, such as: *pälsmatta* ‘fur-carpet’. Prints are also a frequent way to distinguish among clothes they want to wear, or vehicles that they see, such as the above-mentioned “pizza-car”.

Yet, note that many of the children’s coinages, which involve perception, can do so in an imaginary way, or in other words, as mental imagery. A compound, such as *champagnetröja* ‘champagne-sweater’, was uttered to denote a non-existent sweater that the child just dreamt up when playing.

4.6 Overcategorization

We will use the term “overcategorization” to label some striking features among the children’s compounds. Underextension, often involving redundancy, is one way to arrive at overcategorization, as we see it. For instance, *kogräs* ‘cow-grass’ denotes ‘ordinary grass, that cows eat’ according to one child. Additionally, an ordinary car is referred to as *motorbil* ‘motor-car’, or *handfinger* ‘hand-finger’ is used instead of just finger for the body part. In these three examples, N_2 alone would have been the target like word to use, but the children limit its use further.

A quite odd categorization made by all three children, independently, is to add the goal of a direction to the direction: *kalasväg* ‘party-road’ or *mormorväg* ‘granny-road’. Recall that *väg* ‘road’ was one of the nouns that reoccurred frequently among the novel compounds (cf. 4.1). Hence, the three children seem to find it important to name particular roads.

Furthermore, nearly 20 of the children’s compounds contain one of the words *mamma* ‘mommy’, *pappa* ‘daddy’ or *bebis* ‘baby’ as N_1 or N_2, such as *mammfluga* ‘mommy-fly’, *fågelpappa* ‘bird-daddy’ or *bebismyra* ‘baby-ant’. All three children coined such compounds, which we interpret as a kind of emergent categorization, as well as of overcategorization. There were two types of relations involved in these compounds: animals or insects subcategorized according to human kinship terms as in the preceding examples; mommy or daddy subcategorized according to some habit, such as *cigarettpappa* ‘cigarette-daddy’.

4.7 Ad hoc categorization

Barsalou (1983) uses the label ad hoc categories for categories constructed on the spot to achieve certain goals, such as “things to sell at a garage sale”. These categories are much less established in memory than common categories. We interpret ad hoc categories to encompass compounds such as Downing’s (1977) “apple-juice seat”, and also the examples from Clark, Gelman and Lane (1985) claimed to involve a temporal relation, in contrast to compounds with inherent relations. According to Clark, Gelman and Lane (1985), children would more often use novel compounds to express inherent relations among objects. The opposite stand is taken by Mellenius (1997), supported by Berman (2009), who claims children’s novel compounds are “highly ‘context-dependent’ and hence more likely to express temporary rather than intrinsic relations” (2009:311).

Some innovations in our data can be analyzed as ad hoc instances that the children coin spontaneously without a real naming demand. They are typically difficult to understand, or does not make sense, outside the context of the utterance. An example is one child in the data that invents a triplet of compounds with *glass* 'ice cream' with the goal "things that could possibly constitute ice cream": *träglass* 'wood-ice cream', *sockerglass* 'sugar-ice cream' and *glassögon* 'ice cream-eyes'; the latter denotes, according to the child, 'eye-glasses but made of ice cream (*glass*) instead of glass (*glas*)'. Another example is *kungtröja* 'king-sweater', coined on the spot when playing: 'if you wear that sweater you will be the king'.

However, our data points in the direction that the children's innovations more often express inherent relations than temporal relations, but this issue certainly merits further investigation.

5 Conclusion

The study provides evidence of on-line categorization based on spontaneous production of novel NN compounds from three Swedish children. Compared to experimental situations, limited by the material used and the children's will and energy to participate, our collection of data is unique. It shows that high contrast perceptual features give rise to much subcategorization, however not at the expense of conceptual subcategorization, equally important in our data.

Since we lack clear longitudinal facts of how object categorization emerges within the children, the structure view is hard to apply. We can state that L_3 and L_4 categories appear around age 2, but lack numbers about their overall frequency in relation to more inclusive categories. Given that the children show cognitive flexibility in their categorization of an object in a particular way by producing an NN compound, the processing view conforms better to our data. To conclude, the children often categorize objects in a much more detailed way than adults do.

References

Lawrence W. Barsalou. 1983. Ad Hoc Categories. *Memory & Cognition*, 11(3):211-227.

Ruth A. Berman. 2009. Children's Acquisition of Compound Constructions. In Rochelle Lieber and Pavol Štekauer (eds.), *The Oxford Handbook of Compounding*, 298-322. Oxford: Oxford University Press.

Marc H. Bornstein and Martha E. Arterberry. 2010. The Development of Object Categorizaion in Young Children: Hierarchical Inclusiveness, Age, Perceptual Attribute, and Group versus Individual Analyses. *Developmental Psychology*, 46(2):350-365.

Eve V. Clark. 2004. How Language Acquisition Builds on Cognitive Development. *TRENDS in Cognitive Sciences*, 8(10):472-478.

Eve V. Clark and Ruth A. Berman. 1987. Types of Linguistic Knowledge: Interpreting and Producing Compound Nouns. *Journal of Child Language*, 14(3):547-567.

Eve V. Clark, Susan A. Gelman and Nancy M. Lane. 1985. Compound Nouns and Category Structure in Young Children. *Child Development* 56(1):84-94.

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

Anna V. Fisher. 2011. Processing of Perceptual Information Is More Robust than Processing of Conceptual Information in Preschool-Age Children: Evidence from Costs of Switching. *Cognition*, 119(2):253-264.

Eleanor J. Gibson. 1969. *Principles of Perceptual Learning and Development*. New York: Appleton-Century-Crofts.

Stevan R. Harnad (ed). 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Susan S. Jones and Linda B. Smith. 1993. Place and Perception in Children's Concepts. *Cognitive Development*, 8:113-139.

Frank C. Keil and Michael H. Kelly. 1987. Developmental changes in category structures. In S. Harnad (ed.), *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.

Jean M. Mandler. 2000. Perceptual and Conceptual Processes in Infancy. *Journal of Cognition and Development*, 1(1):3-36.

Ingmarie Mellenius. 1997. *The Acquisition of Nominal Compounding in Swedish*. [Travaux de l'Institut de Linguistique de Lund 31]. Lund: Lund University Press.

Gregory L. Murphy. 2002. *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Simone P. Nguyen and Gregory L. Murphy. 2003. An Apple is More Than Just a Fruit: Cross-Classification in Children's Concepts. *Child Development*, 24(6):1783-1806.

Eleanor Rosch. 1978. Principles of Categorization. In Eleanor Rosch and Barbara B. Lloyd (eds.), *Cognition and Categorization*, 27-48. Hillsdale, NJ: Erlbaum.

Robert Schreuder and Harald B. Baayen. 1997. How Complex Simplex Words Can Be. *Journal of Memory and Language*, 37:118-139.

Linda B. Smith. 1984. Young Children's Understanding of Attributes and Dimensions: A Comparison of Conceptual and Linguistics Measures. *Child Development*, 55(2):363-380.

Sandra R. Waxman and Dana B. Markow. 1995. Words as Invitations to Form Categories: Evidence from 12- to 13-Month-Old Infants. *Cognitive Psychology*, 29:257-302.

Using distributional data to explore derivational under-markedness: a study of the event/property polysemy in nominalization

Fabio Montermini
CLLE-ERSS
CNRS & Université de Toulouse 2 Jean Jaurès
fabio.montermini@univ-tlse2.fr

Abstract

This paper proposes a corpus-based analysis of deverbal suffixed nouns in Italian displaying an ambiguity between a clear event reading (*partenza* 'departure') and a clear property reading (*intelligenza* 'intelligence'). It focuses, in particular on words derived with the suffixes *-nza* and *-zione*. Three sets of syntactic contexts for words containing the two suffixes (high- and low-frequency *-nza* words and high-frequency *-zione* words) were extracted from a large corpus of contemporary Italian and coded according to their semantic reading. The comparison of the three datasets, on the one side, confirms an evolution, already observed in the literature, of *-nza* from a typically deverbal action suffix to a typically deadjectival property suffix, and, on the other side, shows that the same ambiguity is observed with *-zione* nouns, although, unlike the case of *-nza*, in this case it remains a marginal feature. The results obtained show the interest of large-scale empirical observations for the analysis of morphological phenomena, and militate in favour of a model in which (regular) polysemy should be considered as a constituting property of derived words.

1 Introduction

Although having a strong empirical basis is an important feature of most current studies of morphological derivational phenomena, these are often realized on (sometimes very large) series of complex words taken in isolation, or on the basis of some examples which are intended to exemplify the totality of the uses a derived lexeme can enter into, or at least the most common, 'unmarked', ones. This approach is reductive, however, especially in the study of the semantics of derivational processes, given the pervasiveness and systematicity of such phenomena as polysemy, semantic underspecification, etc. The first goal of this talk is thus to present arguments in favor of an usage-based model of derivational morphology, i.e. an approach in which the properties of complex lexemes, and the rules by which they are formed, are investigated via a thorough observation of their real contexts of use. The perspective adopted here is an exemplar-based one, in the sense that morphological competence is considered to emerge on the basis of the linguistic material speakers are exposed to, and that this dynamics can be simulated by taking into account large amounts of real usage data. The analysis presented can also be qualified as distributional, since it is inspired, in its fundamental assumption, by distributionalist approaches which are current in semantics (cf. Lenci, 2008 for an overview), according to which there is a correlation between a unit's meaning and its syntactic distribution.

The second goal of the talk is to provide evidence in favor of a non-compositional view of morphological derivation, according to which the semantic properties of complex lexemes cannot be simply computed on the basis of the subelements they contain, but rather on the basis of the lexical relations they enter into. The lack of full isomorphism between the form and the meaning of complex lexemes has been observed and investigated in many cases and in many languages. These include cases of over-marking, where an element (e.g. an affix) is present without carrying any evident meaning (cf. Roché, 2009, among others, for several examples in French), and parallel cases of under-marking, where a relevant semantic differentiation lacks an overt formal counterpart. The existence of the latter has been observed since a long time, and is linked with several other phenomena which are well known in the literature on morphology and

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

semantics, including (regular) polysemy (Booij, 2010: 78-79), multifunctionality (Luschützky and Rainer, 2013), morphological recycling (Hathout, 2011), etc. Morphological under-marking, precisely, constitutes the main focus of the analysis proposed here.

2 The data

In several languages, deverbal nouns present several instances of systematic polysemy, some of which are well described in the literature (e.g. action / result, cf. Rainer, 1996, Bisetto and Melloni, 2007). In particular, this paper is focused on cases of nominalization which, in spite of their frequency, have received less attention (but cf. Kerleroux, 2008 on French) namely deverbal nouns displaying an ambiguity between an event and a property reading, as in the following examples for the lexeme *vigilanza* in Italian:

(1)
la polizia ha effettuato una vigilanza continua
'police guaranteed a continuous control'

vs.

la sua vigilanza è calata del 50%
'his/her attention decreased of 50%'

Although Italian is the main focus of this paper, it should be observed that the same ambiguity can be observed in other Romance languages (and in English), involving several cognate affixes, such as those derived from Latin *-antia*, *-atio*, *-mentum*, *-tura*. In fact, this ambiguity should probably be ascribed to a specific property deverbal suffixes possessed in Latin (cf. (2)), since it is not observed with other morphological processes which cannot be directly linked to corresponding Latin constructions, such as verb-noun conversions or the (Germanic) deverbal suffix *-al* in English:

(2)
Lat.: adaequatio ('adequacy'), observantia ('observation')

The polysemy in question can also be linked to the larger spectrum of meanings that have been observed for deverbal nouns; the typical event reading and the typical property reading, in fact, can be considered as the two poles of a continuum which includes the nominalization of more or less permanent states (cf. Fradin, 2011, 2014):

(3)
una partenza / *latitanza / *intelligenza istantanea
'an instant departure / lam / intelligence'

una *partenza / latitanza / *intelligenza di due mesi
'a two-month departure / lam / intelligence'

una *partenza / *latitanza / intelligenza ammirevole
'an admirable departure / lam / intelligence'

Roughly, we can distinguish the three types above according to four dimensions, as exemplified in Table 1.

	action	punctual	bound	quantifiable
partenza	+	+	+	+
latitanza	+	–	+	–
intelligenza	–	–	–	–

Table 1: Types of deverbal nouns.

More specifically, the analysis presented has been carried on on nominalizations containing the two suffixes *-nza* and *-zione*[1], which share the property that, when they are constructed on a verb, they are linked, formally and semantically, to its participle (respectively, the present and the past participles) or to the homophonous adjective (*accogliere* / *accogliente* ⇒ *accoglienza* 'acceptance'; *educare* / *educato* ⇒ *educazione* 'education'). In addition, they can also be constructed on an adjective lacking a verbal counterpart (cf. *frequente* ⇒ *frequenza* 'frequency'; *perfetto* ⇒ *perfezione* 'perfection'), and in this case, base adjectives more often correspond to an individual-level predicate. In spite of their similarities, however, derived nouns in *-nza* and in *-zione* present several important differences. The most relevant one is probably the fact that while *-nza* is mainly attached to stative verbs (cf. Gaeta 2002), i.e. verbs which are semantically closer to (individual-level) adjectives (cf. Chierchia 1995: 177), no such tendency is observed with *-zione*, which, on the contrary, seems to display a preference for active event verbs. Consequently, apart from some exceptions (cf. *partenza* 'departure'), the property reading can be virtually applied to all *-nza* nouns, at least in some of their uses, while for *-zione* the situation is reversed:

[1] In fact, both suffixes may present several different forms in surface, whose selection depends on the form of the base they attach to. The forms given are intended to be labels for more abstract formal representations (on the formal problems posed by *–nza* and *–zione* cf., respectively, Gaeta, 2002: 127-129; Gaeta 2004, 346-348; Thornton, 1990; Montermini, 2010).

most of them do not allow this reading, while others accept it, a behavior for which no clear systematicity can be identified:

(4)
a. determinazione
‘determination’ / ‘determinedness’

educazione
‘education’ / ‘educatedness’

b. istruzione
‘instruction’ / *‘educatedness’

risoluzione
‘resolution’ / *‘determinedness’

In addition to the general features described above, some empirical observations motivate a deeper large-scale observation of the two derivational processes in question. First, for some of the *-nza* nouns displaying an event reading there exists a corresponding noun containing extra morphological material denoting a property (cf. *assistenza* ‘assistance’ ⇒ *assistenzialità*); similarly, to a past participle can correspond a derived noun denoting a property, either in concurrence with a *-zione* noun or not (cf. *risoluto* ⇒ *risolutezza*, *determinato* ⇒ *determinatezza* (vs. *determinazione*). Second, the observation of real language use shows that lexemes with a typical event meaning can be used as property nouns, and vice-versa, like in the following examples taken from the Web:

(5) La produzione basata sulla **concorrenza** del prezzo tende a tagliare i costi sostenuti dalla produzione basata sulla qualità.
‘Production based on low prices (lit. price concurrence) tends to cut the costs incurred by quality based production’.

Paolo […] era un uomo di estrazione nobile, di grande educazione e **istruzione** ed estremamente religioso e timorato di Dio.
‘St. Paul […] was a man of noble lineage, highly educated and very cultivated (lit. of great education and instruction), and extremely religious and God-fearing’.

Finally, as shown by Benincà and Penello (2005), and as confirmed by the data I have analyzed, while nouns with a pure event reading were the predominant output for *-nza* in ancient Italian, it is more employed today for the construction of property nouns. On the other hand, no comparable shift can be observed for *-zione*.

3 The analysis

In order to test the distribution of meanings for *-nza* and *-zione* nouns, in particular along the event / property divide, I extracted the 61 most frequent lexemes in *-anza* and *-enza* (the two possible formal variants)[2] in the CorIs[3], a large corpus of written Italian. For each of the lexemes in question, 100 contexts of occurrence were randomly selected, each of which was semantically coded according to its compatibility with one of the two meanings in question. In particular, the coding was based on such properties as the possibility of being determined by quantification or a measure adjective, or the presence / absence of temporal boundaries.

Figure 1 shows the distribution of meanings according to the class of the base (verb *vs.* adjective), and, as expected, a strong correlation between verbal bases and event reading, on the one side, and adjectival bases and property reading, on the other, are observed. The diagram also shows that, for the most frequent *-nza* nouns, the two schemes are more or less equally available.

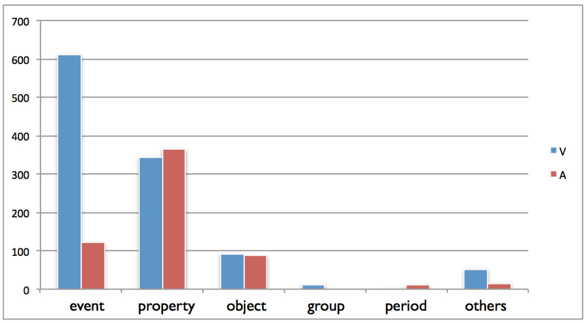


Figure 1: Distribution of meanings for the most frequent *-nza* nouns according to the class of the base.

In order to measure the functioning of this morphological process in the speakers’ synchronic competence, the same procedure was applied to low-frequency words containing *-nza* in the same corpus (62 lexemes overall having a frequency ≤ 3, 88 contexts overall). In this case, as shown in Figure 2, a strong preference of *-nza* for adjectival bases and for the property reading can be observed.

[2] The lexemes in question range from *presenza* (frequency 19,671) to *indifferenza* (frequency 1,671). Of course, the least was cleared from all lexemes ending in *-nza* that could not be clearly linked to a synchronically existing word.
[3] http://corpora.dslo.unibo.it/coris_ita.html.

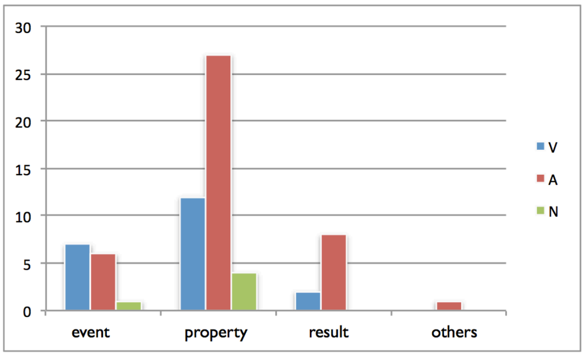


Figure 2: Distribution of meanings for the low-frequency *-nza* nouns according to the class of the base.

This result confirms the observation, mentioned above, that in the history of Italian *-nza* evolved from a (mainly) deverbal suffix forming event nouns to a (mainly) deadjectival suffix forming property nouns.

Moreover, in order to determine whether the features identified for *-nza* are specific to this construction or belong to deverbal nominal suffixes in general, the same procedure was further applied to a comparable set of words in *-zione* (including some possible formal variants, like *-sione*), namely the 61 most frequent forms in the CorIs[4].

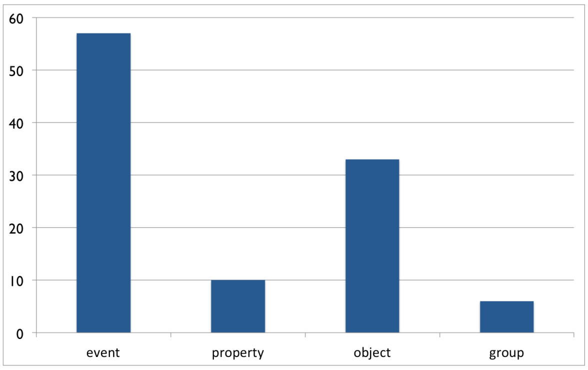


Figure 3: Distribution of meanings for *-zione* nouns.

As Figure 3 shows, the event / property polysemy remains marginal for *-zione*, thus suggesting that, while this polysemy can be considered as a constitutive property of the *-nza* word formation pattern, while it constitutes a rare and marked subpattern for *-zione*.

[4] Ranging from *amministrazione* (frequency 17,139) to *previsione* (frequency 3,987).

References

Paola Benincà, and Nicoletta Penello. 2005. Il suffisso *-anza/-enza* tra sincronia e diacronia. In M. Grossmann and A.M. Thornton (eds.), *La formazione delle parole. Atti del XXXVII Congresso internazionale di studi della Società di linguistica italiana. L'Aquila, 25-27 settembre 2003*, 69-86. Roma: Bulzoni.

Antonietta Bisetto, and Chiara Melloni. 2007. Result nominals: A lexical-semantic investigation. In G. Booij, B. Fradin, E. Guevara, S. Scalise and A. Ralli (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5). Fréjus, 15-18 September 2005*. Bologna: Università di Bologna.

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

Gennario Chierchia. 1995. Individual-level predicates as inherent generics. In: G.N. Carlson and F.J. Pelletier (eds.) *The Generic Book*, 176-223. Chicago / London: University of Chicago Press.

Bernard Fradin. 2011. Remarks on state denoting nominalizations. *Recherches linguistiques de Vincennes*, 40: 73-99.

Bernard Fradin. 2014. La variante et le double. In: F. Villoing, S. David and S. Leroy (eds.) *Foisonnements morphologiques. Études en hommage à Françoise Kerleroux*, 109-147. Nanterre: Presses Universitaires de Paris Ouest.

Livio Gaeta. 2002. *Quando i verbi compaiono come nomi. Un saggio di morfologia naturale*. FrancoAngeli, Milano.

Nabil Hathout. 2011. Une approche typologique de la construction des mots. In: M. Roché, G. Boyé, N. Hathout, S. Lignon, M. Plénat, *Des unités morphologiques au lexique*, 251-318. Paris: Hermès – Lavoisier.

Françoise Kerleroux. 2008. Des noms indisctincts. In: B. Fradin (ed.), *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, 113-132. Amsterdam / Philadelphia: John Benjamins.

Alessandro Lenci. 2008. Distributional semantics in linguistic and cognitive research. A foreword. *Rivista di Linguistica*, 20.1: 1-30.

Hans C. Luschützky, and Franz Rainer. 2013. Instrument and place nouns: A typological and diachronic perspective. *Linguistics*, 51.6: 1301-1359.

Fabio Montermini. 2010. The lexical representation of nouns and adjectives in Romance languages. *Recherches linguistiques de Vincennes*, 39: 135-161.

Franz Rainer. 1996. La polysémie des noms abstraits: historique et état de la question. In: N. Flaux, M. Glatigny, and M. Samain (eds.), *Les noms abstraits. Histoire et théorie. Actes du colloque de Dunkerque (15-18 septembre 1992)*, 116-126. Villeneuve d'Ascq: Presses Universitaires du Septentrion.

Michel Roché. 2009. Pour une morphologie *lexicale*. *Mémoires de la Société de Linguistique de Paris. Nouvelle série*, 17: 65-87.

Anna M. Thornton 1990/1991. Sui deverbali italiani in *-mento* e *-zione*. *Archivio glottologico italiano*, LXXV (2): 169-207 / LXXVI (1): 79-102.

A Distributional Semantics Approach to Implicit Language Learning

Dimitrios Alikaniotis **John N. Williams**
Department of Theoretical and Applied Linguistics
University of Cambridge
9 West Road, Cambridge CB3 9DP, United Kingdom
{da352|jnw12}@cam.ac.uk

1 Introduction

Vector-space models of semantics (VSMs) derive word representations by keeping track of the co-occurrence patterns of each word when found in large linguistic corpora. By exploiting the fact that similar words tend to appear in similar contexts (Harris, 1954), such models have been very successful in tasks of semantic relatedness (Landauer and Dumais, 1997; Rohde et al., 2006). A common criticism addressed towards such models is that those co-occurrence patterns do not explicitly encode specific semantic features unlike more traditional models of semantic memory (Collins and Quillian, 1969; Rogers and McClelland, 2004). Recently, however, corpus studies (Bresnan and Hay, 2008; Hill et al., 2013b) have shown that some 'core' conceptual distinctions such as animacy and concreteness are reflected in the distributional patterns of words and can be captured by such models (Hill et al., 2013a).

In the present paper we argue that distributional characteristics of words are particularly important when considering concept availability under implicit language learning conditions. Studies on implicit learning of form-meaning connections have highlighted that during the learning process a restricted set of conceptual distinctions are available such as those involving animacy and concreteness. For example, in studies by Williams (2005) (W) and Leung and Williams (2014) (L&W) the participants were introduced to four novel determiner-like words: *gi*, *ro*, *ul*, and *ne*. They were explicitly told that they functioned like the article '*the*' but that *gi* and *ro* were used with near objects and *ro* and *ne* with far objects. What they were not told was that *gi* and *ul* were used with living things and *ro* and *ne* with non-living things. Participants were exposed to grammatical determiner-noun combinations in a training task and afterwards given novel determiner-noun combinations to test for generalisation of the hidden regularity. W and L&W report such a generalisation effect even in participants who remained unaware of the relevance of animacy to article usage – semantic implicit learning. Paciorek and Williams (2015) (P&W) report similar effects for a system in which novel verbs (rather than determiners) collocate with either abstract or concrete nouns. However, certain semantic constraints on semantic implicit learning have been obtained. In P&W generalisation was weaker when tested with items that were of relatively low semantic similarity to the exemplars received in training. In L&W Chinese participants showed implicit generalisation of a system in which determiner usage was governed by whether the noun referred to a long or flat object (corresponding to the Chinese classifier system) whereas there was no such implicit generalisation in native English speakers. Based on this evidence we argue that the implicit learnability of semantic regularities depends on the degree to which the relevant concept is reflected in language use. By forming semantic representations of words based on their distributional characteristics we may be able to predict what would be learnable under implicit learning conditions.

2 Simulation

We obtained semantic representations using the `skip-gram` architecture (Mikolov et al., 2013) provided by the `word2vec` package,[1] trained with hierarchical softmax on the British National Corpus or on a Chinese Wikipedia dump file of comparable size. The parameters used were as follows: window size: B5A5, vector dimensionality: 300, subsampling threshold: $t = e^{-3}$ only for the English corpus.

The `skip-gram` model encapsulates the idea of distributional semantics introduced above by

[1] https://code.google.com/p/word2vec/

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

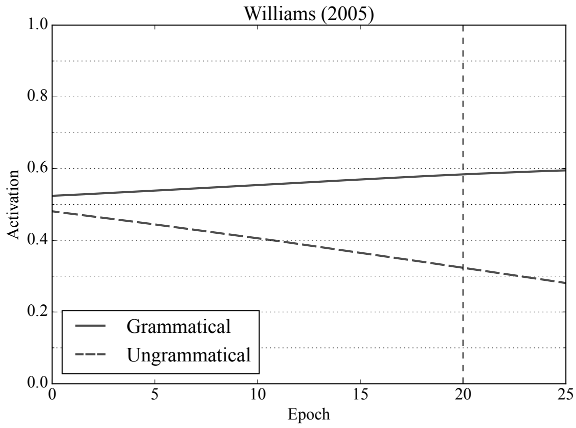


Figure 1: Generalisation gradients obtained from the Williams (2005) dataset. The gradients were obtained by averaging the output activations for the grammatical and the ungrammatical pairs, respectively. The network hyperparameters used were: learning rate: $\eta = 0.01$, weight decay: $\gamma = 0.01$, size of hidden layer: $\mathbf{h} \in \mathbb{R}^{100}$. For this and all the reported simulations the dashed vertical lines mark the epoch in which the training error approached zero. See text for more information on the experiment.

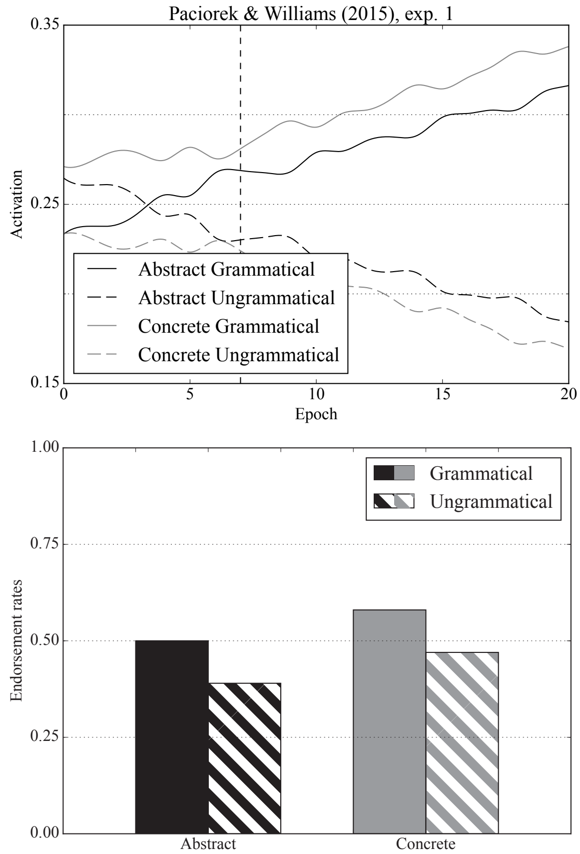


Figure 2: Results of our simulation along with the behavioural results of Paciorek and Williams (2015), exp. 1. The hyperparameters used were the same as in the simulation of Williams (2005).

learning which contexts are more probable for a given word. Concretely, it uses a neural network architecture, where each word from a large corpus is presented in the input layer and its context (i.e. several words around it) in the output layer. The goal of the network is to learn a configuration of weights such that when a word is presented in the input layer the nodes in the output that become more activated correspond to those words in the vocabulary, which had appeared more frequently as its context.

As argued above, the resulting representations will carry, by means of their distributional patterns, semantic information such as concreteness or animacy. Consistent with the above hypotheses, we predict that given a set of words in the training phase, the degree to which one can generalise to novel nouns will depend on how much the relevant concepts are reflected in the former words. If, for example, the words used during the training session do not encode animacy based on their co-occurrence statistics, albeit denoting animate nouns, then generalising to other animate nouns would be more difficult.

In order to examine this prediction, we fed the resulting semantic representations to a non-linear classifier (a feedforward neural network) the task of which was to learn to associate noun representations to determiners or verbs, depending on the study in question. During the training phase, the neural network received as input the semantic vectors of the nouns and the corresponding determiners/verbs (coded as 1-in-N binary vectors, where N is the number of novel non-words)[2] in the output vector. Using backpropagation with stochastic gradient descent as the learning algorithm, the goal of the network was to learn to discriminate between grammatical and ungrammatical noun – determiner/verb combinations. We hypothesise that this could be possible if either specific features of the input representation or a combination of them contained the relevant concepts. Considering the *distributed* nature of our semantic representations, we explore the latter option by adding a $\tanh$ hidden layer, the purpose of which was to extract non-linear combinations of features of the

[2]All the studies reported use four novel non-words.

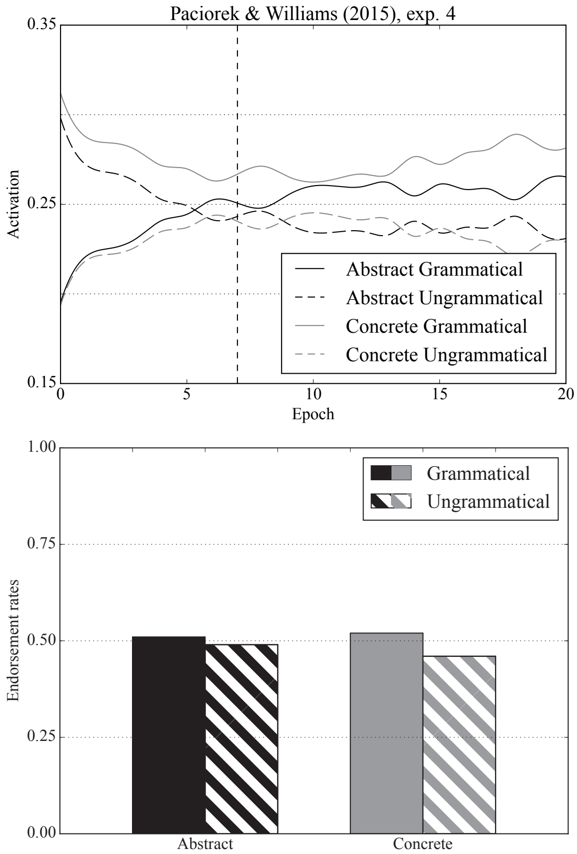


Figure 3: Results of our simulation along with the behavioural results of Paciorek and Williams (2015), exp. 4. The hyperparameters used were the same as in the simulation of Williams (2005).

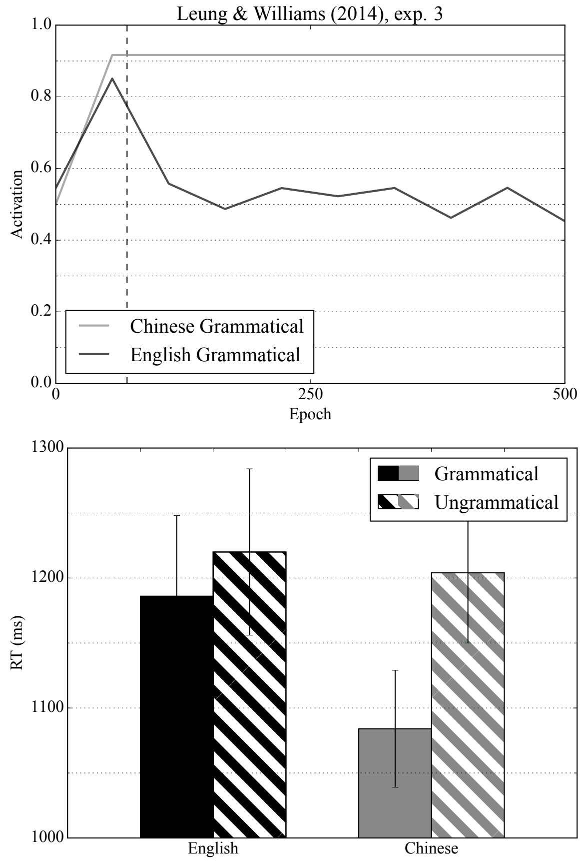


Figure 4: Results from Leung and Williams (2014), exp. 3. See text for more info on the measures used. The gradients for the ungrammatical combinations are (1 − grammatical). The value of the weight decay was set to $\gamma = 0.05$ while the rest of the hyperparameters used were the same as in the simulation of Williams (2005).

input vector. We then recorded the generalisation ability through time (epochs) of our classifier by simply asking what would be the probability of encountering a known determiner k with a novel word $\vec{w}$ by taking the softmax function:

$$p(y = k|\vec{w}) = \frac{\exp{(net_k)}}{\sum_{k' \in K} \exp{(net_{k'})}}. \tag{1}$$

3 Results and Discussion

Figures 1-4 show the results of the simulations across four different datasets which reflect different semantic manipulations. The simulations show the generalisation gradients obtained by applying eq. (1) to every word in the generalisation set and then keeping track of the activation of the different determiners (W, L&W) or verbs (P&W) through time. For example, in W where the semantic distinction was between animate and inanimate concepts ‘gi lion’ would be considered a grammatical sequence while ‘ro lion’ an ungrammatical one.

If the model has been successful in learning that ‘gi’ should be activated more given animate concepts then the probability $P(y = \text{gi}|\vec{w}_{\text{lion}})$ would be higher than $P(y = \text{ro}|\vec{w}_{\text{lion}})$. Fig. 1 shows the performance of the classifier on the testing set of W where, in the behavioural data, selection of the grammatical item was significantly above chance in a two alternative forced choice task for the unaware group. The slopes of the gradients clearly show that on such a task the model would favour grammatical combinations as well.

Figures 2-3 plot the results of two experiments from P&W which focused on the abstract/concrete distinction. P&W used a *false memory* task in the generalisation phase, measuring learning by comparing the endorsement rates between novel grammatical and novel ungrammatical verb-noun pairs. It was reasoned that if the participants had some knowledge of the system they would endorse more novel grammatical sequences. Expt 1 (Fig. 2) used generalisation items that were higher in semantic similarity to trained items than was the case in Expt 4 (Fig. 3). The behavioural results from the unaware groups (bottom rows) show that this manipulation resulted in larger grammaticality effects on familiarity judgements in Expt 1 than Expt 4, and also higher endorsements for concrete items in general in Expt 1. Our simulation was able to capture both of these effects.

L&W Expt 3 examined the learnability of a system based on a long/flat distinction, which is reflected in the distributional patterns of Chinese but not of English. In Chinese, nouns denoting long objects have to be preceded by a specific classifier while flat object nouns by another. L&W's training phase consisted of showing to participants combinations of thin/flat objects with novel determiners, asking them to judge whether the noun was thin or flat. After a period of exposure, participants were introduced to novel determiner – noun combinations, which either followed the grammatical system (*control* trials) or did not (*violation* trials). Participants had significantly lower reaction times (Fig. 4, bottom row) when presented with a novel grammatical sequence than an ungrammatical sequence, an effect not observed in the RTs of the English participants. The corresponding results of our simulations plotted in Fig. 4 show that indeed the regularity was learnable when the semantic model had only experienced a Chinese text, but not when it experienced the English corpus.

While more direct evidence is needed to support our initial hypothesis, our results seem to point to the direction that semantic information encoded by the distributional characteristics of words when found in large corpora can be important in determining what could be implicitly learnable.

References

Bresnan, J. and Hay, J. (2008). Gradient grammar: An effect of animacy on the syntax of give in New Zealand and American English. *Lingua*, 118(2):245–259.

Collins, A. M. and Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2):240–247.

Harris, Z. (1954). Distributional structure. *Word*, 10(23):146–162.

Hill, F., Kiela, D., and Korhonen, A. (2013a). Concreteness and Corpora: A Theoretical and Practical Analysis. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 75–83.

Hill, F., Korhonen, A., and Bentz, C. (2013b). A Quantitative Empirical Analysis of the Abstract/Concrete Distinction. *Cognitive Science*, 38(1):162–177.

Landauer, T. K. and Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.

Leung, J. H. C. and Williams, J. N. (2014). Crosslinguistic Differences in Implicit Language Learning. *Studies in Second Language Acquisition*, 36(4):733–755.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

Paciorek, A. and Williams, J. (2015). Semantic generalisation in implicit language learning. *Journal of Experimental Psychology: Learning, Memory and Cognition*.

Rogers, T. T. and McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. MIT Press.

Rohde, D., Gonnerman, L. M., and Plaut, D. C. (2006). An improved model of semantic similarity based on lexical co-occurrence. *Communications of the ACM*.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, 27:269–304.

Suffixation and the expression of time and space in Modern Greek

Anna Anastassiadis-Symeonidis
Aristotle University of Thessaloniki
ansym@lit.auth.gr

Abstract

This paper draws a comparison, through semasiological and onomasiological methods, of three Modern Greek (MG) suffixes *-in(os), -iatik(os)* and *-isi(os),* which construct denominal adjectives of time and/or space. Following D. Corbin's model (1987; 1991 and forthcoming) of Construction Morphology, an in depth analysis of these suffixes' semantics will be presented. The results suggest that, in order to construct a denominal adjective following the relational Lexeme Construction Rule (LCR_{REL}), a categorical, semantic and pragmatic compatibility are necessary between the base-noun and the suffix, as well as between the suffixed adjective and the noun of the noun phrase (NP); there are no synonyms even if the same noun is used as a base-noun. The three suffixes differ with respect to their semantic and pragmatic features; as a consequence, they are used in different genres. The data has been drawn from many dictionaries and especially from the *Reverse Dictionary of Modern Greek* (Anastassiadis-Symeonidis, 2002) as well as the *Corpus of Greek Texts* (Goutsos 2003).

1 The suffixes

1.1 The *-in(os)* suffix

This suffix is applied to a nominal base, or an adverbial one which could, however, be considered as a nominal one, given that these adverbs function also as nouns (Berthonneau 1989: 493). Consequently, we suggest a unified nominal base. In our corpus' base-nouns (87%) belong to the category of temporal or spatial nouns, e.g., *proinos* 'of early morning', *vradinos* 'of the evening', *kalokairinos* 'of the summer', *pashalinos* 'of Easter', *aprilianos* 'of April', *simerinos* 'today's/of today', *pantotinos* 'of ever - everlasting' - *vorinos* 'north', *antikrynos* 'of the opposite side', *brostinos* 'of the front', *makrinos* 'distant'.

The temporal sense base-nouns can label one of the denominations of the internal structure of the time unit YEAR, e.g., *kalokairi* 'summer', *theros* 'summer', *fthinoporo* 'autumn', or DAY, e.g., *proi* 'morning', *vradi* 'evening', or designate one of their special denominations, e.g., *Aprilios* 'April'. Aside from these base-nouns, we observe that the base can be selected from the names of important celebrations e.g., *Pasha* 'Easter', and that the specific deictic (NOW) denominations construct denominal adjectives exclusively with the suffix *-in(os),* e.g., *simerinos* 'of today', *apopsinos* 'of this evening', *htesinos* 'of yesterday', *torinos* 'of now', *fetinos* 'of this year', *persinos* 'of last year', *pantotinos* 'of ever - everlasting'.

Following our observation of spatial sense base-nouns we operate a distinction between: (i) a group of nouns referring to geographical terms, e.g., *vorras* 'north', *oros* 'mountain', *thalassa* 'sea'; (ii) toponyms, e.g., *Alexandria* 'Alexandria'; and (iii) adverbs constructing denominations within the deictic system (HERE), e.g., *antikry* 'across', *konta* 'near', *makria* 'far', *piso* 'behind'.

Finally, based on the context, the remaining nouns in the corpus (13%) can be categorized as conveying spatial meaning (provenance), e.g., *agheladhino ghala* 'cow's milk', *vodhino/hoirino kreas* 'bovine (beef)/pig (pork) meat', *kreatini/tyrini evdhomadha* 'Meatfare/Cheesefare week', *anthropini symperifora* 'human behaviour'. The same principles hold for the adjectives *foteinos* 'bright', *faeinos* 'brilliant', *skoteinos* 'dark', *alithinos* 'real', that originate in ancient Greek, where the base-noun functioned as a spa-

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

tial noun; relevant passages are preserved where the nouns *fos* 'light' and *skotos* 'darkness' refer to the source that transmits light and darkness respectively (Giannakis, 2001). Similarly, *alithinos* 'real' refers to location, since –according to Plato– truth originates from the real world.

1.2 The *-iatik(os)* suffix

From a semantic point of view, we notice that approximately 85% of the corpus consists in bases which are temporal nouns referring to time-measure units, e.g., *hronos* 'year' *minas* 'month' *(e)vdhomadha* 'week' as well as their reanalyses, including two subsets: (i) denominations of special units, e.g., *Dheftera* 'Monday', *Triti* 'Tuesday', *Ianouarios* 'January', *Fevrouarios* 'February'; and (ii) denominations related to the internal structure of the above units, e.g., *proï* 'morning', *mesimeri* 'midday', *anoiksi* 'spring' (Berthonneau, 1989).

In addition, the base can be selected among important days of public holidays or religious celebrations with which people mark time, and which are therefore categorized as temporal nouns, e.g., *Protomaghia* 'First of May', *Protaprilia* 'First of April', *Protochronia* 'New Year's Day', *Pasha* 'Easter', *Hristoughenna* 'Christmas', *Aghio-Vasilis* 'the feast day of Saint Vasilios', *Aï-Dhimitris* 'the feast day of Saint Demetrios', *Kathari Dheftera* 'Clean/Ash Monday', *apokria* 'Carnival festivities', *paramoni* 'Eve'. Finally, the suffix *-iatik(os)* is attached to the base form of 7 nouns, seemingly not associated with a temporal sense: *paidh(i)* 'child', *ghiort(i)* 'celebration', *skol(i)* 'leisure', *feggar(i)* 'moon', *ghampr(os)* 'groom', *nyf(i)* 'bride', *kefal(i)* 'head'. However, these nouns can be encountered in contexts that associated to important moments of people's lives, e.g., *ghampriatiko kostoumi* 'bridegroom's suit', *nyfiatiko traghoudi* 'wedding song', *paidhiatika kamomata* 'childish antics'.

1.3 The *-isi(os)* suffix

The suffix *-isi(os)* is associated with the notion of 'provenance' (Tsopanakis, 1994), which is diachronic in nature, particularly since the suffix *-isi(os)* is derived from the latin suffix *-ēnsis* which is associated with this notion (Meyer, 1895). This is a spatial provenance (where the base is a proper or common noun referring to the natural landscape or to man-made places (Le Pesant, 2011), e.g., *vounisios aeras* 'mountain air', *limnisio psari* 'fish of the lake'); even if the base-noun refers to an animal, e.g., *arnisia païdakia* 'lamb cutlets', *ghidhisio ghala* 'goat milk', *katsikisio tyri* 'goat cheese', to a plant, e.g., *kalampokisio alevri* 'corn flour', *thymarisio meli* 'thyme honey', to an artefact, e.g., *varelisia bira* 'draught', to a human or human-like being (human entity) or to parts of the human body, through extension, e.g., *flevisio aima* 'veins' blood' or through an intension reading, related to a stereotypical meaning, e.g., *gherontisia foni* 'elderly's voice'.

The availability of the suffix *-isi(os)* in contemporary language use is rather restricted, as it is not encountered in cases where it is possible to construct non-attested lexemes which constitute nothing more than coincidental gaps (Corbin, 1987: 177).

2 Is there synonymy?

We argued that the *-in(os)* suffix constructs denominal adjectives related to space and time, that the *-iatik(os)* suffix constructs denominal adjectives related to time and that the *-isi(os)* suffix constructs denominal adjectives of provenance, related to the notion of space. The question will thus be the following: can we talk about synonymy between the temporal and spatial denominal adjectives constructed with the aforementioned suffixes and the same base-noun?

If we take into account the pragmatic feature [learned], a feature with a non-binary value (Anastassiadis-Symeonidis and Fliatouras, 2004), we notice that for the base-nouns with a [+learned] value, only the *-in(os)* suffix is applied, that for the base-nouns with a [-learned] value only the suffixes *-iatik(os)* and *-isi(os)* are applied, and, that for the base-nouns with a [+/-learned] value all three suffixes *-in(os), -iatik(os)* and *-isi(os)* are applied. The reason is that the suffix *-in(os)* constructs denominal adjectives localizing in space and time objectively, i.e., free of prototypical or stereotypical perceptions (Geeraerts, 1985), contrary to the suffixes *-iatik(os)* and *-isi(os),* that are associated with the individual's everyday life. Consequently, the derived adjectives are not synonymous, even if the aforementioned suffixes are attached to the same base, e.g., *vradino/*vradhiatiko dheltio eidhiseon* 'the evening news report', or to a synonymous base, e.g., *arnisia/*provatisia païdhakia* 'lamb cutlets'. This is the reason for which only adjectives in *-in(os)* are encountered in scientific and religious discourse, in greater percentages in premeditated speech on television and the radio, as well as in newspapers. This means, semantic/pragmatic factors determine the genre of text where a lexeme may be encountered. It is not by chance that the pragmatic feature [learned] is attributed to a suffix found in ancient Greek and the feature [-learned] to suffixes that appeared later, during the Hellenistic era.

3 Compatibility

A categorical as well as semantic and pragmatic compatibility are therefore necessary between the base-noun and the suffix as well as between the derived noun and the modified noun. For instance, there would be an issue of categorical compatibility if the suffix *-in(os)* or the suffix *-iatik(os)* were attached to a verb-base. There would be an issue of semantic compatibility if the suffix *-in(os)* were attached to a non-temporal/spatial base-noun or if the suffix *-iatik(os)* were attached to a non-temporal base-noun. Lastly, there would be an issue of pragmatic compatibility if the suffix *-in(os)* were attached to a [-learned] base-noun or if the suffix *-iatik(os)* were attached to a [+learned] base-noun, e.g., if the adjective *aniksiatikos* 'of spring' modified the noun *isimeria* 'equinox'.

Therefore, each of the aforementioned suffixes is characterized by their categorical, semantic, and pragmatic/stylistic specifications and, according to this "genetic inheritance", it participates in the LCR_{REL}. Subsequently, within the framework of Construction Morphology, the notion of compatibility constitutes the key to grammaticality judgements.

4 Predictions

Starting from the semantic function of each suffix at the word-construction level of words that belong to the same onomasiological field, on one hand, similarities as well as differences at both the semantic and pragmatic level can be explained. For example, terms such as: *kalokairiatikos – kalokairinos* 'of the summer', *kampisios – pedhinos* 'of/in a plain'; on the other hand, predictions can be formulated, in the sense that restrictions are imposed, e.g., *avrianos - *avriatikos* 'of tomorrow', *kontinos - *kontaios* 'near' (similarly: *mesaios* 'middle'), *ghenarisios - *ianouarisios* 'of January'.

According to this model we are able to explain:

a) The reason why it is possible to derive adjectives with different suffixes from the same base-noun e.g., *vradhinos – vradhiatikos* 'of the evening', *agheladhisios – agheladhinos* 'of a cow/cow's (milk/meat)': the suffix *-in(os)* selects certain properties from the anaphoric/descriptive meaning of the base-noun, whereas the suffixes *-iatik(os)* and *-isi(os)* select from the base-nouns those properties that correspond to an experiential meaning associated with everyday life. We can thus explain why the adjectives in *-in(os)* and *-iatik(os)*, or those in *-in(os)* and *-isi(os)* are not synonyms.

b) The reason why certain suffixes cannot be attached to certain base-nouns: compatibility is required between the two. The adjectives in *-in(os)* are likely derived from the [+learned] or [+/-learned] allomorph of the base-noun, whereas the adjectives in *-iatik(os)* and *-isi(os)* are derived from the [-learned] or [+/-learned] allomorph of the base-noun, e.g., *mesimvrinos* and *mesimeriatikos* but **mesimvriatikos* 'midday', *pedhinos* and *kampisios* but **pedhisios*/**kampinos* 'of/in a plain', *therinos* but **theriatikos* 'of the summer', *heimerinos* but **heimeriatikos* 'of the winter', *omfalios* and *afalisios* but **omfalisios,* **afalios* 'umbilical'.

c) The reason why both the adjectives *kalokairinos* and *kalokairiatikos* 'of the summer' are grammatical without being synonymous: they both share the [+/-learned] feature.

d) The reason why it is grammatical to say *praghmatika anoiksiatikos kairos* 'real spring weather', *praghmatika vounisios aeras* 'real mountain air', but we do not say **praghmatika earini isimeria* 'real vernal equinox*, *praghmatika oreinos oghkos* 'real mountain massif': the adverb *praghmatika* 'real/proper' modifies qualifying adjectives but not taxonomic/relational ones.

e) The reason why the suffix *-in(os)* is selected in utterances that refer to the speaker's "HERE and NOW", within the deictic system: adjectives in *-in(os)* merely denote a location in space and time; that is, within the NP, they create a temporal or spatial relationship between the modified noun and the time period or the location signified by the base-noun. Conversely, the suffix *-iatik(os)* is associated with a subjective, experiential and/or stereotypical temporal meaning, while the suffix *-isi(os)* is experientially associated with the notion of provenance, e.g., *brostinos - *brostisios* 'of the front', *simerinos – *simeriatikos, *simerisios* 'of today'.

f) The reason why the adjectives *tritiatikos* 'of Tuesday', *tetartiatikos* 'of Wednesday', *pemptiatikos* 'of Thursday' (and the corresponding adverbs) are not encountered in written texts: are they potential or non-grammatical words?

According to the theoretical framework followed throughout this article, the aforementioned words are constructed according to the LCR_{REL} and are, therefore, potential words. However, they are not encountered in written texts due to pragmatic factors, as individuals – marking time and demarcating their life according to a sum – in our case, a sum of days –, are inclined to pay attention only to the beginning and the end, that is, for people, the days that mark the beginning and the end of the week are of particular importance.

Based on what I have stated above, I suggest the following categorization of the three suffixes according to semantic criteria:

	experiential	objective
space	*-isi(os)*	*-in(os)*
time	*-iatik(os)*	

Table 1: Semantic distribution of the suffixes *-iatik(os), -in(os), -isi(os)*

5 Impact on the theory of derivation

Every suffix is characterized by their categorical, semantic, and pragmatic/stylistic specifications and, according to this "genetic inheritance", they participate in the LCRs. Consequently, within the field of Construction Morphology, the notion of compatibility is key notion for grammaticality judgements. Thus, it seems to me that it is a bit far-fetched to attribute anomalies/exceptions, or even a lack of productivity, to lexicon merely because the study of lexicon constitutes unmapped territory (see also Anastassiadis-Syméonidis, 2003).

Similarly, as there is no synonymy between lexemes, there is neither synonymy between suffixes nor between their derivatives, even if the related suffixes are attached to the same base or if the same suffix is attached to a synonymous base.

Lastly, semantic/pragmatic reasons determine the genre of text wherein a derived lexeme will appear, due to semantic/pragmatic features of both the base as well as the suffix.

6 Conclusion

Since the lexicon does not constitute a separate level of linguistic analysis, but horizontaly cuts through all levels, the properties of those levels are to be taken into consideration, that is, phonological, morphological, syntactic, semantic and pragmatic.

This study examines the abstract system – in the form of LCRs and the suffixes' semantic instruction, which, according to several theories, is homogeneous. However, the present study is based on actual language use, since it takes into consideration rich authentic language data within context, linguistic production of native speakers, as well as metalinguistic texts. In particular, the study of concordances in the *Corpus of Greek Texts* illustrated the breadth of use of derivatives that carry the suffixes in question.

The homogeneity of the abstract system is contrasted to the linguistic variety characterizing the use of the system, and simultaneously, it constitutes an essential linguistic attribute.

In our case, variety is associated with the varying degrees of availability of the suffixes in question, as well as with the [+/- learned] feature. This simultaneous examination is beneficial to both, as it bridges the gap between theory and practice to the extent that one fuels the other. This is a dynamic, dialectical relationship that explains language change, which has been a topic of interest either in the form of borrowing, during earlier times, or through the non-frequent occurrence of the *-isi(os)* suffix in contemporary language.

Furthermore, an association has been attempted between the onomasiological method – which, in our case, originates from the notion of time and space – and the semasiological method. The latter, starting from the form of the suffixes *-iatik(os)*, *-in(os)* and *-isi(os)*, focused on the extensive analysis of their semantic instruction, unlike other studies that are limited to a basic presentation of semantic features.

Within D. Corbin's theoretical framework of Construction Morphology, meaning occupies a central role, since the units that contribute to it are meaning-bearing units. The constructed lexemes demand a more complex analysis at the semantic level in comparison to simple ones. The reasons are multiple: (i) because two elements participate – the base and the suffix; (ii) because the suffix is encountered in many other constructed lexemes; (iii) because the base is part of other constructed lexemes with a different suffix; and, (iv) because the meaning and the behavior at the level of anaphora of constructed lexemes are associated with their morphological structure. Through implementing this theoretical framework, it was possible to compare the semantic instruction of the suffixes *-iatik(os), -in(os)* and *-isi(os)*, the interpretation of semantic similarities and differences between derived words that carry those suffixes, as well as the interpretation of grammaticality through the notion of compatibility between the base-noun and the suffix with regard to grammatical category, meaning, and pragmatic level.

References

Anastassiadis-Syméonidis, A. 1996. A propos de l'emprunt suffixal en grec moderne. *Cahiers de Lexicologie 68/1,* 79-106.

Anastassiadis-Syméonidis, A. 2002. *Reverse Dictionary of Modern Greek.* Institute of Modern Greek Studies, Aristotle University of Thessaloniki, www.komvos.edu.gr, www. greek-language.gr

Anastassiadis-Syméonidis, A. 2003. Inflexion and derivation: myths and truths. *Studies on Greek Linguistics – Proceedings of the 24th Annual Meeting of the Department of Linguistics, School of Philology, Faculty of Philosophy,* Aristotle University of Thessaloniki, 43-54.

Anastassiadis-Syméonidis, A. 2008. Les adjectifs temporels suffixés en *-in(os)* et *-iatik(os)* en grec moderne. In Bernard Fradin (éd.) *La raison morphologique - Hommage à la mémoire de Danielle Corbin,* Amsterdam, Benjamins, 17-27.

Anastassiadis-Syméonidis, A. 2009. Suffix *-isi(os)* in Modern Greek. *Studies in Greek Linguistics – Proceedings of the 29th Annual Meeting of the Department of Linguistics, School of Philology, Faculty of Philosophy,* Aristotle University of Thessaloniki, 58-73.

Anastassiadis-Syméonidis, A. 2010. Pourquoi une langue emprunte-t-elle des suffixes ? L'exemple du grec et du latin. *META 55, 1.* Mélanges en hommage à André Clas, 147-157.

Anastassiadis-Symeonidis, A. & Fliatouras, A. 2004. The distinction between learned and non learned in Modern Greek. *Proceedings of the 6th International Conference of Greek Linguistics, University of Crete,* 110-120.

Babiniotis, G. 1998, 2002. *Dictionary of Modern Greek,* Athens, Centre of Lexicology.

Baldinger, K. 1964. Sémasiologie et onomasiologie, *Revue de linguistique romane 28,* 249-272.

Berthonneau, A.-M. 1989. *Composantes linguistiques de la référence temporelle. Les compléments de temps. Du lexique à l'énoncé.* Thèse d'Etat, Paris VII.

Buck, C. 1933. *Comparative Grammar of Greek and Latin,* Σικάγο, University of Chicago.

Corbin, D. 1987. *Morphologie dérivationnelle et structuration du lexique.* Villeneuve d'Ascq: Presses Universitaires de Lille 1991^2.

Corbin, D. 1991. Introduction – La formation des mots: structures et interprétations. *Lexique 10,* 7-30.

Corbin, D. forthcoming. *Le lexique construit. Méthodologie d'analyse*.

Dictionary of Modern Greek 1933. Athens, newspaper *Proïa,* St. Dimitrakos.

Dictionary of Standard Modern Greek 1998. Thessaloniki, Institute of Modern Greek Studies, Aristotle University of Thessaloniki.

Filos, P. 2008. *Studies in the Morphology of Latin Loanwords into Greek: Evidence from the Papyri.* PhD, University of Oxford.

Geeraerts, D. 1985. Les données stéréotypiques, prototypiques et encyclopédiques dans les dictionnaires. *Cahiers de Lexicologie 46/1,* 27-43.

Giannakis, G. 2001. Light is Life, Dark is Death: An Ancient Greek and Indo-european Metaphor, *Dodoni-Philologia 30,* 127-153.

Goutsos, D. 2003. Corpus of Greek Texts (CGT). *Proceedings of the 6th International Conference of Greek Linguistics, University of Crete,* 930-939.

Gross, G. 1994. Classes d'objets et description des verbes, *Langages 115,* Paris, Larousse, 15-31.

Haspelmath, M. & Sims, A. 2010^2. *Understanding Morphology.* London, Hodder Education.

Klairis, Chr. & Babiniotis, G. 2005. *Grammar of Modern Greek. Structural/functional – communicative.* Athens, Ellinika Grammata.

Kleiber, G. 1990. *La sémantique du prototype. Catégories et sens lexical,* Paris, P.U.F.

Le Pesant, D. 2001. *Les noms locatifs.* HDR, Paris XIII.

Liddell, H. G. & Scott, R. 1996. *A Greek-English Lexicon,* Oxford University Press.

Meyer, G. 1895. *Neugriechische Studien* III. *Die lateinischen Lehnworte im Neugriechischen.* Sitzungsberichte, usw., Band 132, Vienna.

Modern Greek Grammar (of Demotic Greek) 1941. Thessaloniki, Institute of Modern Greek Studies, Aristotle University of Thessaloniki, 1978^2.

Palmer, L.R. 1946. *A Grammar of the Post-Ptolemaic Papyri.* London, Oxford University Press.

Petrounias, E. 1998. *Dictionary of Standard Modern Greek* [Etymological Part], Thessaloniki, Institute of Modern Greek Studies, Aristotle University of Thessaloniki.

Psaltes, St. 1913. *Grammatik der Byzantinischen Chroniken.* Göttingen, Vandenhoeck & Ruprecht.

Rey, A. 1992. *Dictionnaire historique de la langue française.* Paris, Dictionnaires Le Robert, 2 vol.

Taylor, J.R. 1989. *Linguistic Categorization – Prototypes in Linguistic Theory.* New York, Oxford University Press.

Temple, M. 1993. *Le sens des mots construits : pour un traitement dérivationnel associatif,* PhD, University of Lille III.

Same same but different: Type and typicality in a distributional model of complement coercion

Alessandra Zarcone[1]
Universität des Saarlandes
Saarbrücken, Germany

Sebastian Padó[2]
Universität Stuttgart
Stuttgart, Germany

Alessandro Lenci[3]
Università degli Studi di Pisa
Pisa, Italy

[1]`zarcone@coli.uni-saarland.de`, [2]`pado@ims.uni-stuttgart.de`, [3]`alessandro.lenci@unipi.it`

Abstract

We aim to model the results from a self-paced reading experiment, which tested the effect of semantic type clash and typicality on the processing of German complement coercion. We present two distributional semantic models to test if they can model the effect of both type and typicality in the psycholinguistic study. We show that one of the models, without explicitly representing type information, can account both for the effect of type and typicality in complement coercion.

1 Introduction: Complement Coercion

Complement coercion (*The author began the book* → *reading the book*) has been shown to cause an increase in processing cost (Pylkkänen and McElree, 2006; Katsika et al., 2012), which has been ascribed to a *type clash* between an event-selecting verb (*begin*) and an entity-denoting object (*book*). The increase in processing costs is found in comparison with a baseline condition, where the same verb is combined with an event-denoting object (*journey*), which does not trigger a type clash.

A second influence on processing cost is the *thematic fit* or *typicality* of the fillers of the verb's argument slots (Bicknell et al., 2010; Matsuki et al., 2011): high-typicality combinations are processed more quickly than low-typicality ones (*the mechanic checked the brakes / the spelling*).

Distributional semantic models (DSMs) can successfully model a range of psycholinguistic phenomena, including the effect of typicality on complement coercion (Zarcone et al., 2012). However, they generally do not include a notion of type. Can a DSM account for effects both of type and typicality?

In this paper, we consider experimental results from a study on complement coercion in German that manipulates both type and typicality. We discuss the performance of existing DSMs and a novel DSM combination. We also discuss how type information can be emerge from distributional information.

2 Manipulating Type and Typicality

In a self-paced reading study on German complement coercion (Zarcone et al., in preparation), we have manipulated both type and typicality. The dataset consists of 20 pairs of subjets (S) and aspectual verbs (V). Each pair is combined with four nominal objects (O) in SOV order:

[$_S$ Das **Geburtstagskind**] hat [$_O$ mit den Geschenken
[$_S$ The **birthday boy**] has [$_O$ with the presents
/ der Feier / der Suppe / der Schicht] [$_V$ **angefangen**].
/ party / soup / work shift] [$_V$ **begun**].

The objects are: a high-typicality entity (*presents*); a high-typicality event (*party*); a low-typicality entity (*soup*); and a low-typicality event (*work shift*). The low-typicality objects are drawn from the high-typicality objects of other S-V pairs.

The self-paced reading study yielded the following significant effects: (1) an effect of typicality on reading times ($t = 2.28, p = .02$) at the object region (indicating subject-object integration), (2) an effect of object type on reading times ($t = -2.5, p = .01$) at the verb region (the region of the type clash), (3) an interaction of type and thematic fit at the verb region ($t = 2.04, p = .04$). Mean reading times per condition are reported in Table 1. In sum, the study shows that complement coercion involves both type and typicality. Thus, computational models of complement coercion need to account for both.

3 Modeling the Experimental Results

Distributional semantic models (DSMs) represent word meaning as high-dimensional vectors recording co-occurrences with elements of their

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

	Object region *mit den Geschenken* *with the presents*	**Verb region** *angefangen* *began*
high-fit entity	642	819
high-fit event	655	736
low-fit entity	667	802
low-fit event	710	806

Table 1: Mean reading times per condition (in ms) in the self-paced reading study.

usage contexts. Semantic similarity is defined in terms of a vector similarity metric such as cosine.

Distributional Memory (DM, Baroni and Lenci (2010)) is a DSM that includes syntactic knowledge into the word representations. More concretely, the TypeDM version of DM records word-relation-word tuples $\langle w_1 \ \mathtt{r} \ w_2 \rangle$. The tuples are weighted by *Local Mutual Information* (Evert, 2005), which can be employed to model predicate-argument typicality. For example, the weight of ⟨*book* `obj` *read*⟩ is higher than ⟨*label* `obj` *read*⟩, which in turn is higher than ⟨*elephant* `obj` *read*⟩. TypeDM has been shown to be versatile and effective in several semantic tasks, including predicting verb-argument plausibility.

3.1 Complement Coercion and DSMs.

DM has been extended into the Expectation Composition and Update model (ECU, Lenci (2011)), a family of procedures that can be used to predict the typicality of one sentence part given other sentence parts. E.g., to model the typicality at the verb region in a German sentence with SOV word order (e.g. *Das Geburtstagskind hat mit dem Geschenk angefangen / The birthday boy has with the present begun*), ECU determines the thematic fit for the verb given subject and object:

- compute an expectation for the verb given the subject s, as the distribution over verbs v defined by the weights of the tuples $\langle s \ \mathtt{subj} \ v \rangle$
- compute an expectation for the verb given the object o, as the distribution over verbs v defined by the weights of the tuples $\langle o \ \mathtt{obj} \ v \rangle$.

To combine the subject and object expectations, we combine the two distributions component by component, typically either by sum or products. This distribution is then represented in a vector space by computing the centroid or prototype of the vectors of the 20 most expected verbs. Finally, the thematic fit for a verb v given the subject s and the object o is its cosine similarity to the centroid.

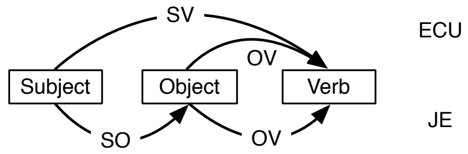


Figure 1: ECU vs. Joint Expectations for the verb

ECU. We call the models following the ECU procedure *SOV+* and *SOV**, depending on their combination operation (sum and product, respectively). Simpler models only consider the influence of subject or object on the verb (*SV* and *OV* respectively), just by leaving out the combination step. These models can successfully account for reading time results on a dataset of complement coercion in German that manipulates typicality but not type (Zarcone et al., 2012).

In order to test ECU on a dataset which manipulates both type and typicality, we evaluate the following ECU models on the complement coercion data in (Zarcone et al., in preparation): *SO* to model effects at the object given the subject; *SOV+*, *SOV** and *OV* to model effects at the verb. We expect these models to account for the typicality effect at the object (1), but not for the type effects at the verb (2,3).

The results are summarized in Table 2 (left and middle). In accordance with our prediction, *SO* correctly yields the typicality effect at the object ($F = 7.38$, $p < 0.01$). Neither *SOV+*, *SOV**, nor *OV* can model the type-typicality interaction at the verb (3). Surprisingly, though, *SOV** and *OV* yield (2), an effect of type at the verb ($F = 5.3228$, $p < 0.05$ and $F = 20.388$, $p < 0.001$, respectively).

Joint Expectations. The reading time study found that the subject-object typicality effects linger at the verb, interacting with type. The main shortcoming of ECU is its inability to model the typicality effects at the verb. This is due to the architecture of the *SOV* models (cf. Fig. 1, top): they compute the expectations for the verb first from the subject (*SV*) and update them with the object's expectations (*OV*). They ignore the interaction between subject and object (SO) – the source of typicality effects (1,3) – corresponding to the assumption that this interaction should only matter at the object. In order to account for this, we draw an analogy to the concept of *joint probability*:

$$P(S, O, V)$$

	non-compos.		ECU		JE	
	SO	*OV*	*SOV+*	*SOV**	SO+OV	*SO*OV*
(1) effect of typicality at the object region (SO interaction)	✓	×	×	×	✓	✓
(2) effect of type at the verb region (type clash)	×	✓	×	✓	×	✓
(3) type x thematic fit interaction at the verb region	×	×	×	×	×	×

Table 2: Overview of the results of the different DSMs: non-compositional, ECU and JE.

which is equivalent (by the chain rule), to

$$P(S)P(O|S)P(V|O)$$

Treating the first term as a constant prior, we obtain

$$P(O|S)P(V|O)$$

which we can interpret distributionally as motivation to *reweight* the typicality of the verb given the object with the typicality of the object given the subject, thus re-introducing the subject-object interaction into the verb prediction (cf. Figure 1, bottom).

In the *Joint Expectation (JE)* model, the thematic fit score assigned to the target verb is influenced both by the verb's thematic fit with the object (the verb's initial thematic fit score, equivalent to the ECU weight for the ⟨*object* obj *verb*⟩ tuple) and by the object's thematic fit with the subject (equivalent to the ECU weight for the ⟨*subject* verb *object*⟩ tuple), which in turn is used to reweight the verb's score.

Similar to ECU, there is a choice of combination operations in JE (sum or product). Since JE can be formulated as a simple wrapper around ECU, ECU can be used to compute the individual components (e.g. SO, OV, or more complex ones) and these then just need to be combined additively (SO+OV) or multiplicatively (SO*OV).

The right-hand side of Table 2 shows the results for JE. *SO+OV* yields an effect of typicality ($F = 6.777$, $p < 0.05$) but no effect of type (2) or interaction (3). *SO*OV* yields two main effects of (2) type ($F = 7.2359$, $p < 0.05$) and typicality ($F = 7.2359$, $p < 0.01$), although no interaction (3).

Comparing the two models, we see that ECU *SO* accounts for the results obtained at the object (1), but the *SOV* models cannot explain the interaction with typicality on the verb (2,3). JE (*SO * OV*) models the effects of both type (2) and typicality at the verb, but does not (yet) account for their interaction (3).

4 Discussion: Type and Typicality

We found that the *SO* model successfully accounts for the effect of typicality at the object. This is not surprising: one of the most typical tasks successfully performed by distributional models such as ECU is predicting verb-argument plausibility, and ECU had already been successful in modeling effects of typicality on reading times in German complement coercion (Zarcone et al., 2012).

On the other hand, the ECU *SOV* models were not able to account for the type–typicality interaction at the verb. The JE model (*SO * OV*), which we presented as an alternative to the ECU model to better account for the typicality effects at the verb, yielded effects of both type and typicality at the verb, but did not account for their interaction.

Our most surprising result is that the *OV*, *SOV**, and *SO*OV* models explain the effect of type. As DSMs do not represent this concept explicitly, a possible interpretation suggested by our results is that type and typicality are not distinct categories, but capture properties of predicate-argument combinations at different granularity levels.

Distributional models can account for types because they emerge from the observed corpus distributions. Specifically, for the aspectual verbs used in the present data set, the distribution over their objects – namely that event nouns occur much more frequently that object nouns (Zarcone et al., 2013) – corresponds more naturally to an interpretation in terms of types than of typicality. A compositional distributional model where semantic types emerge as patterns of behavior has the advantage of relying on minimal assumptions regarding the granularity of the type ontology, which is intriguing, as pattern recognition is a key aspect of human cognition (Rumelhart and McClelland, 1987; Saffran et al., 1996; Tomasello, 2009).

In conclusion, the picture that emerges from our experiments is one where (1) expectations for predicate-argument combinations have a hierarchical structure, with types as a high-level distinction and typicality as a low-level distinction, (2) both levels are different, but interact early during processing, influencing reading times, and (3) both type and typicality can emerge from the "same same" distributional model.

Acknowledgments

This research was funded by the German Research Foundation (DFG) as part of SFB 732 "Incremental Specification in Context" and SFB 1102 "Information Density and Linguistic Encoding".

References

Marco Baroni and Alessandro Lenci. 2010. Distributional Memory: a general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

Klinton Bicknell, Jeffrey L Elman, Mary Hare, Ken McRae, and Marta Kutas. 2010. Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63:489–505.

Stefan Evert. 2005. *The statistics of word cooccurrences*. Ph.D. thesis, Universität Stuttgart.

Argyro Katsika, David Braze, Ashwini Deo, and Maria Mercedes Piñango. 2012. Complement coercion: Distinguishing between type-shifting and pragmatic inferencing. *The Mental Lexicon*, 7(1):58–76.

Alessandro Lenci. 2011. Composing and updating verb argument expectations: A distributional semantic model. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, pages 58–66, Portland, OR.

Kazunaga Matsuki, Tracy Chow, Mary Hare, Jeffrey L Elman, Christoph Scheepers, and Ken McRae. 2011. Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4):913–934.

Liina Pylkkänen and Brian McElree. 2006. The syntax-semantics interface: On-line composition of sentence meaning. In M. Traxler and M. A. Gernsbacher, editors, *Handbook of Psycholinguistics*, pages 539–579. Elsevier, Amsterdam, The Netherlands, 2nd edition.

David E. Rumelhart and James L. McClelland. 1987. Learning the past tenses of English verbs. Implicit rules or parallel distributed processing. In *Mechanisms of language acquisition*, pages 249–308. Lawrence Erlbaum Associates, Hillsdale, NJ.

Jenny R Saffran, Richard N Aslin, and Elissa L Newport. 1996. Statistical learning by 8-month-old infants. *Science*, 274(5294):1926–1928.

Michael Tomasello. 2009. *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press, Cambridge, MA.

Alessandra Zarcone, Jason Utt, and Sebastian Padó. 2012. Modeling covert event retrieval in logical metonymy: probabilistic and distributional accounts. In *Proceedings of the 3rd Workshop on Cognitive Modeling and Computational Linguistics*, pages 70–79, Montréal, Canada.

Alessandra Zarcone, Alessandro Lenci, Sebastian Padó, and Jason Utt. 2013. Fitting, not clashing! a distributional semantic model of logical metonymy. In *Proceedings of the 10th International Conference on Computational Semantics*, Potsdam, Germany.

Alessandra Zarcone, Alessandro Lenci, Ken McRae, and Sebastian Padó. in preparation. Type and thematic fit in logical metonymy resolution.

Identifying Existing and Novel Compound Words in Reading Finnish: An Eye Movement Study

Jukka Hyönä
Department of Psychology
University of Turku
Turku, Finland
hyona@utu.fi

Minna Koski
Department of Psychology
University of Turku
Turku, Finland
mmjkos@utu.fi

Alexander Pollatsek
Department of Psychology
University of Massachusetts
Amherst, MA, USA
pollatsek@psych.umass.edu

1 Introduction

According to the dual-route race model of compound word identification (Pollatsek, Hyönä, & Bertram, 2000), the holistic route and the morphological decomposition route operate in tandem. Bertram and Hyönä (2003) posited that word length modulates the interplay between the two access routes. When a compound word is sufficiently short so that all or most of its letters fall on the foveal region when fixating it during reading, the holistic route gets a head start and completes faster than the morphological route and thus the word is more likely to be identified as a whole. On the other hand, when a compound word is so long that a subset of letters is beyond foveal reach, the identification is initiated by first recognizing the initial constituent followed by the recognition of the second constituent and that of the whole word.

In their study examining the processing of novel compound words, Pollatsek et al. (2011) demonstrated that the decomposition route played even a more prominent role in processing novel than lexicalized compound words. Pollatsek et al. (2011) compared the processing of novel and existing Finnish compound words by manipulating the frequency of first constituent as an independent word, separately for long (average length of 13 letters) existing and novel compound words. The length of the first constituent as well as the frequency of the second constituent was matched across conditions. For first fixation duration, which indexes early effects in word processing, an effect of first-constituent frequency was observed that was similar in size for existing and novel compound words. For gaze duration (i.e. the summed duration of fixations made on the word before exiting to the right or left) first-constituent frequency was greater for novel than existing compound words. For the latest stages of processing during the first-pass reading, indexed by fixation time spent on the target word after fixating away from the first constituent but before exiting the word, only a main effect of novelty was observed. As regards to the processing of long novel compound words, the pattern of results was taken to suggest a two-stage process. During the first stage, lexical access is achieved for the compound word constituents. During the second stage, the meaning of the novel compound word is composed out of the constituent meanings. The second stage is assumed to take longer when the frequency of the first constituent is low, because the prototypical relationships that the low-frequency first constituent would be engaged in compounding are not firmly established.

In the present study, we further investigated the processing of novel and lexicalized Finnish two-noun compound words. This time we manipulated the frequency of the second constituent (the compound head). It was done separately for existing and novel compound words. Moreover, we also manipulated the length of the compound words. If indeed word length strongly determines the interplay between the holistic and decomposition route in compound word identification, as argued by Bertram and Hyönä (2003), the manipulation of the second-constituent frequency tapping into the decomposition process should result in different types of processing especially for short existing versus novel compound words. Short existing compound words are more likely be identified by the holistic route, whereas short novel compound words have to be processed via the morphological decomposition route. For long compound words, on the other hand, the manipulation of the second-constituent frequency should lead to less dramatic differences between existing and novel compounds, as the decomposition

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

route is assumed to be in operation for both word types.

Adult readers read sentences silently for comprehension while their eye movements were registered. The target compound words were embedded somewhere in the middle of the sentences. The frequency of the second constituent as a separate word was manipulated for short (7-9 letters) lexicalized (e.g., *savukala* = smoked fish) and novel (e.g. *hymykisa* = smile contest) compounds as well as for long (12-16 letters) lexicalized (e.g., *hiekkapaperi* = sand paper) and novel (e.g., *skandaalivaali* = scandal election) compound words. Thus, the experimental design was a 2 (low vs. high frequency second constituent) x 2 word type (existing vs. novel) x 2 word length (short vs. long) within-participants design. Comprehensibility of the novel compound words was secured by a rating test conducted prior to the experiment proper. Only novel compound words whose meaning could be computed without providing any linguistic context were chosen for the study. The frequency of the first constituent was matched across the conditions, as was the frequency of the short and long existing compound words.

2 Results

Several eye fixation measures were used to tap into the time course of compound word processing. The earliest effects were measured by first fixation duration. Early, but less immediate effects were measured by second fixation duration and gaze duration. Still later effects were measured by total fixation time, which is the sum of all fixations, both first-pass and second-pass, made on the target word.

First fixation duration: In the earliest stages of foveal word processing, indexed by first fixation duration, no effects of novelty or second-constituent frequency were observed.

Second fixation duration: A bit later in the processing timeline, main effects of novelty and second-constituent frequency were obtained. These effects were modified by interactions involving word length, including the three-way interaction. This interaction was broken down by computing a separate 2x2 ANOVA for short and long compounds, respectively. These analyses revealed no effect of novelty or second-constituent frequency for long compounds, whereas for short compounds both main effects and their interaction proved significant. The interaction reflected the fact that the second-constituent frequency effect was only observed for short novel compounds.

Gaze duration: In gaze duration, summing up all fixations made during the first-pass reading, the main effect of word type, word length and second-constituent frequency were all significant. Gaze duration was significantly longer for novel than existing words, longer for long than short words, and longer for compounds containing a low-frequency than high-frequency second constituent. Similarly to second fixation duration, gaze duration also revealed a reliable three-way interaction between the manipulated factors. In order to examine in more detail the interaction, it was broken down into two separate 2x2 ANOVAs, one for the short and another for the long compound words.

For the long compound words, there was a main effect of word type and second constituent frequency, but no reliable interaction between them, suggesting that the second-constituent frequency effect was of similar magnitude for existing (an effect size of 91 ms) and novel (an effect size of 111 ms) compound words. However, for short words, the Word Type x Second-Constituent Frequency proved significant. This interaction reflected the fact that the second-constituent frequency effect was considerably greater for novel (an effect size of 155 ms) than for existing (an effect size of 42 ms) compound words.

Total fixation time: We also analyzed the total fixation time spent reading the target words. This measure indexes late effects; it sums up the duration of all fixations made on the word during the first-pass and second-pass reading. In this measure, the three-way interaction obtained for second fixation duration and gaze duration was no longer significant. However, the interaction between word type and second-constituent frequency was almost significant. This interaction reflects the fact that in total fixation time the effect of second constituent frequency was greater for novel than existing compound words, regardless of word length. The size of the second-constituent frequency effect was 51 ms for the existing compounds and 151 ms for the novel compounds.

Summary of results: The following picture emerges from the pattern of results presented above. In the earliest stages of word processing, no signs of either novelty or second-constituent frequency were seen, which suggests that these effects took some time to develop during compound word identification. For long compounds, these effects were still absent in second fixation duration but emerged in gaze duration. For short compounds, the effects were already visible in second fixation duration. Finally, the measure indexing second-pass reading demonstrated a greater second-constituent effect for novel than existing compounds. All in all, the pattern of results suggests that meaning composition takes place with more delay for long than short compound words.

3 Conclusions

The present study provided further evidence for the view (Bertram and Hyönä, 2003, 2013), according to which word length modifies the relative role of the holistic versus the morphological decomposition route in compound word identification. The decomposition route is an integral part in identifying long compound words, because holistic processing is not viable due to visual acuity constraints. This became apparent in the effect of the second constituent frequency indexing access via morphological constituents being similar in magnitude for the novel and lexicalized compound words. On the other hand, when lexical access via the holistic route is a viable option, as is the case with short existing compound words that fit in the foveal area of the eyes when the word is fixated, the novelty effect emerged relatively early (during second fixation) and the second-constituent frequency effect was considerably smaller for existing than novel compound words during first-pass reading. Finally, the second-pass reading measure demonstrated a greater effect of constituent frequency for novel than lexicalized compounds. This may be taken to suggest that meaning composition takes longer when the frequency of the second constituent is low, since the typical relationships low-frequency constituents are engaged in compounds are less firmly established.

There are also two findings that are not completely in line with the visual acuity principle proposed by Bertram and Hyönä (2003). One is the absence of an early novelty effect for short compound words. If the holistic route is immediately activated when making the first fixation on the word, there should have been a novelty effect in first fixation duration. Second, there was a 42 ms effect of second constituent frequency in gaze duration even for existing short compound words, suggesting that the decomposition route also becomes active when identifying short lexicalized compounds.

A possible theoretical framework that can account for the obtained pattern of results is a dual route cascade model assuming that identification always starts out with the decomposition route with the process quickly cascading into the holistic access in the case of short compounds and with some delay in the case of long compounds. When a compound word is short, its constituents are also short and may be accessed rapidly. On the other hand, when the word is long, not only the constituents are likely to be longer and may thus lengthen their access, but the morphological segmentation process may also need additional time. Hence, the holistic route is activated with some delay after the decomposition route is activated. The suggested model may be further tested by replicating the present study by manipulating the frequency of first constituent separately for long and short, novel and lexicalized compound words.

References

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of Memory and Language, 48*, 615-634.

Bertram, R., & Hyönä, J. (2013). The role of hyphen at the constituent boundary in compound word identification: Facilitative for long, detrimental for short compound words. *Experimental Psychology, 60*, 157-163.

Pollatsek, A., Bertram, R., & Hyönä, J. (2011). Processing novel and lexicalised Finnish compound words. *Journal of Cognitive Psychology*, *23*, 795-810.

Pollatsek, A., Hyönä, J., & Bertram (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 820-833.

Electrophysiological correlates of idiom comprehension: semantic composition does not follow lexical retrieval

Paolo Canal[a,b], Francesca Pesciarelli[a], Francesco Vespignani[c], Nicola Molinaro[d,e] & Cristina Cacciari[a]

a Department of Biomedical Sciences, Università degli Studi di Modena e Reggio Emilia, Italy
b NEtS Center for Neurocognition Epistemology and Theorethical Syntax, IUSS, Pavia, Italy
e Ikerbasque, Basque Foundation for Science, Bilbao, 48001, Spain
c Department of Cognition and Formation Sciences, Università degli Studi di Trento, Italy
d BCBL, Basque center on Cognition, Brain and Language, Donostia/San Sebastian, Spain

1 Introduction

Idiomatic expressions, such as *break the ice*, are pervasive in everyday communication. They are frequently co-occurring sequences of words with a *conventional* meaning that is not derived from word-by-word semantic composition, but rather can be retrieved as such from semantic memory. Idioms are often read faster compared to literal sentences [e.g., Siyanova-Chanturia et al., 2011] and also lexical decision times are faster on idiom related words than on literal related targets [e.g., Cacciari & Tabossi, 1988]. Recent EEG data further suggest that semantic composition processes of idiomatic constituents might be not fully engaged during comprehension [Rommers et al, 2013]. Finally brain-imaging studies reported stronger and more widespread activation of the language network when reading idioms compared to non-idiomatic sentences [Zempleni et al., 2007; Lauro et al., 2008; Boulenger et al., 2009], suggesting that idiom comprehension might involve more cognitive resources. From these fragmented results, it is not clear yet how idiomatic semantic processing differs from literal semantic processing and this might be due to the paradoxical nature of idioms [e.g., Libben & Titone, 2008], which seem to be at the same time amenable of direct memory retrieval and word-by-word compositional analysis.

The two main questions of the present research thus concern two aspects of idiom comprehension: one relates to how the meaning of the whole is retrieved and integrated in the sentence representation; the second relates to what happens to word-by-word semantic composition of the literal meanings of the expression: is it carried out or suspended? To answer these questions we used EEG measures (with the analysis of Event-Related Potentials and oscillatory dynamics of Time-Frequency representations) because of their temporal precision [e.g., Luck, 2014], and because of the possibility of disentangling between memory juggling between memory retrieval and semantic integration processes [e.g., Hoecks & Brower, 2014].

2 The present Study

We carried out two Experiments in which short and literally plausible idioms (e.g., *break the ice*), i.e. having a literal well-formed meaning and a conventional meaning, were embedded in literal or idiomatic contexts. Notably, materials were designed in such way that the sentential context would constrain expectations on the upcoming target words to a similar extent across conditions. By doing so we minimized the impact of differential sentence constraints, known to elicit N400 effects, and we carried out a comparison between sentences that were semantically well-formed and for which contextual expectations on upcoming words were always fulfilled. Experiment 1 used EEG measures as dependant variable to investigate the time course of idioms comprehension and was followed up by Experiment 2 in which a cross modal priming paradigm was implemented, in order to confirm the activation of the literal meaning of the idiomatic constituents in both types of contexts.

On the basis of the previous ERP literature we hypothesized that meaning retrieval processes would affect the N400 component [e.g., Federmeier, 2007]: more demanding retrieval processes should be associated to larger N400 effects. The debate about the role of the N400 in semantic integration vs. retrieval mechanisms [see semantic unification processes in Hagoort & Van Berkum, 2007] makes it hard to exclude that the N400 component is not associated with the semantic integration of the meaning of the whole; however, given the available evidence on figurative language processing, we could also expect an effect on later occurring positivities, previously associated with metaphor (Late Positive Complex, LPC) [e.g., Coulson & Van Petten, 2002; Lai et al., 2009] or irony (P600) [Regel et

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

al., 2010] processing, or semantic pragmatic reanalysis (frontal Post-N400 Positivity) [e.g., Van Berkum et al., 2009; Molinaro et al., 2012]. Another result that has been previously reported in the ERP literature of idioms processing is the finding of an involvement of the P300 component. The P300 is generally associated with cognitive mechanisms of context update [Donchin & Coles, 1988] or context closure [Verleger, 1988]: Vespignani et al. (2010) found that the brain's electrical response to the correct idiom constituent was different if recorded before or after the idiom recognition point (RP, e.g., *prendere il toro per le*$_{\mathrm{RP}}$ *... corna -- take the bull by the*$_{\mathrm{RP}}$ *... horns*). The match to the correct idiom word was associated with an N400 reduction before recognition, but the electrophysiological response led to a P300 effect after the recognition of the idiom. Such effect would mirror a qualitative change in readers' expectations about upcoming words, after the expression has been recognized. We also expected to replicate Rommers et al (2013) results in the time-frequency domain of the EEG. The authors observed a power increase in the upper gamma frequency band after the presentation of the expected target words in literal but not in idiomatic contexts, supporting the hypothesis that semantic unification mechanisms are less engaged in idioms comprehension.

3 Method

3.1 Participants

380 students at Università degli studi di Modena e Reggio Emilia participated to the study set up to norm the experimental materials. 32 different students took part in Experiment 1. 42 students volunteered in experiment 2.

3.2 Materials

Experiment 1 materials were 90 idiomatic expressions of similar structure (VP+NP idioms) embedded in sentences. Idioms were selected for being highly Familiar and correctly paraphrased. Three sentential contexts for each expression were created so that the last word of the expression was highly predictable in the three contexts (above 85% cloze probability). ERPs were time-locked to the presentation of the first word of the expression (W1), and epochs comprising W1, W2 and W3 were extracted from the EEG. In Experiment 2 a subset of 44 idioms was used.

1a) La maestra aveva notato che Nicola disturbava i compagni, ma la prima volta ***chiuse un occhio*** *e continuò la lezione.*
(The teacher saw Nick was bothering his desk mate but for the first time she closed an eye (turned a blind eye) and kept on teaching.)

1b) Alla visita oculistica Enrico, prima di leggere le lettere indicate sulla lavagna luminosa, ***chiuse un occhio*** *per valutare la miopia.*
(At the Ophthalmological visit, before starting to read the letters on the panel aloud Henry closed an eye in order to evaluate his nearsightedness.)

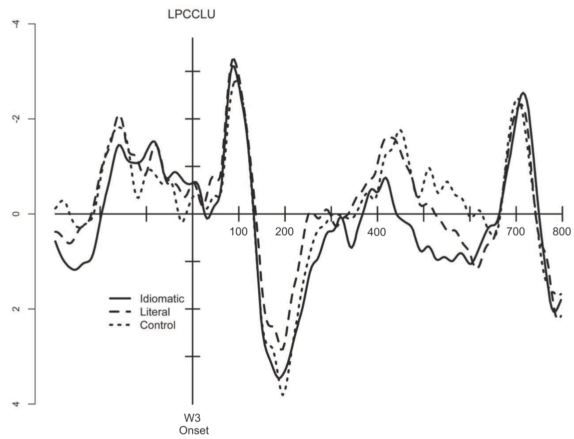
1c) Giovanni ha rotto gli occhiali durante la rissa perché ha preso un pugno in un ***occhio*** *e gli sono caduti a terra.*
(Jack broke his glasses during the fight because got a punch in his eye and fell on the ground.)

3.3 Procedure

In Experiment 1 sentences were presented word-by-word at the centre of the screen (SOA=600ms). In Experiment 2, contexts sentences were auditorily presented via headphones until the last word of the expression. Targets that could be related or unrelated to the literal meaning of the last word of the expression were visually presented at the offset of the audio file.

4 Results

Fig.1 Grand Average ERPs from a pool of 7 frontal electrodes (AF3, AF4, F3, FZ, F4, FC1, FC2) in which frontal PNP effects are usually reported (negative voltage is plotted upwards). Idiomatic condition (solid line), Literal condition (dashed line) and Control condition (dotted line) are compared at the onset of the last word of the idiomatic



expression (e.g., *ice*).

Experiment 1 showed that:

- No N400 differences emerged between literal and idiomatic context, during the processing of the three constituent words.
- Differences between Idiomatic vs. Literal, and Idiomatic vs. Control conditions emerged during the presentation of the last word of the expression (e.g., *ice*), and occurred in the 400 to 600 ms time interval.
- Consistently with Rommers et al (2013) study, the Time-Frequency analysis of the EEG revealed power differences in the higher gamma frequency band (60-80Hz) between expressions embedded in literal vs. idiomatic contexts: no power increase was associated with the idiomatic condition.

Experiment 2 showed that:

- Target words related to the literal meaning of the idiomatic constituents obtained faster lexical decision times with respect to unrelated targets, regardless of type of context.

5 Discussion

Concerning the question related to how the meaning of the whole idiom is integrated in the sentence representation, our results suggest that integration mechanisms occur only upon presentation of the last constituent word, when the idiomatic expression has very likely been recognized. On the last constituent, ERP differences between idiomatic and literal contexts emerged between 400 and 600 ms in frontal electrodes. The timing and scalp distribution of the effect suggest that it affected a positive component (the frontal Post-N400 Positivity) occurring soon after the peak of the N400. These results could be accommodated elaborating the framework proposed by the Retrieval-Integration hypothesis [Hoecks & Brower, 2014], which holds that semantic - pragmatic integration processes are reflected in P600 like positivities. One possible interpretation is that the observed frontal positive shift might be part of a larger family of positive components reflecting the engagement of a semantic/pragmatic wrap-up mechanism that is performed at end of the expression to assign a full interpretation to the incoming input.

Concerning the second experimental question related to the composition of individual constituent words we argue that Experiment 2 showed that the literal meaning of the last word of the expression was at least accessed, and confirms other evidence supporting the idea that readers process the literal meaning of idiomatic constituents (Boulenger, Shtyrov & Pulvermüller, 2012). Moreover, the lack of N400 differences across conditions and word positions, suggests that lexical retrieval processes similarly occurred in literal and idiomatic contexts. However, the analysis of the frequency domain replicated Rommers et al's findings of a larger power increase in the high gamma frequency band for literal compared to idiomatic contexts, which, consistently with their interpretation, could signal that word-by-word composition mechanisms are less engaged in idioms comprehension.

Conclusions

When presented with idiomatic expressions readers retrieve the literal meaning of the constituent words. However, word-by-word semantic composition mechanisms are idling, and, only at the end of the expression, a semantic/pragmatic wrap-up of the idiom is carried out to update the sentence representation.

Reference

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *27*(2), 251–272.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683.

Zempleni, M.-Z., Haverkort, M., Renken, R., & A. Stowe, L. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *NeuroImage*, *34*(3), 1280–1291.

Lauro, L. J. R., Tettamanti, M., Cappa, S. F., & Papagno, C. (2008). Idiom Comprehension: A Prefrontal Task? *Cerebral Cortex*, *18*(1), 162–170.

Boulenger, V., Hauk, O., & Pulvermüller, F. (2009). Grasping Ideas with the Motor System: Semantic Somatotopy in Idiom Comprehension. *Cerebral Cortex*, *19*(8), 1905–1914.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2012). Context-dependent Semantic Processing in the Human Brain: Evidence from Idiom Comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6),1103–1121.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Hoeks, J. C. J. and Brouwer, H. (2014). Electrophysiological Research on Conversation and Discourse Processing. In: Holtgraves, T. (Ed.), Oxford Handbook of Language and Social Psychology, pp. 365-386. New York: Oxford University Press.

Hagoort, P., & Berkum, J. van. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801–811.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491–505.

Coulson, S., & Petten, C. V. (2002). Conceptual integration and metaphor: An event-related potential study. *Memory & Cognition*, *30*(6), 958–968.

Lai, V. T., Curran, T., & Menn, L. (2009). Comprehending conventional and novel metaphors: An ERP study. *Brain Research*, *1284*, 145–155.

Regel, S., Gunter, T. C., & Friederici, A. D. (2010). Isn't It Ironic? An Electrophysiological Exploration of Figurative Language Processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293.

Berkum, J. J. A. V., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or Wrong? The Brain's Fast Response to Morally Objectionable Statements. *Psychological Science*, *20*(9), 1092–1099.

Molinaro, N., Carreiras, M., & Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *NeuroImage*, *59*(4), 3488–3501.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? Behavioral and Brain Sciences, 11(03), 357–374.

Verleger, R. (1988).Event-related potentials and cognition: A critique of the context-updating hypothesis and an alternative interpretation of the P300.Behavioral and Brain Sciences,11, 343–427.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive Mechanisms in Idiom Comprehension. Journal of Cognitive Neuroscience, 22(8), 1682–1700.

Boulenger, V., Shtyrov, Y., & Pulvermüller, F. (2012). When do you grasp the idea? MEG evidence for instantaneous idiom understanding. NeuroImage, 59(4), 3502–3513.

Metaphorical priming in a lexical decision task in high functioning autism

Chahboun[1], S., Vulchanov[1], V., Saldaña[2], D., Eshuis, R.[1] and Vulchanova[1], M.

[1]Language Acquisition and Language Processing Lab, NTNU, Norway.

[2]Departamento de Psicología Evolutiva, University of Seville, Spain.

1 Introduction:

The difficulties experienced by autistic individuals with regard to communication and language are widely known and well documented. Individuals with High functioning autism (ASD) are distinguished by relative preservation of linguistic and cognitive skills. However, problems with pragmatic language skills have been consistently reported across the autistic spectrum, even when structural language is intact. Many studies establish failure to understand metaphors, idioms and other forms of figurative language (Gold & Faust, 2010; Vulchanova, Talcott, Vulchanov & Stankova, 2012). Figurative language takes many forms, conceptual metaphors being one of the most common. On the cognitive level, conceptual metaphors are the mental representations we establish in order to map between two domains (Lakoff & Johnson 1980; Fauconnier 1985; Vulchanova, Saldaña, Chahboun & Vulchanov 2015). In other words, the logic of one conceptual domain is applied to another.

Several studies have shown impaired figurative language in ASD populations. One of the first studies in figurative language in autism for instance was that of Happé (1995). She used 3 types of expressions: synonyms, similes, and metaphors. The underlying assumption of this study is that, in order to understand these kinds of expressions, we need to be able to "decode" the intentions and ideas of person to whom we are talking. The findings from this study showed that metaphor comprehension is impaired in individuals with autism.

Our hypothesis in this study is that this deficient metaphorical ability might depend, not only in the type of figurative expression (regarding the novelty or conventionality of it), but also on the way these expressions are perceived. This is especially relevant for individuals with ASD who need specific ways of integrating inputs, such as the ways in which the type of instruction can drastically change the reading comprehension in this population (Micai, Vulchanova & Saldaña 2015). In the current study, we test responses to metaphorical expressions and whether or not metaphors solicit priming for literal or rather the appropriate figurative interpretation in high-functioning children and adolescents with ASD.

These tests are carried out through a cross modal priming task. Priming is a process occurring outside conscious awareness, and thus differs from direct retrieval. It is an effect of retrieval from implicit memory, creating a heightened sensitivity to certain stimuli. In general, priming effects are found between lexical items which share a semantic component or a semantic association. For example, *angel* is recognized quicker, if it is followed by *wings* than,

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

say, *table*. Here we exploit priming to reveal how metaphorical expressions are associated with figurative as opposed to literal interpretations in individuals with ASD. We are also interested in whether or not the modality of presentation of the stimuli (auditory vs. written) has an effect on their processing, as already established in on-going research (Chahboun, Vulchanov, Saldaña & Vulchanova, 2015).

2 Method:

2.1 Participants:

Two age groups of high-functioning ASD participants (N=48) and controls (N=39) were included (all native speakers of Spanish),each group has 2 age ranges

- **Group 1**: Age range 10-12.

Control group (N=18) and ASD group (N=26).

- **Group 2:** Age range 16-20.

Control group (N=21) and ASD group (N=22).

Participants and their legal tutors (usually the parents) provided written consent for entry into the study. Most of the individuals had participated in an earlier study (Chahboun et al 2015).

The diagnosis of ASD was confirmed according to the Autism Diagnostic Observation Schedule (ADOS) and also with the Autism Quotient (AQ).

We also made sure the participants do not have any structural language deficit. In addition to measuring the general IQ with the Weschler Scale (WISC IV or WAIS) we measured the participants' receptive vocabulary (British Picture Vocabulary Scale), their grammatical language level (CEG: Test of comprehension of grammatical structures) and theory of mind.

2.2 Apparatus and Stimuli:

Stimuli were displayed on a color monitor controlled by E-prime software implemented on a Dell compatible laptop.

Responses were collected with a response box; response accuracy (ACC) and reaction times (RTs) were measured by the E-prime software. The stimuli included 36 prime expressions classified into 3 different types: novel metaphors, conventional metaphors and free combinations (non- metaphorical expressions), all comprising a noun and a modifier. The target words were semantically related to the prime expressions. On half of the instances for each group of expressions, targets were related to the figurative interpretation of the prime, the remaining half were related to the literal meaning (cf. Figure1.)

In a pilot study with 150 adult native speakers of Spanish, we determined the degree of familiarity of the metaphors. This allowed us to verify the conventionality of the metaphors or their novelty, and their inclusion in the test stimuli. The same number of filler expressions (N=36) were added, respectively as primes, and non-words served as targets. Thus, each participant responded in total to 72 trials, 36 in each modality: visual modality (stimuli presented orthographically) and auditory (stimuli presented auditorily). The experiment was designed as a lexical decision task on the target word.

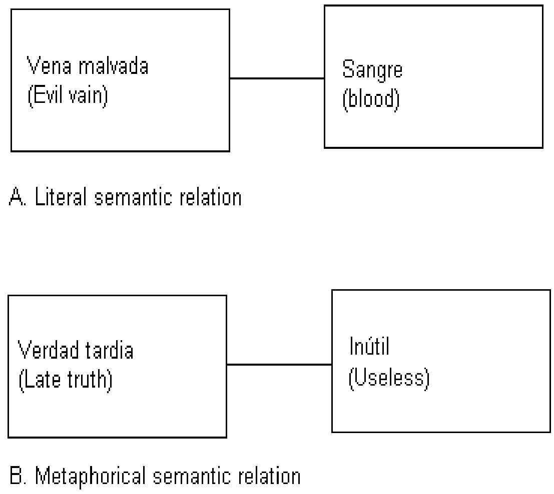


Figure 1: Examples of the targets semantically related used: Literal or metaphorical relation

2.3 Procedure:

Each participant was tested individually in a single session. Participants either saw the prime expression on a computer screen or heard it via loud-speakers. The timing of the specific stimulus events on each trial was as follows: (1) The prime is presented as visual text on the screen or auditorily via the loud-speakers (depending on the experimental block); (2) a fixation point is presented followed by a delay of 400 ms as a latency; (3) a target is presented as word or non-word; (4) Finally, participants have to decide whether the target is a word or not in Spanish (cf. Figure 2.).

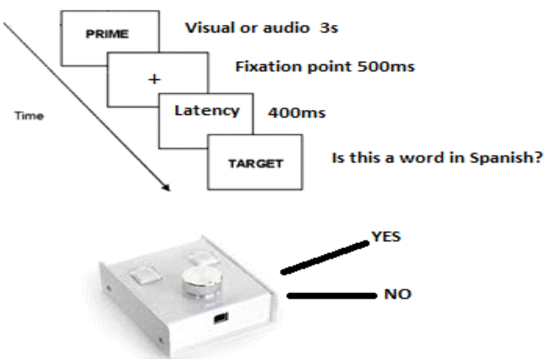


Figure 2: Sequence of events for the trials of the experiment.

3 Preliminary results:

The data of both the control and experimental group (N=19) were analysed with R. A linear mixed model analysis on RTs revealed a significant interaction between presentation modality and conventionality of the metaphors ($p<.05$), with poorer performance of the ASD group when the prime was presented auditorily. Furthermore, there was an interaction between group and age, with younger groups taking more time to respond. Finally, the results showed a significant interaction between modality, type of target and age. The younger groups' performance was slower when the prime was presented auditorily, and when the target relationship with the prime was figurative. Regarding accuracy, with a generalized linear mixed model (R) we found significant interactions depending on the modality of the prime. The ASD groups were less accurate in the auditory modality, in contrast with the control groups. Moreover, the results show a significant interaction between conventional metaphors and age in both groups. There was a significant interaction between the type of target, modality and age. Finally, a main effect of group, a main effect of age and an interaction of age and group were observed The typically developing participants were more accurate in both age ranges. In both the experimental and the control group, the older participants performed better than the younger ones, and the difference in performance between the age ranges in the ASD group was greater than in the control group.

4 Conclusions:

Most of the available literature and previous studies, using a range of different methodologies, consistently demonstrate that figurative language is demanding for ASD populations.

In particular, metaphors present a difficulty in terms of processing for the ASD group. The preliminary results of this study confirm our earlier findings that the auditory modality is more demanding for the ASD group. Surprisingly, the significant effect we found for accuracy was confined to the conventional metaphors. An explanation can be sought in the difference between conventional and novel metaphors.

Conventional metaphors are less transparent, making them more problematic compared to novel metaphors, as these might be processed without the need for prior familiarity.

These results support the findings in Chahboun et al (2015), where a similar effect was found for idioms contra novel metaphors. Idioms are similar to conventional metaphors in that both types of expression are less transparent than both literal expressions and novel metaphors.

Acknowledgments:

This project has received funding from the European Union's Seventh Framework Programme for research; technological development and demonstration under grant agreement no 316748.

References:

Chahboun, S., Vulchanov, V., Saldaña, D., Eshuis, H. & Vulchanova, M. (2015). Contextual and visual cues in the interpretation of idioms in high functioning autism. Manuscript in preparation.

Fauconnier, G. (1985). *Mental spaces*. Cambridge, MA: MIT Press.

Gold, R. & Faust, M. (2010). Right Hemisphere Dysfunction and Metaphor Comprehension in Young Adults with Asperger Syndrome. *Journal of Autism and Developmental Disorders*, 40, 7, 800-811.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.

Micai, M., Vulchanova, M., & Saldaña, D. (2015). Can targeted instructions change the reading strategy of children with ASD? An eye link study. Manuscript in preparation.

Vulchanova, M., Talcott, J., Vulchanov, V., and Stankova, M. (2012). Language against the odds, or rather not: The Weak Central Coherence hypothesis and language. *Journal of Neurolinguistics, 25,* 1,13-30.

Vulchanova, M., Saldaña, D., Chahboun, S., & Vulchanov, V. (2015): Figurative language processing in atypical populations: The ASD perspective. *Frontiers in Human Neuroscience*, special issue *The Metaphorical Brain*. doi:10.3389/fnhum.2015.00024.

ERP correlates of letter-case in visual word recognition

Barbara Leone-Fernandez
Departamento de Metodología, Facultad de Psicología
Av. Blasco Ibáñez, 21, Universitat de València, 46010, Spain
barbara.leone@uv.es

Manuel Perea
Departamento de Metodología, Facultad de Psicología
Av. Blasco Ibáñez, 21, Universitat de València, 46010, Spain
manuel.perea@uv.es

Marta Vergara-Martínez
Departamento de Psicología Evolutiva, Facultad de Psicología
Av. Blasco Ibáñez, 21, Universitat de València, 46010, Spain
marta.vergara@uv.es

1 Introduction

Visual word recognition is a key element of language comprehension. The vast majority of current models assume that the recognition of a printed word is based on the activation of abstract letter identity representations. The hierarchical neural accounts of letter/word recognition of Dehaene, Cohen, Sigman, and Vinckier (2005) and Grainger, Rey, and Dufau (2008) posit that, early in the process of lexical access, there are neuronal assemblies that respond to the word's case-specific letters (e.g., they respond to 'e' but not to 'E'). Later in processing, there are neuronal assemblies that respond to the abstract representation of the letter identity (e.g., they respond to the same degree to 'e' and to 'E').

Behavioral evidence using masked priming (i.e., a paradigm that taps onto early word processing; Forster & Davis, 1984; see Grainger, 2008, for review) has revealed that there is a rapid access to case-invariant letter representations. Specifically, the advantage of the identity condition over the unrelated condition is independent of the letter-case (similar advantage for kiss-KISS and EDGE-edge; see Bowers, Vigliocco, & Haan, 1998). Furthermore, response times to matched-case identical prime-target pairs (EDGE-EDGE) are virtually similar as the response times to mismatched-case identical prime-target pairs (edge-EDGE; see Jacobs, Grainger, & Ferrand, 1995; Perea, Jiménez, & Gómez, 2014).

To our knowledge, only a previous experiment investigated the temporal processing of letter-case using event-related potentials in an unmasked paradigm (Lien, Allen, & Crawford, 2012). Lien et al. compared the processing of lowercase-printed vs mIxEd-cAse-printed words of different frequency (high and low). They found that the N170 amplitude, related to structural encoding, was sensitive to case mixing, but the P3, related to stimulus categorization, was sensitive to lexicality and word frequency. They proposed that case mixing affects early processing stages of visual word recognition.

The Lien et al. experiment is important, but it does not respond to the question of whether letter-case plays a role during visual-word recognition with visually familiar words –note that mIxEd-cAsE stimuli are visually unfamiliar and difficult to process. In contrast, lowercase and uppercase words are the usual format when reading words. Indeed, experiments on visual-word recognition employ either lowercase or uppercase words with no explicit justification.
Importantly, there is one account that does assume that letter-case information may form an integral part of a word's lexical representation. Specifically, Peressotti, Cubelli, and Job (2003) claimed that 'while size, font and style (cursive or print) affect the visual shape of letters, the uppercase–lowercase distinction is abstract in nature as it is an intrinsic property of letters' (p. 108). In the framework of Peressotti et al.'s 'orthographic cue' account, a given lexical unit would not be retrieved only on the basis of the letter identity and letter position, but also on the basis of letter-case information. Given that most printed words are presented in lowercase, this should provide an advantage for the processing of lowercase vs. uppercase words (see Mayall & Humphreys, 1996; Perea & Rosa, 2002, for behavioral evidence of a lowercase advantage in visual-word recognition).

The main aim of this study is to examine the time course of letter-case on lexical access. The ERPs may help to disentangle whether letter case is an attribute that is only relevant in early perceptual

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

processing or whether it is also relevant in the retrieval of lexical representations. To attain this goal, we examined whether the effects of letter-case (lowercase vs. UPPERCASE) are modulated by word-frequency (a factor that indicates lexical/semantic activation; see Vergara-Martínez, Perea, Gómez, & Swaab, 2013) tracking the ERP waves in well-studied time windows (N/P150: 100-170 ms; P200: 170-250 ms; N400: 255-450 ms) in a lexical decision task.

2 Method

Twenty-two healthy, right-handed, native Spanish-speaking Valencia University students, naïve to the manipulation of the stimuli, participated in the study in exchange for a small gift.

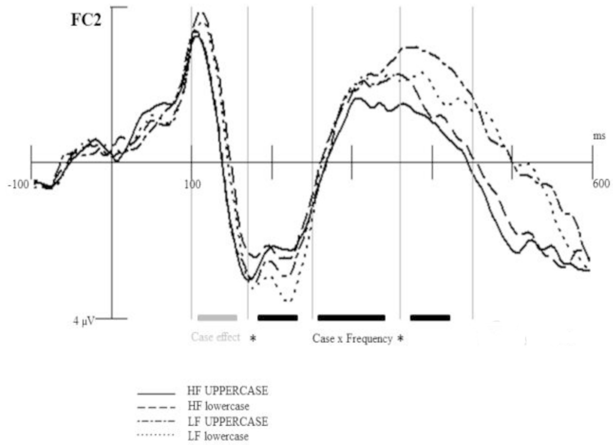
We selected a set of 160 words from the Web-accessible EsPal database (Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013). Half of the words were of high frequency and half were of low frequency. The two groups of words were matched in relevant psycholinguistic factors (length, orthographic neighborhood, concreteness, imageability…). Half of the words were presented in uppercase and half in lowercase (MOTHER; mother). In addition, a list of 160 pseudowords (half in lowercase, half in uppercase) was included for the purposes of the lexical decision task.

Participants were instructed to decide as accurately and rapidly as possible whether or not the stimulus was a Spanish word. They pressed one of two response buttons (YES/NO). The electroencephalogram (EEG) was recorded from 29 electrodes, averaged separately for each of the experimental conditions, each of the subjects and each of the electrode sites. For each time window, we conducted ANOVAs with word-frequency (high, low), case (lowercase, UPPERCASE), and AP (anterior, central-anterior, central, central-posterior and posterior) as factors in the design.

Results and Conclusions

The behavioral data revealed significantly faster responses for high-frequency than for low-frequency words (656 vs. 702 ms) and significantly faster responses for lowercase than for uppercase words (675 vs. 683 ms). There were no signs of an interaction between the two factors. The error data revealed the same pattern as the response time data.

In the N/P150, larger negative values were observed for lowercase than for uppercase words, with a central scalp distribution, whereas the effect of word-frequency was not significant. In the P200, and only for low-frequency words, larger positive values were observed for the lowercase than for uppercase words in frontal/central scalp areas. With respect to the N400, the ERP waves revealed a dissociation of the letter-case effect for low- and high-frequency words. High-frequency words showed an effect of letter-case in an early stage of the N400, whereas low-frequency words showed an effect of letter-case (in the opposite direction; see Figure 1) in a later stage of the N400.



As expected, there was an early pre-lexical effect of letter-case that did not interact with word-frequency. Importantly, we found an interaction between letter-case and word-frequency not only in the N400 time window –which is commonly associated to lexical-semantic processing, but also the P200 time window, thus supporting the hypothesis that letter-case may affect the mapping of visual-orthographic information onto word representations. Taken together, the present ERP data provide empirical support to the hypothesis that letter-case information may be stored in the abstract word representations (Peressotti et al., 2003), thus posing some problems for current computational and neural models of visual-word recognition.

"Figure 1. Grand average ERP waves to Frequency and case manipulations in one representative electrode. Different columns mark the four epochs under analysis"

References

Andrew Duchon, Manuel Perea, Nuria Sebastián-Gallés, Antonia Martí and Manuel Carreiras. 2013. EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods*, 45:1246–1258 doi:10.3758/s13428-013-0326-1

Arthur M. Jacobs, Jonathan Grainger and Ludovic Ferrand. 1995. The incremental priming technique: A method for determining within-condition priming effects, *Perception & Psychophysics*, 57:1101–1110. doi:10.3758/BF03208367

Francesca Peressotti, Roberto Cubelli and Remo Job. 2003. On recognizing proper names: The orthographic cue hypothesis. *Cognitive Psychology*, 47:87–116. doi:10.1016/S0010-0285(03)00004-5

Jeffrey S. Bowers, Gabriella Vigliocco and Richard Haan. 1998. Orthographic, phonological, and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance*, 24:1705–1719. doi:10.1037/0096-1523.24.6.1705

Jonathan Grainger, Arnaud Rey and Stéphane Dufau. 2008. Letter perception: from pixels to pandemonium. *Trends in Cognitive Sciences*, 12:381-387. doi:10.1016/j.tics.2008.06.006

Kate Mayall and Glyn W. Humphreys. 1996. Case mixing and the task-sensitive disruption of lexical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22:278–294. doi:10.1037/0278-7393.22.2.278

Kenneth I. Forster and Chris Davis. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 10:680–698. doi:10/1037/0278-7393.10.4.680

Manuel Perea and Eva Rosa. 2002. Does "whole word shape" play a role in visual word recognition? *Perception and Psychophysics*, 64:785–794. doi:10.3758/BF03194745

Manuel Perea, Maria Jiménez and Pablo Gomez. 2014. A challenging dissociation in masked identity priming with the lexical decision task. *Acta Psychologica*, 148:130–135. doi:10.1016/j.actpsy.2014.01.014

Marta Vergara-Martínez, Manuel Perea, Pablo Gómez and Tamara Y. Swaab. 2013. ERP correlates of letter identity and letter position are modulated by lexical-frequency. *Brain and Language*, 125: 11–27. doi: 10.1016/j.bandl.2012.12.009

Mei-Ching Lien, Philip A. Allen and Caitlin Crawford. 2012. Electrophysiological evidence of different loci for case-mixing and word-frequency effects in visual word recognition. *Psychonomic Bulletin & Review*, 19:677–684. doi: 10.3758/s13423-012-0251-9

Stanislas Dehaene, Laurent Cohen, Mariano Sigman and Fabien Vinckier. 2005. The neural code for written words: A proposal. *Trends in Cognitive Sciences*, 9:335–341. doi:10.1016/j.tics.2005.05.004

Morphotactic effects on the processing of Italian derivatives

Pier Marco Bertinetto
Scuola Normale Superiore
P.zza dei Cavalieri 7
Pisa, Italy
bertinetto@sns.it

Chiara Celata
Scuola Normale Superiore
P.zza dei Cavalieri 7
Pisa, Italy
celata@sns.it

Luigi Talamo
Università di Bergamo
Piazzetta Verzieri 1
Bergamo, Italy
luigi.talamo@unibg.it

Abstract

This paper investigates the processing of Italian affixed forms differing for morphotactic transparency. A lexical decision task with immediate priming was used. Following the principles of morphotactic transparency and Natural Morphology, the priming effect was hypothesized to be stronger for items with a higher degree of morphotactic transparency. However, the predictions were not totally met. The paper discusses possible explanations from the theoretical and methodological points of view, and highlights potential developments of the research.

1 Introduction

According to Dressler (1985, 2005), morphotactic transparency is one of the main parameters within the universal markedness theory of the so-called System-Independent Morphological Naturalness. It assumes the widespread existence of "opacifying obstructions" (Dressler, 2005: 272) in inflectional, derivational or compounding processes, and is expressed by preference degrees along a naturalness scale. The most natural forms are those without opacifying obstructions, followed by those based on mildly opacifying phonological processes (such as resyllabification), while allomorphic rules and suppletion are the most opaque and least natural morphological operations. In this approach, natural is synonymous with cognitively simple, iconic and therefore easy to acquire and process.

This work investigates the native speakers' processing of Italian affixed forms differing for morphotactic transparency. Derived forms pertaining to two different classes of morphotactic transparency but matching for length, average frequency, stress pattern, as well as morphosemantic transparency were used as immediate primes in a lexical decision task; the corresponding underived words were used as targets. Following the principles of morphotactic transparency and Natural Morphology, the priming effect was hypothesized to be stronger for items with a higher degree of morphotactic transparency.

2 Morphotactic Transparency

To date, the only database for morphotactic transparency of derivational processes is *derIvaTario*, an open-source annotated lexicon of about 11,000 Italian derivatives (<derivatario.sns.it>; see Talamo & Celata, 2011; Talamo et al., submitted). The lexical source of *derIvaTario* is CoLFIS, *Corpus e Lessico di Frequenza dell'Italiano Scritto* (Bertinetto et al. 2005), a fully lemmatized three-millions word corpus of written Italian, sampled out of a carefully balanced variety of books, journals and newspapers. CoLFIS was created with the purpose of representing the mental lexicon of the ideal Italian speaker – or, more exactly, reader – as reliably as possible (Laudanna et al., 1995).

derIvaTario takes into account several morphological properties of the base and of each affix involved in the derivational cycles, crucially including morphotactic and morphosemantic transparency (see Libben, 1998 and Dressler, 2005 for the latter). With respect to the former, *derIvaTario* provides a value according to the Universal Scale of Morphotactic Transparency (Dressler, 1985 and 2005). The scale values

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

range from mt1 to mt8, as shown in Appendix A. The items used in the present experiment belonged to two sets of derivatives, respectively characterized by full transparency (mt1) and relative opacity (mt4).

3 Experiment

3.1 Materials and methods

Adult native Italian speakers participated in a speeded lexical decision task with orthographic stimuli. 32 words and 32 nonwords functioned as targets. Each target (consisting of an underived word) was immediately preceded by a prime in three different conditions: morphological (e.g. *ribellione/ribelle*, 'rebellion/rebel'), identity (*ribelle/ribelle*) and unrelated (*xxxxxx/ribelle*). Participants saw each target in only one of the three conditions. The test items are listed in Appendix B.

All primes were morphosemantically fully transparent. Half of them were classified as mt1 according to *derIvaTario* (full transparency), the other half as mt4 (with intervening morphophonological opacifying process). The two groups were carefully balanced for: (a) average lexical frequencies of both primes and targets, (b) length of prime and target (as measured by N of phonemes and N of graphemes), and (c) type of base. The last point needs clarification. As is well-known, Italian morphology is not word-based, i.e. the base does not correspond to an actual word. Since *derIvaTario* assumes 7 base types, it was necessary to control for the possible effect of this variable. Only the two most frequent base types were used in the present experiment: (i) root, i.e. an underived word without inflectional ending (e.g. *bellezza* 'beauty' as based on the root *bell-* of *bello* 'beautiful.M.SG.'), (ii) verbal theme, i.e. a verb root plus the thematic vowel (e.g. *battimento* 'beat' as based on the verbal theme *batti-* of *battere* 'to beat'). These two base types were equally distributed within the two word sets: 11 verbal themes, 5 roots.

Nonwords were created by replacing one phoneme in real Italian derivatives and the corresponding underived words. They had the same average length as the test words.

The order of words and nonwords was randomized across participants. Before performing the task, the participants were trained on a list of 8 items (4 words, 4 nonwords).

The priming effect of the derivatives was assessed as the average RT difference between the morphological condition and the identity and unrelated conditions. A statistically significant interaction between priming condition (morphological, identity, unrelated) and morphotactic transparency (mt1 vs. mt4) would suggest that the morphotactic contrast is cognitively salient.

3.2 Results

Repeated measure ANOVAs were run with priming condition as within-subject factor and morphotactic transparency as between-subject factor. The mean results are shown in Table 1. Comparing the morphological and the unrelated conditions, mt1 primes facilitated target recognition to a larger extent than mt4 primes. Similarly, comparing the morphological condition with the identity condition, mt4 primes slowed down target recognition to a larger extent than mt1 primes. Although the general tendency was consistent with the experimental hypothesis, the interaction condition x morphotactic transparency was not significant (Pillai's trace F=0.547, $p > .05$). Thus, although the priming effect exerted by mt4 derivatives onto the corresponding underived words was weaker than the one yielded by mt1 derivatives, the current experiment does not support the initial hypothesis.

Table 1. Average reaction times and differential priming (ms) across conditions and transparency levels.

		identity		morphological		unrelated
mt1		491		547		631
	diff.		*56*		*84*	
mt4		502		573		637
	diff.		*71*		*64*	

4 Discussion

The purpose of this experiment was to investigate whether morphotactic transparency is a cognitively relevant factor in the processing of Italian base forms when primed by corresponding derivatives. A significant differential priming effect was expected between mt1 and mt4 primes, which would have lent support to the Universal Scale of Morphotactic Transparency as implemented by *derIvaTario*. The experiment, however, did not produce the expected result, despite encouraging tendencies.

A possible explanation for this result is the strictly on-line character of the technique used (immediate priming). As Laudanna et al. (2004) have shown for verbal inflection, the effect of complex morphological properties on the processing of isolated words is more likely to be detected in off-line techniques, such as free recall tasks, implying a short-term and/or episodic memory component.

In addition, the assumed difference between transparent and partially opaque derivatives in priming their base forms might surface to a larger extent when the morphological condition is compared with a phonological priming condition (e.g. *colazione/colare* 'breakfast/percolate'), in which no morphological relatedness is found between the prime and the target, although their formal relationship is the same as in a morphologically related pair (e.g. *formazione/formare* 'formation/form'). This hypothesis is currently under investigation.

References

Pier Marco Bertinetto, Cristina Burani, Alessandro Laudanna, Lucia Marconi, Daniela Ratti, Claudia Rolando, and Anna Maria Thornton. 2005. *Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS).* http://linguistica.sns.it/CoLFIS/Home.htm

Wolfgang U. Dressler 1985. On the Predictiveness of Natural Morphology. *Journal of Linguistics*, 21(2): 321–337.

Wolfgang U. Dressler 2005. Word-Formation in Natural Morphology. In: Štichauer P., Lieber R. (eds) *Handbook of Word-Formation*, Springer: 267–284.

Alessandro Laudanna, Simone Gazzellini, and Maria de Martino 2004. Representation of grammatical properties of Italian verbs in the mental lexicon. *Brain and Language* 90: 95–105.

Alessandro Laudanna, Anna Maria Thornton, Giorgina Brown, Cristina Burani, and Lucia Marconi 1995. Un corpus dell'italiano scritto contemporaneo dalla parte del ricevente. In: Bolasco S., Lebart L., Salem A. (eds) *III Giornate internazionali di Analisi Statistica dei Dati Testuali*, Roma: Cisu: 103–109.

Gary Libben 1998. Semantic Transparency in the Processing of Compounds: Consequences for Representation, Processing, and Impairment. *Brain and Language* 61: 30–44.

Luigi Talamo and Chiara Celata 2011. Toward a morphological analysis of the Italian lexicon: developing tools for a corpus-based approach. *Quaderni del Laboratorio di Linguistica della Scuola Normale Superiore di Pisa,* 10.

Luigi Talamo, Pier Marco Bertinetto, and Chiara Celata C. (submitted) DERIVATARIO: An annotated lexicon of Italian derivatives.

Appendix A. Scale of morphotactic transparency.

DEGREE	NATURE OF PHENOMENON
mt1	none
mt2	purely prosodic and phonological (e.g. resyllabification, assimilation)
mt3	phonological, with neutralization of phonetic constituents (e.g. flapping)
mt4	morpho-phonological, without loss of constituents (e.g. articulatory weakening)
mt5	morpho-phonological, with loss of constituents (e.g. deletion)
mt6	purely morphological (e.g. paradigmatic alternation of affixes)
mt7	lexical: weak suppletion
mt8	lexical: strong suppletion

Appendix B. Experimental words.

Mt1		Mt4	
Prime	Target	Prime	Target
disegnatore	*disegnare*	*traducibile*	*tradurre*
bruciatore	*bruciare*	*discutibile*	*discutere*
suggeritore	*suggerire*	*tessitura*	*tessere*
cancellazione	*cancellare*	*competitore*	*competere*
esclamazione	*esclamare*	*emettitore*	*emettere*
dominazione	*dominare*	*roditore*	*rodere*
nuotatore	*nuotare*	*scommettitore*	*scommettere*
accentuzione	*accentuare*	*perseguibile*	*perseguire*
bollitura	*bollire*	*godibile*	*godere*
piegatura	*piegare*	*cedimento*	*cedere*
fregatura	*fregare*	*spargimento*	*spargere*
intrusione	intruso	rassegnazione	rassegnato
perversione	perverso	concitazione	concitato
ribellione	ribelle	desolazione	desolato
introversione	introverso	discrezione	discreto
avversione	avverso	depravazione	deparavato

Are you reading what I am reading?
The impact of contrasting alphabetic scripts on reading English

Tatiana Iakovleva
CNRS, France
59, rue Pouchet
75017 Paris
tatiakovleva@yahoo.fr

Anna E. Piasecki
UWE Bristol
Coldharbour Lane
Bristol BS16 1QY
Anna.Piasecki@uwe.ac.uk

Ton Dijkstra
Radboud University
Montessorilaan 3
6500 HE Nijmegen
T.Dijkstra@donders.ru.nl

1 Introduction

This study examines the impact of the cross-linguistic similarity of translation equivalents on word recognition by Russian-English bilinguals, who are fluent in languages with two different but partially overlapping writing systems. Current models for bilingual word recognition, like BIA+, hold that all words that are similar to the input letter string are activated and considered for selection, irrespective of the language to which they belong (Dijkstra and Van Heuven, 2002). These activation models are consistent with empirical data for bilinguals with totally different scripts, like Japanese and English (Miwa et al., 2014). Little is known about the bilingual processing of Russian and English, but studies indicate that the partially distinct character of the Russian and English scripts does not prevent co-activation (Jouravlev and Jared, 2014; Marian and Spivey, 2003; Kaushanskaya and Marian, 2007).

Many Russian-English translation equivalents are in part composed of shared letters that can potentially activate both Russian and English word candidates. Often, these letters have ambiguous phonemic mappings across the two languages. The degree of ambiguity is high especially when shapes of block-letters and letters in italics overlap across languages. For instance, a printed Russian letter '*и*' does not look like any letter of the English alphabet, but the shape of its handwritten equivalent '*u*' perfectly coincides with the English hand-written grapheme. We identified 5 overlapping pairs of printed English block-letters and Russian letters in italics (g, r, m, n, u).

Our study started from the assumption that even when a bilingual reads English words in printed font, letter shapes also activate handwritten Russian letters with similar shapes in a bottom-up way. We focused on the impact of convergence and divergence in Russian and English script coding for cognates and non-cognates. Cognates are translation equivalents with significant cross-linguistic form overlap in phonology and/or orthography (e.g., 'marriage' in English, 'mariage' in French). Cognates are generally processed more quickly by bilinguals than matched control words (for an overview of studies, see Dijkstra, Miwa et al., 2010). However, as far as we know, cognate processing for the Russian-English language pair has not been examined before.

2 Predictions

We are making the following predictions about English word recognition by Russian-English bilinguals:

1. In English word processing, Russian-English bilinguals will activate lexical candidates that are similar to the input word in both Russian and English (language non-selective lexical access).
2. English-Russian cognates will be recognized more quickly than English control words, due to co-activation and convergence (cognate facilitation effect, Dijkstra, Miwa et al., 2010; Lemhöfer and Dijkstra, 2004).
3. Cognates with ambiguous orthography, i.e. shared letters mapping onto different phonemes in the two languages, will be processed more slowly than cognates with mismatching orthography, due to decreased facilitation from the other cognate member.

The following two predictions are more speculative and exploratory in nature.

4. Response times to cognates with transparent orthography, i.e. shared letters mapping onto the same phonemes in the two languages, will be about equal to those for cognates with mismatching orthography, because transparent orthography and shared phonology will lead to increased lexical competition, but, at the same time, the

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

transparency will lead to increased semantic co-activation of cognates in the two languages.

5. English control words with mismatching orthography will be processed more quickly than words with ambiguous orthography, because less interference from the Russian alphabet is expected in the first case.

3 Method

To test these hypotheses, we first constructed a large database of Russian-English cognates with three, four, five or six letters in length. To our knowledge, no such database is currently available to the community of researchers. Next, 75 English cognates were selected as test words in a lexical decision task. Orthographic coding was performed on English cognate words written in lower-case block letters in Arial font. The resulting items were allocated to three categories: 1) Cognates with Ambiguous Orthography (CAO=Minus condition), composed of letters that have different phonological mappings in English and Russian (e.g. ‘guru’ might be read as /digi/ if a Russian monolingual was asked to read this string of letters); 2) Cognates with Transparent Orthography (CTO=Positive condition), composed of letters that largely share their orthographic-phonological mappings with letters of the Russian alphabet (e.g. in ‘koala’ the only mismatch with the Russian alphabet is the grapheme ‘l’); 3) Cognates with Mismatching Orthography (CMO=Base condition), composed mostly of letters that do not exist in the Russian alphabet (e.g. ‘filter’). The cognate types were matched across conditions (CAO/CTO/CMO) in word length, frequency, and degree of cross-linguistic orthographic overlap between Russian and English alphabets. Three groups of control words were then selected that matched the cognates of each type with respect to these three dimensions. Finally, each cognate and non-cognate was matched with a pseudo-word generated with the help of the Wuggy-software (crr.ugent.be).

Next, 20 Russian-English bilinguals were asked to rate the visual similarity between the English cognates and their Russian translation equivalents. They also rated the semantic similarity of all selected item pairs. Rating results showed that bilinguals mostly considered orthographic congruence (as opposed to incongruence) between the orthography of Russian and English translation equivalents and gave higher ratings to English words that have shared orthography with the Russian alphabet. Ratings also indicated that bilinguals considered not only block-letters but also corresponding handwritten graphemes when rating the visual similarity between words.

In total, 37 Russian-English participants (10 male vs. 27 female; age: 19-60 years) took part in the study. At the moment of testing, all participants were residing in English-speaking countries: 11 participants in Bristol, UK, 21 participants in Sheffield, UK, and 5 participants in New Zealand. After the experiment, all participants rated their proficiency in English on a scale from 1 (the lowest) to 6 (the highest). Average ratings for reading, writing, speaking, and listening varied between 4.4 and 5. Except for two participants, ages of L2 acquisition (AoA) ranged between 6 and 19 years. Length of residence in an English-speaking country varied between 3 months and 21 years (mean = 33 years, SD = 11 years).

Participants performed an English lexical decision task, in which they pressed a “yes” or a “no” button depending on whether a presented word was English or not. They were asked to press a button as quickly and accurately as possible. The items were presented in a pseudo-randomized order to each participant. The experiment was programmed in E-Prime. Reaction times (RTs) and accuracy of responses were measured. Only correct responses to real words were included in the analyses of reaction times.

4 Results

First, all responses faster than 300 ms and slower than 3 s were removed from the data set, because they were not considered as valid measurements. Next, the data from 9 participants were excluded from analysis, because they had a response accuracy below 70%. We removed 5 cognates, 8 control words, and 14 non-words from the items, because these items had an accuracy below 70 % or had extremely slow responses. For the remaining 28 participants, after removing these items, cognate and control word conditions were still matched with respect to length and frequency (as shown by non-significant t-tests). None of the remaining responses were further apart than 2.5 SDs from the participant mean in each condition. The mean RT for non-words was 892 ms. Table 1 presents the mean RTs for words in each cognate and control word condition, as well as their accuracy.

Condition Type	Cognates	Controls	RT difference
Base	661 (82.2) .97	727 (112.7) .95	66
Minus	711 (105.7) .94	734 (106.4) .93	23
Plus	656 (89.01) .97	730 (113.1) .92	74

Table 1. Mean reaction times and accuracies for word categories (standard deviations between parentheses).

The word data were analyzed by means of a repeated-measures Analysis of Variance (ANOVA), using cognate type (3, MO vs. AO vs. TO) and cognate status (2, cognate vs. control) as within-subject factors. This analysis resulted in main effects of Cognate Status ($F(1, 27) = 94.11$, $p<.001$), Item Type ($F(2, 54) = 9.89$, $p<.001$), and an interaction of Cognate Status with Item Type ($F(2, 54) = 10.22$, $p<.001$). Next, we did planned comparisons to test the Cognate Minus (CMO) and Cognate Plus (CTO) conditions against the Cognate Base (CMO) condition. Significant differences were found between the RTs between the Cognate Base condition and the Cognate Minus condition ($t(27)=-5.0$, $p<.001$ two-tailed) but not between the Cognate Base and the Cognate Plus condition ($t(27)=.60$, $p=.55$). There was a significant difference between the Cognate Base condition and the Control Base condition ($t(27)=-6.54$, $p<.001$). Finally, no significant differences arose between the different control conditions (Control Base vs. Control Minus, $t(27)=-.67$, $p=.51$; Control Base vs. Control Plus $t(27)=-.36$, $p=.72$).

5 Discussion

Russian-English bilinguals performed an English lexical decision task with purely English control words and English-Russian cognates 1) with mismatching orthography or 2) shared orthography with a) transparent or b) ambiguous mappings on phonemes in Russian and English.

Responses to cognates were faster than to English controls (see Table 1). This cognate facilitation effect is in line with prediction 1 that lexical candidates in both Russian and English are activated during Russian-English bilingual word recognition.

It also confirms prediction 2 that language non-selective lexical access takes place in Russian-English word recognition. Because the effect is also observed in cognates with (partially) mismatching orthography, the cognate effect may in part be ascribed to the phonological and semantic overlap in these cognates. Thus, the orthographic input representation quickly leads to an activation of sublexical and lexical phonological representations (cf. Peeters et al., 2013).

In line with prediction 3, the cognate facilitation effect is modulated by the degree of shared transparent overlap between Russian and English alphabets. Cognates with transparent orthography were processed faster than cognates with ambiguous grapheme to phoneme mappings. This finding can be explained by assuming that Russian words are co-activated with English words to the extent that they match the English letter input, irrespective of whether this matching is in terms of block letters or handwritten visual similarity. In other words, it is purely a bottom-up (signal-driven) effect.

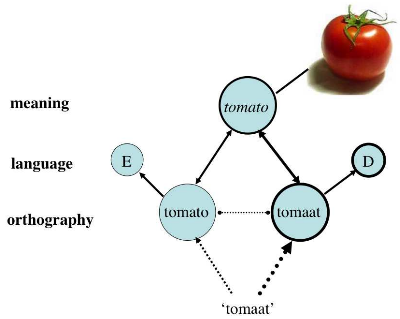


Figure 1. Localist connectionist illustration of cognate representation and processing, adapted from Dijkstra, Miwa et al. (2010).

The finding that cognates with mismatching orthography and shared orthography with transparent grapheme-to-phoneme mappings are responded to about equally fast, is in line with prediction 4, which is based on the representation for cognates that has been proposed by Dijkstra, Miwa et al. (2010). As Figure 1 indicates, both form representations of cognates are assumed to be activated based on the input and they spread activation to convergent semantic representations. The co-activation of form representations results in lexical competition and interference (Dijkstra, Hilberink-Schulpen et al., 2010), whereas the convergence on semantics results in facilitation. As a result, the RT difference between cognates with mismatching orthography and shared transparent orthography may be relatively small, due to a cancelling out of the effects of increased lexical form competition and increased semantic co-activation.

Finally, in contrast to prediction 5, English control words with mismatching orthography were not processed more quickly than control words with ambiguous orthography. Apparently, mismatching orthography in general did not result in any systematic interference on word processing speed. Said differently, the noise introduced by spuriously activated word candidates from Russian with overlapping letters in the other control conditions did not systematically affect the lexical decision to the English target word, although it may have affected the participants' general decision-making strategies in the experiment. In terms of interactive activation models, the increase in noise could be cancelled out by a somewhat higher reliance on semantic codes or global lexical activation (Grainger and Jacobs, 1996) for making the lexical decision.

In all, the obtained patterns of results are in support of interactive activation models for bilingual word recognition, such as the BIA+ model (Dijkstra and Van Heuven, 2002) when the assumption is made that cognates are represented in terms of overlapping but lexically competing form representations and largely shared semantic representations in the two languages (Dijkstra, Miwa et al., 2010), see Figure 1. Even the somewhat counter-intuitive prediction 4 can find a reasonable explanation in terms of such models. Prediction 5 was not confirmed, but the actually obtained result can be interpreted in terms of slightly shifted lexical decision criteria.

This study confirms the presence of language non-selective lexical access in visual word recognition by different script-bilinguals, in line with, e.g., for Korean-English Kim and Davis (2003) and for Japanese-English Hoshino and Kroll (2008), Miwa et al. (2014), and Ando et al. (2015). Moreover, it bridges research on shared scripts and different scripts by considering the partially overlapping Latin and Cyrillic scripts of English and Russian. It is innovative in showing that cross-linguistic effects depend on the degree of overlap in scripts depending on the exact characteristics of the words involved.

The study also provides indirect support for various types of models that assume co-activation of word candidates that are orthographically similar to the input letter string. The set of such candidates is often referred to as the neighbourhood (Grainger and Dijkstra, 1992). Van Heuven et al. (1998) have shown that the number of neighbours within and between languages affects bilingual word recognition. This result has recently been confirmed by Mulder and Dijkstra (under revision). The present study provides confirmation for these models from a completely independent perspective, that of cross-linguistic similarity effects in scripts.

To conclude, we presented evidence in favor of language non-selective lexical access in Russian-English bilinguals, showing an English-Russian cognate facilitation effect, the size of which depended on whether there was overlap in orthography or not, and on whether this overlap was ambiguous or transparent relative to phonology. These effects were shown to be lexical in nature, because mismatching orthography in control target words with translations that are completely different in form did not show any evidence of differential processing.

Acknowledgments

This research was made possible with support from NetWordS, the «European Network on word structure in the languages in Europe» (research grant n° 09-RNP-089). The authors are also deeply indebted to the EPSRC's RefNet research network that enabled us to collect a large part of our data. We also wanted to thank the anonymous reviewers for their helpful comments on an earlier version of this paper.

References

Eriko Ando, Kazunaga Matsuki, Heather Sheridan, and Debra Jared. 2015. The locus of Katakana-English masked phonological priming effects. *Bilingualism: Language and Cognition*, 18: 101-117.

Ton Dijkstra, Béryl Hilberink-Schulpen, and Walter J. B. Van Heuven. 2010. Repetition and masked form priming within and between languages using word and nonword neighbors. *Bilingualism: Language and Cognition*, 13: 341-357.

Ton Dijkstra, Koji Miwa, Bianca Brummelhuis, Maya Sappelli, and Harald R. Baayen. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62: 284-301.

Ton Dijkstra and Walter J. B. Van Heuven. 2002. The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5: 175-197.

Jonathan Grainger and Arthur M. Jacobs. 1996. Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103: 518-565.

Jonathan Grainger and Ton Dijkstra. 1992. On the representation and use of language information in bilinguals. In R. J. Harris (ed.), *Cognitive Processing in Bilinguals*, 207-220. Amsterdam: Elsevier.

Noriko Hoshino and Judith F. Kroll. 2008. Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, 106: 501-511.

Olessia Jouravlev and Debra Jared. 2014. Reading Russian–English homographs in sentence contexts: Evidence from ERPs. *Bilingualism: Language and Cognition*, 17: 153-168.

Margarita Kaushanskaya and Viorica Marian. 2007. Bilingual Language Processing and Interference in Bilinguals: Evidence From Eye Tracking and Picture Naming. *Language Learning*, 57: 119-163.

Jeesun Kim and Chris Davis. 2003. Task effects in masked cross-script translation and phonological priming. *Journal of Memory and Language*, 49: 484-499.

Kristin Lemhöfer and Ton Dijkstra. 2004. Recognizing cognates and interlingual homographs: Effects of code similarity in language specific and generalized lexical decision. *Memory and Cognition*, 32: 533-550.

Viorica Marian and Michael Spivey. 2003. Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, 24: 173-193.

Koji Miwa, Ton Dijkstra, Patrick Bolger, and Harald R. Baayen. 2014. Reading English with Japanese in mind: Effects of frequency, phonology, and meaning in different-script bilinguals. *Bilingualism: Language and Cognition*, 17: 445-463.

Kimberley Mulder and Ton Dijkstra (under revision). Revisiting the neighbourhood: The processing of cross-language hermits and neighbours in different tasks. *Bilingualism: Language & Cognition*.

David Peeters, Ton Dijkstra, and Jonathan Grainger. 2013. The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, 68: 315-332.

Walter J. B. Van Heuven, Ton Dijkstra, and Jonathan Grainger. 1998. Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, 39: 458-483.

A study of relations between associative structure and morphological structure of Hungarian words

Dániel Czégel
Budapest University of Technology and Economics
czegel_d@yahoo.com

Zsolt Lengyel
Pannon University, Veszprém

Csaba Pléh
Central European University, Budapest
vispleh@ceu.hu

The paper mainly aims to reanalyze data with the presently available corpus linguistics tools from a relatively large scale paper-and-pencil based Hungarian verbal association dictionary with regard to two aspects. *i)* The mental lexicon issue. How are associative overlaps representing structural relations in the mental lexicon? *ii)* The systemic variability of the associative fields mobilized by the stimulus words: how variable the responses are, and how these associative entropies are related to morphological entropies of the same words.

1 Methods and materials

For the associative corpora, two dictionaries of Lengyel [3] were used. They are based on the responses of 2000 students between $10-14$ and $18-24$ to about 200 stimulus words. Digitized responses from this dictionary were related to the frequency distribution of 800 million web-based Hungarian words from the MOKK corpus [2].

2 Results

2.1 Associative overlaps and lexical fields

Based on the associative overlap measure introduced by Deese [1], a multidimensional scaling method was used to obtain associative fields depicting the pairwise associative distance of stimulus words in a two-dimensional figure. The results indicate that young adults have a more dense structure, their associative clusters are more tight compared to those of children of age 10-14, as illustrated in figure (1) and (2) and shown quantitatively in figure (3) and (4).

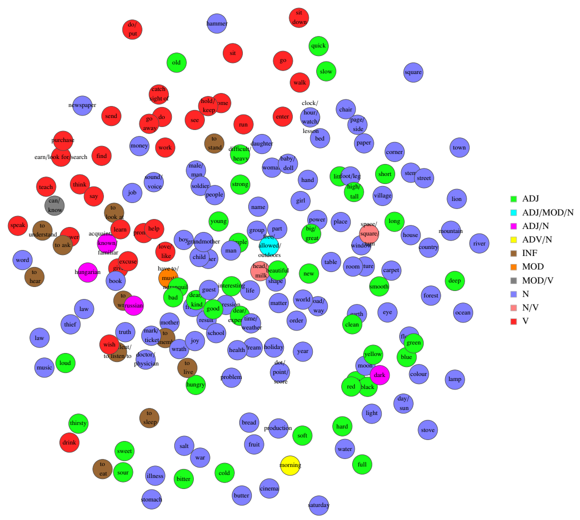


Figure 1: Associative field of children (age 10-14)

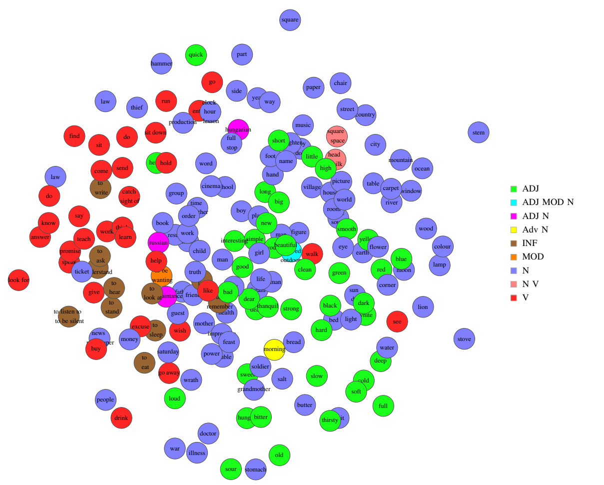


Figure 2: Associative field of young adults (age 18-24)

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

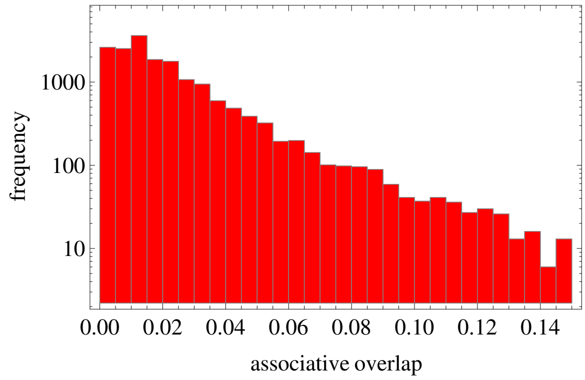


Figure 3: Histogram of pairwise associative overlaps (age 10-14)

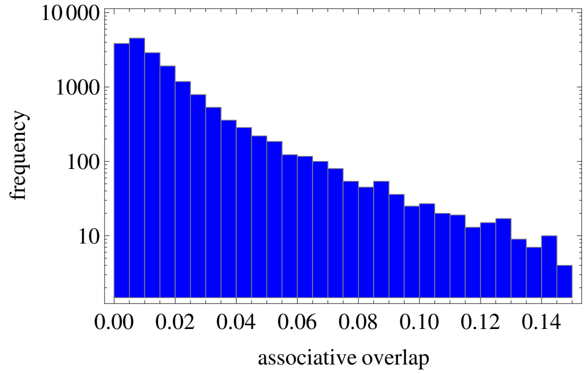


Figure 4: Histogram of pairwise associative overlaps (age 18-24)

2.2 Relationship between associative entropy, morphological entropy and frequency

We followed the methods introduced by Osgood [5] for analyzing the variability of morphological and associative structure of words. As shown in figure (5), there is an interesting difference of the relations between associative entropy (defined as $H_A = -\sum_i p_i \log_2 p_i$, where p_i is the relative frequency of the ith associated word) and corpus morphological entropy (defined as $H_M = -\sum_i q_i \log_2 q_i$, where q_i is the relative frequency of form i in the MOKK corpus) between nouns and verbs. In nouns, the more varied the morphology of the noun is in the corpus, the more variable the associative field is ($r = 0.202$, 0.175 in the two ages). That can be interpreted as implying that the more varied the suffixation of a noun is, the more variable associative relations it enters with other words. In verbs, however, if the verb has a more varied morphology, it has less associations ($r = -0.194$, both groups). As figure (6) shows, a similar relationship has been obtained between associative entropy and the logarithm of corpus frequency: In the case of nouns the correlation is positive ($r = 0.134$ and 0.367 in the two ages) while the correlation is negative ($r = -0.281$, -0.222) for verbs. This peculiar relation would be further studied with considering morphological entropy in light of the argument frames of the verbs on the one hand, and the role of syntagmatic associations in the associative fields of verbs on the other [4].

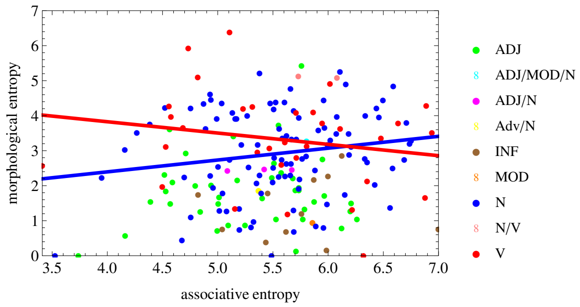


Figure 5: Relation between associative and morphological entropy (age 18-24). The blue regression line corresponds to nouns, while the red line corresponds to verbs.

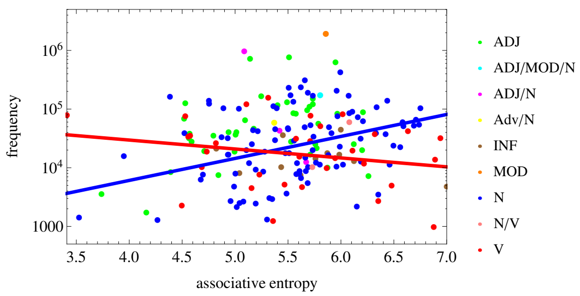


Figure 6: Relation between associative entropy and corpus frequency (age 18-24). The blue regression line corresponds to nouns, while the red line corresponds to verbs.

References

[1] James Deese. *The structure of associations in language and thought*. Johns Hopkins University Press, 1966.

[2] Péter Halácsy, András Kornai, László Németh, András Rung, István Szakadát, and Viktor Trón. Creating open language resources for hungarian. In *LREC*, 2004.

[3] Zsolt Lengyel. *Magyar asszociációs normák enciklopédiája I-II.* TINTA könyvkiadó, Budapest, 2008-2010.

[4] Katherine Nelson. The syntagmatic-paradigmatic shift revisited: a review of research and theory. *Psychological Bulletin*, 84(1):93, 1977.

[5] Charles Egerton Osgood. *Cross-cultural universals of affective meaning*. University of Illinois Press, 1975.

Suffix perceptual salience in morphological processing: evidence from Italian

Hélène Giraudo
Laboratoire CLLE (Equipe ERSS)
CNRS & Université Toulouse Jean Jaurès
giraudo@univ-tlse2.fr

Serena Dal Maso
Dip. Lingue e Letterature Straniere
serena.dalmaso@univr.it

Abstract

The goal of the present research is to determine the role of suffixes and morphological schemas in the access and processing of Italian complex words and to investigate whether (and possibly to what extent) suffix salience affects such processes. Two experiments using the masked-priming methodology will contribute to verify if native speakers of Italian organize lexical items according to morphological series as they do according to morphological families.

1 Introduction

In usage-based approaches to language representation and process (mainly Bybee's Network Model and Booij's Constructional Morphology), morphology is generally conceived as organizing the lexicon according to two main dimensions: i) morphological families, i.e. words connected because sharing the same root: *kind/ kindness/ kindly/ unkind/ kind-hearted*, etc. and ii) morphological series, i.e. words connected because sharing the same affix *kindness/ happiness/ sadness/ abruptness*, etc. Psycholinguistic research has mostly confirmed this view, demonstrating with experimental data that words in the mental lexicon are stored according to formal and semantic similarity, thus following morphological principles.

More specifically, the relationship between morphologically complex words and their roots (or other members of the same morphological family) has been extensively investigated by means of the masked-priming experimental paradigm (i.e. Stanners, Neiser, Hernon & Hall, 1979; Rastle, Davis, Marslen-Wilson & Tyler, 2000; Clahsen, Sonnenstuhl & Blevins, 2003; Rastle, Davis & New, 2004; Frost, Kugler, Deutsch & Forster, 2005). This technique focuses on the effect of the (visual) presentation of a stimulus word (the 'prime') on the recognition of a target word. Experimental results indicate that the recognition of the target word is faster when it is preceded by a morphologically related prime (e.g. *kindness/ KIND*), compared to cases where it is preceded by an unrelated word (e.g. *raw/ KIND*) or by an only orthographically similar word (e.g. *kin/ KIND*; *kite/ KIND*). According to Forster, these results show that "the cortical representations of the prime and the target are interconnected or overlap in some way such that the representation of the prime automatically activates the representation of the target word" (Forster, 1999).

On the other hand, the relationship between words with the same suffix and the same morphological schema (in constructional terms), like *kindness/ happiness/ sadness*, has been scarcely investigated yet and results do not allow a consistent and univocal interpretation. Marslen-Wilson et al. 1996 investigated the role of suffixes in English with a cross-modal technique and found a significant priming effect for morphologically related words (e.g. *darkness/ TOUGHNESS*) and no hints of orthographic priming when the overlap did not involve real suffixes (e.g. *darkness / HARNESS*). More recently, Duñabeitia, Perea & Carreiras 2008 found significant facilitation effects on the recognition of suffixed words in Spanish employing a series of experiments with different degrees of prime segmentation: 1) er/ WALKER; 2) %%%%er/ WALKER; 3) baker/ WALKER. The experiments revealed priming effects in all the conditions (independently from the degree of segmentation of the prime) and a clear dissociation between orthographic and morphological priming (e.g. *brevidad* primes *igualdad* but *volumen* does not prime *certamen*). Taken together these results were interpreted as a strong evidence in favor of an early prelexical morphological decomposition (e.g., Duñabeitia et al., 2007; Rastle et

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

al, 2004) of all forms that can be potentially split into two "surface morphemes" (see for details Rastle & Davis, 2008) acknowledging to both stems and affixes an equal status of access units during word recognition.

However, when Giraudo & Grainger 2003 addressed this issue using French materials and an experimental design controlling the effect of morphological primes relative to formal primes, results did not show any reliable morphological priming effect, i.e. both priming conditions produced significant priming effects relative to the unrelated baseline but the morphological condition did not yield significantly faster RTs with respect to the orthographic condition. Note that, according to within priming comparisons, the effect of morphological primes is compared to the effect of the orthographic primes on the same targets, e.g., *fumet* 'scent' - *MURET* 'down wall' vs. *béret* 'beret'- *MURET* 'down wall', considering that *fumet* and *muret* share the same functional suffix *–et*, while *béret* and *muret* do not because *béret* is a monomophemic word in French and *ber-* is not a possible stem. Giraudo and Grainger, who conversely found in the same study clear morphological priming effects when manipulating prefixed words, interpreted these asymmetrical results on the base of different semantic and syntactic functions carried by prefixes and suffixes in French. An alternative explanation for the results of Giraudo & Grainger study could be linked to the issue of perceptual salience of suffixes (i.e. their size and segmental-prosodic features) and to the connected degree of suffix likelihood (the probability for a word to be a suffixed word). As a matter of fact, it seems that the more a word ending is salient and functionally consistent, the stronger the probability it is a suffix.

2 The present study

On such premises, in the present research we verify by means of a masked priming experiment and a within-comparison design whether the processing of morphologically complex words is affected by the morphological schema and, more specifically, whether the processing is affected by the formal salience of the suffix.

We choose to run the experiments on Italian not only because Italian has a rich, productive and relatively regular morphology, but also because, being a phonetically 'conservative' language, at least significantly more conservative than French, Italian has relatively long suffixes (e.g. lat. *-ĭttu(m)* > it. *-etto vs. fr. –et*, realized phonetically as [e] as in it. *muretto*/fr. *muret*).

Moreover, as a result of the fact that Italian has undergone little phonological reduction, it has a high degree of orthographic transparency and consistency, which can contribute to the perception and representation of functional word endings (Taft 2003).

Finally, although in Italian the great majority of suffixed words are paroxytone, i.e. stressed on the penultimate syllable, as suffix generally carry the word stress, there is a limited number of proparoxytone words (i.e. stressed on the third to last syllable, with a suffix which does not carry the word stress). Consequently, suffixed words in Italian can have different prosodic contours and suffixes can show different degrees of perceptual prominence at the prosodic level. For these reasons, we considered Italian as an ideal test situation to verify the role of salience on suffixed word processing and access.

More precisely, for our experiments we selected some productive suffixes *–tore*, *–ico* and *-etto* because they show different segmental and prosodic features.

Moreover, they have different degrees of functional consistency, i.e. a different proportion between suffixed and non-suffixed words (i.e. monomophemic words) in a series of words ending with a given letter string (Laudanna et al. 1994). As a matter of fact, while 78% of the words ending with *–tore* and 52,04% of words with *–ico* are suffixed, only 20% of the words ending with *–etto* is suffixed (quantitative data are taken from COLFIS and Derivatario). The criteria according to which we defined the perceptual salience of the suffixes are:

i. **size** of the suffix (number of phonemes and graphemes);

ii. **different degrees of morpho-tactic transparency** (Dressler 1985) and of phonological integration of the suffix to the base, in particular in relation to the phenomenon of:

- resyllabification: no resyllabification takes place with *-tore* which has always two syllables, independently from the root, whereas *–ico*, and *–etto*, starting with a vowel, are more integrated with the stem ([i] and [e] become the coda of the last syllable of the stem (*sto.ria/ sto.ri.co*) and the suffixed word is resyllabified);
- morphological boundary: with *-tore* the boundary of the suffix always coincides with

the boundary of the syllable, whereas with *–ico* and *–etto* the suffix is split in the two last syllables. In the Natural Morphology framework, the more the morphology overlaps with the phonological components (i.e. the higher the morpho-tactic transparency) the easier the recognition;

iii. **word stress**: the suffixes *-tore* and *–etto* always carries the word stress, while *-ico* does not. Moreover, in Italian, the stressed syllable has a long vowel [–'to:re] which, although not phonological, may constitute a perceptual hint for an easier identification. Finally, words with *–tore* and *–etto* show the more frequent stress pattern in Italian (about 80% of the words have the word stress on the penultimate syllable, Thornton, Iacobini & Burani 1997, see Burani & Arduino 2004 and Giraudo & Montermini 2010 on the effect of stress regularity and stress consistency in stress assignment for Italian words).

According to these criteria – *tore* is the most salient suffix and *–etto* is more salient than *–ico*.

In the first experiment we will verify: a) whether words with a perceptually salient suffix like *–tore* are recognized faster than words with a less salient suffix like *–ico*. If this would be the case, the word *lavoratore* should prime *viaggiatore* better than *ironico* primes *metallico*; b) whether a word belonging to a more consistent word ending series (like *–tore*) is recognized faster than a word belonging to a less consistent word ending series (like *–etto*). According to this hypothesis, we expect higher priming effect for words with *–tore* than for words with *–etto*.

The affix condition (our test condition), i.e. the effect of the presentation of a suffixed word as a prime on the recognition of a complex target word with the same suffix (*servitore/ EDUCATORE, sinfonico / NOSTALGICO, boschetto/ PEZZETTO*), will be considered in relation to 3 other conditions: the identity condition (*educatore/ EDUCATORE, nostalgico/ NOSTALGICO, pezzetto/ PEZZETTO*) which should yield the main facilitation effect and consequently the shortest RTs and the unrelated condition (*colomba / EDUCATORE, approccio/ NOSTALGICO, ombelico/ PEZZETTO*) which, on the contrary, is expected to yield the smallest facilitation effect and the longest RTs. These two conditions are considered as baselines to assess RTs obtained in the test condition. Moreover, in the stem condition we will contrast the strength of the connection between words with the same suffix and morphological schema (test condition) with the strength of the connection between words sharing the same stem (*educare / EDUCATORE, nostalgia / NOSTALGICO, pezzo / PEZZETTO*).

In the second experiment we will focus on the issue of the sequential organization of the word, namely that the access and processing of a suffixed word is affected by the position of the suffix at the end of the word and by the (visual) perception of the final part of the word. In order to verify this aspect, we will use the same critical materials as in the first experiment but we will manipulate the location of the fixation point.

Specifically, in the forward mask which precedes the presentation of the prime/target pairs, the fixation marks (####), whose aim is to focus attention on a certain point of the screen, will overlap with the suffix position.

To sum up, our research will contribute to verify the role of suffixes and morphological schemas in the access and processing of Italian complex words and to investigate whether (and possibly to what extent) suffix salience affects such process. Results will indicate if native speakers of Italian organize lexical items according to morphological series as they do according to morphological families.

References

Alessandro Laudanna, Cristina Burani, Antonella Cermele(1994). Prefixes as processing units. *Language and Cognitive Processes***,** 9, 295-316

Bertinetto, P. M., Burani, C., Laudanna A., Marconi L., Ratti D., Rolando C., Thornton A. M. (2005). *Corpus e Lessico di Frequenza dell'Italiano Scritto* (CoLFIS). http://linguistica.sns.it/CoLFIS/CoLFIS_home.htm.

Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.

Burani, C., & Arduino, L.S. (2004). Stress regularity or consistency? Reading aloud Italian polysyllables with different stress patterns. *Brain and Language, 90,* 318-325.

Bybee, J. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (eds.), *Theoretical Morphology. Approaches to modern linguistics* (pp. 119-142). San Diego: Academic Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes*, *10*, 5, 425-55.

Clahsen, H., Sonnenstuhl, I. & Blevins, J.P. (2003). Derivational morphology in the German mental lexicon: A Dual Mechanism account. In H. Baayen & R. Schreuder (Eds.), *Morphological structure in language processing* (pp. 125-155). Mouton de Gruyter: Berlin.

Derivatario, http://derivatario.sns.it/

Dressler, W. (1985). On the predictiveness of Natural Morphology. *Journal of Linguistics 21*, 321

Duñabeitia J. A., Perea M. & Carreiras M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition* 105, 691–703.

Forster, K.I. & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

Forster, K.I. (1999). The microgenesis of priming effects in lexical access. *Brain and Language, 68,* 5-15.

Frost, R., Kugler, T., Deutsch, A. & Forster, K.I (2005). Orthographic structure versus morphological structure: principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *31*, 1293-1326.

Giraudo, H. & Grainger, J. (2003). On the role of derivational affixes in recognizing complex words: Evidence from masked affix priming. In R. H. Baayen and R. Schreuder (Eds.), *Morphological Structure in Language Processing*. Mouton de Gruyter: Berlin, 209-232)

Giraudo, H. & Montermini, F. (2010). Primary stress assignment in Italian: linguistic and experimental issues. *Lingue e Linguaggio, 2*, 113-129.

Grossmann, M. & Rainer, F. (eds.), *La formazione delle parole in italiano*. Tübingen: Niemeyer.

Nespor, M. & Vogel, I. (2007). *Prosodic Phonology*. Mouton de Gruyter: Berlin.

Orsolini M. & Marslen-Wilson W.D. (1997), Universals in morphological representation: Evidence from Italian. *Language and Cognitive Processes* 12: 1-47.

Rastle, K. & Davis, M. H. 2008 Morphological decomposition based on the analyses of orthography. In : Language and Cognitive Processes. 23, 7-8, p. 942-971

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

Rastle, K., Davis, M.H., Marslen-Wilson, W.D. & Tyler, L.K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes 15* (4-5), 507-537.

Reid, A.A. & Marslen-Wilson, W.D. (2003), Lexical representation of morphologically complex words: Evidence from Polish. In Baayen, R.H. & Schreuder, R. (eds.), *Morphological Structure in Language Processing*, 287-336

Stanners, R.F., Neiser, J.J., Hernon, W.P. & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, *18*, 399-412.

Taft, M. (2003). Morphological representation as a correlation between form and meaning. In E. Assink, & D. Sandra (Eds.) *Reading complex words*. Amsterdam: Kluwer, 113-137

Thornton, A.M. Iacobini, C. & Burani, C. (1997). *BDVDB Una base di dati sul Vocabolario di Base della lingua italiana, seconda edizione riveduta e ampliata.* Roma. Bulzoni

Vroomen, J. & de Gelder, B. (1999). Lexical access of resyllabified words: evidence from phoneme monitoring. *Memory and Cognition 27* (3), 413-21.

A User-Based Approach to Spanish-Speaking L2 Acquisition of Chinese Applicative Operation

Nana Y-H Huang
Cambridge University
yhh28@cam.ac.uk

1 Introduction: Low Applicative Operations

Recent studies of argument structure distinguishes non-core (applied) arguments from core arguments in the sense that non-core ones do not belong to the basic argument structure of verbs and that they enter argument structures through Applicative Operations (AO) introduced by functional heads such as Low Applicative-source (LA-source) or Low Applicative-goal (LA-goal) heads (Pylkkänen, 2000; 2002; 2008; Cuervo, 2003). Because languages make use of different applicative heads, in this study, I examine the acquisition of Chinese AO by Spanish-speaking L2 learners and propose a usage-based approach for the results collected from a comprehension task and an acceptability judgment task.

1.1 Applicative Operations in Spanish

Cuervo (2003) reports that in Spanish a predicate which expresses the transfer of a theme to a goal, such as verbs indicating creation (e.g. cocinar 'cook/bake', construer 'build', and etc.), allows LA-goal, where the applied argument is the dative argument, as in (1).

(1) Valeria le diseñó una pollera a Anna.

Valeria CL designed a skirt DAT Anna

Lit.: 'Valeria designed Anna a skirt.'

A Spanish applied argument can also appear in the environment of a transfer predicate with 'reverse directionality', such as robar 'steal', sacar 'take from', and extraer 'take out from'. In this case the applied argument is understood as the possessive source of the theme object.

(2) Pablo le robó la bicicleta a Anna.

Pablo CL stole the bike DAT Anna

Lit.: 'Pablo stole Anna the bike.'

The source argument appears in dative case which has the same morphosyntactic properties of a recipient argument; therefore, it is predicted that in the context of verbs with underspecified directionality (e.g., vender 'sell' and alquilar 'rent') and verbs of motion (e.g., lanzar 'throw' and pataer 'kick'), the applied argument would be ambiguous between a goal and a source. Cuervo provides such an example as (3).

(3) Valeria le vendió el auto a su hermano.

Valeria CL sold the car DAT her brother

1. Valeria sold the/her car to her brother.
2. Valeria sold her brother's car.

1.2 Applicative operations in Chinese

In Chinese, AO is as productive; nevertheless, unlike Spanish, Chinese only allows LA-source (see (4)) but not LA-goal (see (5)):

(4) Zhangsan tou-le Lili liang tai diannao.

Zhangsan steal-PERF Lili two CL computer

'Zhangsan stole Lili of two computers.'

(5) *Zhangsan sheji-le Lili liang jian qunzi.

Zhangsan design-PERF Lili two CL skirt

'Zhangsan designed Lili two skirts.'

1.3 Research Questions

This study examines Spanish L2ers' acquisition of Chinese AO and considers the learnability problem posed by the superset-subset relation between Spanish and Chinese on this structure (i.e. Spanish allows both LA-goal and LA-source while Chinese allows only LA-source). We predict learners to wrongly transfer LA-goal, which is allowed in L1 Spanish, to L2 Chinese despite the lack of positive evidence for the use of LA-goal in L2 input. Furthermore, due to lack of negative evidence (from the fact that AO do not appear in pedagogical textbooks nor in classrooms designed for L2ers), L2 Chinese input lacks information regarding ungrammaticality of LA-goal, which would be necessary for L2ers to rule out incorrect hypotheses. That is, these learners are expected to show overgeneralization from early on till even at the advanced level.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

2 Methods

To test our prediction on L1 transfer effects we designed two tasks to probe different knowledge of L2 structures: one being implicit and meaning–focused; the other being explicit and form-focused.

2.1 Materials and Procedures

An Animation Matching Task (AMT) was used to probe L2er's implicit knowledge because it called for a focus on meaning. The AMT included 12 items (6 test sentences and 6 fillers). The 6 test sentences included verbs underspecified for directionality of transfer. The 6 fillers bore only surface similarity and served to distract participants' focus in different ways. 2 contained syntactically unacceptable sentences; another 2 contained sentences that matched both animations; the other 2 contained sentences that matched neither of the two animations. See Appendix A.

On each trial, the L2ers first saw 2 animations on the computer screen. Next, they heard the target sentence presented auditorily. Participants were required to match the sentence to the correct animation. For example,

(6) Zhansan reng-le Lisi yi jian waitao.

Zhangsan toss-PERF Lisi one CL coat

Lit: 'Zhangsan tossed Lisi one coat.'

The sentence was preceded by two animations: (a) Zhangsan tossed one coat to Lisi; (b) Zhangsan tossed one of Lisi's coats away. Participants chose which animation was a better match for the sentence by ticking the answer on the answer sheet. They were told at the beginning of the test that if they found both animations matching the sentence, they could select both. If they found neither matching the sentence or if they could not understand the sentence, they could choose 'don't know' option on the side and choose/state the reason. See Appendix B.

Following the AMT was the Acceptability Judgment Task (AJT), which tapped participants' explicit knowledge on forms. 2 different types of verbs that induced opposite directionality of transfer (i.e., grammatical LA-source and ungrammatical LA-goal) were included, 3 items per type. In addition, with 6 control sentences and 6 fillers, the AJT contained 18 items in total, half grammatical and half ungrammatical. Please see Appendix C. Rating scale ranged from very unacceptable (1), unacceptable (2), acceptable (3), to very acceptable (4). A 'don't know' option was provided on the side which learners could choose if they were unsure of the response. See Appendix D.

2.2 Participants

20 L2ers and 10 natives speakers (NS) of Chinese serving as a control group participated in this study. All NS were graduate students born and raised in Taiwan. Most L2ers were undergraduate students with the exception of 3 people being Catholic priests. L2ers had learned Chinese in Taiwan for at least 3 years and came from different Spanish-speaking countries. Spanish was the native language for all L2ers. English was the second most proficient language.

Before the study, L2ers had completed a 40-item Chinese proficiency cloze test developed by Yuan (2014). Based on the scores, they were divided into Advanced (AD) and Intermediate (IN) group. Table 1 summarizes the participants' background information and cloze test scores.

Group	NS	AD	IN
Number of participants	10	10	10
Mean age (ranges in brackets)	26.2 (22-28)	26.9 (23-38)	24.1 (20-36)
Duration (years) of formal instruction	NA	8.4	5.7
Length (years) of residence in Taiwan	NA	5.7 (3-11)	4.8 (3-9)
Cloze test score (ranges in brackets)	39 (38-40)	35 (33-37)	29 (27-32)

Table 1: Participants' Background Information

3 Results and discussion

Table 2 presents the percentage of how often participants chose a certain animation in the AMT (for example, the (a) condition in example (6) above depicts a Goal condition).

Group	Source	*Goal	*Both	Don’t know
NS	100	0	0	0
AD	57	10	33	0
IN	23	17	57	3

Table 2: Percentages of choice in the AMT

A 2-sample z-test was performed separately to compare proportions between any 2 among the 3 groups. The results showed that any 2 groups were significantly different from each other in the choice for Source and for Both, but not significantly different in Goal. IN group as expected showed overgeneralization in wrongly choosing Both, while AD group seemed to be able to overcome overgeneralization and limit the construction of Chinese AO to LA-source from the fact that the choice for Both was greatly decreased and that for Source was greatly increased at the higher proficiency level.

As for the AJT, Table 3 presents the mean scores with the standard deviation in the brackets of each group by verb types. Using an alpha level of 0.05, paired t-tests showed that only NS exhibited significant difference in the responses to 2 types of verbs, while L2 groups did not.

Group	Verb type	
	Consumption	Creation
NS	3.53(0.39)	1.36(0.24)
AD	3.22(0.54)	3.33(0.44)
IN	3(0.34)	3.23(0.38)

Table 3: Mean scores for the AJT

In contrast with the result in Table 2, AD group did not perform better in AJT than IN group in rejecting ungrammatical AO-Goal introduced by verbs of creation. The question is how we can explain for AD group’s inconsistency in overcoming overgeneralization.

Notice that the major difference between the 2 tasks is whether the verb specifies directionality of transfer. Verbs included in the AMT are the verbs that do not favor a particular direction of transfer and therefore the introduced applied argument is inherently ambiguous between Goal and Source in the L1 Spanish. In other words, the verbs that trigger ambiguity in L1 Spanish are where subjects first overcome overgeneralization.

It is, therefore, proposed that subjects’ experience in L1 to resort to context in the face of ambiguity caused by verbs underspecified for directionality helps advanced L2ers overcome overgeneralization. The sensitivity trained in L1 is transferred to L2 learning and displayed in that more attention is paid to the co-occurring applied argument in the face of ambiguous thematic role assigned to applied argument. Advanced L2ers might have accumulated enough indirect statistical information (Reali and Christiansen, 2005) tracked from co-occurrences of recurring sequences of words before being able to overcome overgeneralization. This finding suggests that the effects of L1 transfer result not only from the similarity and/or difference of linguistic facts between the native and the target language, but also from L2ers’ experience gained in their native language.

References

C. Cuervo. 2003. Datives at Large. PhD Thesis. MIT.

C. J. Huang. 2007. Hanyu dongci de tiyuan jiegou yu qi jufa biaoxian (The thematic structures of verbs in Chinese and their syntactic projection). *Yuyan Kexue* (Linguistic Sciences) 6(4): 3-21.

B. MacWhinney. 2004. A multiple process solution to the logical problem of language acquisition. *Journal of Child Language*, 31, 833-914.

L. Pylkkänen. 2000. What applicative heads apply to. In M. Minnick, A. Williams, and E. Kaiser (eds.), working papers in *Proceedings of the 24th Annual Penn Linguistics Colloquium* 7(1).

L. Pylkkänen. 2002. Introducing Arguments. PhD Thesis. MIT.

L. Pylkkänen. 2008. Introducing Arguments. Cambridge, MA: MIT Press.

F. Reali and M. Christiansen. 2005. Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science*, 29, 1007-1028.

T. Q. Sun and Y. F. Li. 2010. Hanyu fei hexin lunyuan yunzhun jiegou chu tan (Licensing non-core arguments in Chinese), Zhongguo Yuwen (*Studies of the Chinese Language*) 334: 21-33.

B. Yuan. 2010. Domain-wide or variable-dependent vulnerability of the semantic-syntax interface in L2 acquisition? Evidence from wh-words used as existential polarity words in L2 Chinese grammars. *Second Language Research* 26: 219-60.

Appendix A: Test Sentences in the AMT

Type of Verbs	Item	Question Number	Target Sentence
Test sentences	na2 'take'	1	小明拿了小華一本雜誌
	ban1 'carry'	6	小張搬了小李一張桌子
	reng1 'toss'	9	張三扔了李四一件外套
	tou1 'steal'	3	阿明偷了阿華兩瓶可樂
	mai4 'buy'	8	張三買了李四一支毛筆
	ying2 'win'	12	小張贏了小李一隻手錶
Fillers	sha1 'kill'	2	小明殺了小華兩頭小羊 (matches both)
	gei3 'give'	7	老李給了老張一隻小鳥 (matches both)
	dao3 'collapse'	10	老王倒了小李一棵小樹 (ungrammatical)
	gei3 'give'	4	老李關了老張一隻小鳥 (ungrammatical)
	song4'give'	5	小華送了小李兩瓶可樂 (matches neither)
	jiao1 'teach'	11	張三教了瑪莉兩題數學 (matches neither)

Appendix B: Sample Answer Sheet of the AMT

Question Number	Which Animation do you choose?		If you tick 'I don't know', please tick or state the reason
1.	☐A ☐B	☐ I don't know	☐ Neither of the two animations is correct. ☐ I do not understand the sentence that I heard. ☐Other reason ________________

Appendix C: Test sentences in the AJT

Type of Verbs	Item	Question Number	Target Sentence
Verbs of Consumption	chi1 'eat'	1	李四吃了張三兩個蛋糕
	he1 'drink'	8	小華喝了小明兩瓶紅酒
	yong4 'use'	17	小李用了小張一支鉛筆
Verbs of Creation	kao3 'bake'	6	*阿華烤了小明一個蛋糕
	zhu3 'cook'	12	*小李煮了老張一頓晚餐
	zao4 'build'	14	*張三造了老李一棟房子
Control Sentences	chi1 'eat'	2	李四吃了兩個蛋糕
	he1 'drink'	9	小華喝了兩瓶紅酒
	yong4 'use'	13	小李用了一支鉛筆
	kao3 'bake'	4	阿華烤了一個蛋糕
	zhu3 'cook'	11	小李煮了一頓晚餐
	zao4 'build'	16	張三造了一棟房子
Fillers	gei3 'give'	3	老李給了老張一隻小鳥
	song4 'give'	7	小華送了小李兩瓶可樂
	jiao1 'teach'	15	張三教了瑪莉兩題數學
	gei3 'give'	5	*老李給了隔壁老張
	song4 'give'	10	*小華送了鄰居小李
	jiao1 'teach'	18	*張三教了同學瑪莉

Appendix D: Sample Answer Sheet of the AJT

	Very Unacceptable	Unacceptable	Acceptable	Very Acceptable	
1. 阿明吃了我兩個蛋糕。	1	2	3	4	◘ I don't know

Visual word recognition of morphologically complex words: Effects of prime word and root frequency

Hélène Giraudo
Laboratoire CLLE (Equipe ERSS)
CNRS & Université Toulouse Jean Jaurès
giraudo@univ-tlse2.fr

Karla Orihuela
Laboratoire CLLE (Equipe ERSS)
CNRS & Université Toulouse Jean Jaurès
karla.orihueal@univ-tlse2.fr

Abstract

The present study aims to investigate the relative role of the surface frequencies (i.e., token frequencies) in base word recognition. A masked priming experiment was carried using two types of suffixed French primes: the effects of words having a surface frequency (SF) higher than their base (e.g., *cigarette – cigare*) were compared with those produced by word primes having a SF lower than their base (e.g., *froideur-froid* 'coldness-cold'). Results show that HighSF are more efficient primes than LowSF relative to both orthographic and unrelated priming baselines. This suggests that despite a highly salient base, whole words matter more than morphemes during the early processes of lexical access.

1 Introduction

Morphological complexity has been extensively explored by psycholinguists in order to shed light on the role of morphology in lexical structuring. Starting from the idea - inherited from the connectionist theory of visual word recognition (see Seidenberg, 1987) - that the lexicon is comprised of different levels of interconnected representations reflecting the linguistics characteristics of the words as well as the cognitive processes by the which complex words are recognized, the main issue regarding lexical morphology concerns its specific role relative to word forms and semantics. Accordingly, morphology can be though as a structuring factor either for the lexicon, morphological relationships being expressed by the mapping between from and meaning reflecting the construction of the words (e.g., Giraudo & Voga, 2007; 2014; Giraudo & Grainger, 2000; 2001; but see also Aronoff, 1994 and Booij, 2010 for the same linguistic view) or for the access ways to the lexicon, morphology influencing the simple development of orthographic representations (e.g., Duñabeitia et al., 2007; Rastle & Davis, 2003; Rastle et al., 2004 and see in the same vein Marantz, 2013). An interesting way to explore this issue is to use the masked priming paradigm (Forster & Davis, 1984) which has been designed to measure the qualitative and the quantitative effects induced by the prior processing of a morphologically complex word presented visually on the subsequent processing of another -target- word. Behavioural data obtained with the masked priming paradigm associated with the lexical decision task revealed clear strong morphological priming effects through various languages (Arabic: Boudelaa & Marslen-Wilson, 2001; Basque: Duñabeitia, Laka, Perea, & Carreiras, 2009; English: Rastle, Davis, Marslen-Wilson & Tyler, 2000; French: Giraudo & Grainger, 2000; German and Dutch: Drews and Zwitserlood, 1995; Greek: Voga & Grainger, 2004; Hebrew: Frost, Deutsch & Forster, 1997) but the results are still controversial when manipulating the relative frequencies of the prime and the target. On the one hand, some studies (Giraudo & Grainger, 2000) have revealed that larger effects are obtained when using high in comparison to low frequency derived primes encouraging the lexeme-based approach; on the other, some authors (McCormick, Rastle, & Davis, 2009) have failed to observe an interaction between the prime frequency and morphological facilitation, strengthening the morpheme-based approach. It has been suggested that these outcomes may be due to the fact that the methodological procedure among experiments varies (Amenta & Crepaldi, 2012), as they each use a

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

different timing for their stimulus onset asynchrony (SOA). The SOAs of the original experiments were of 57ms in Giraudo and Grainger (2000) and 42ms in McCormick, Rastle, & Davis (2009).

2 The study

In order to disentangle such findings, the present study was carried out using the same paradigm and similar SOAs as the previous ones described. The main manipulation was to compare morphological facilitation when the frequency of the complex words (used as primes) and their roots (used as targets) was inverted. More specifically we selected 60 base word targets from French, half having systematically a surface frequency higher than their derived forms (55.82 occ./million) and the other half a surface frequency lower than their derived forms (10.15 occ./million according to *Lexique* database, New, Pallier, & Ferrand, 2001). Each target was primed by (1) a morphologically related word (M, e.g., *mariage-marier 'wedding – to marry'*), (2) an orthographically related word (O, e.g., *marine-marier, 'navy-marry'*) and (3) an unrelated word (U, e.g., *courage-marier, 'courage-marry'*). In both the HighSF condition and the LowSF condition, primes were matched in number of letters (respectively 6.4 and 7 letters in average) and surface frequency (respectively 6.48 and 40.64 occ./million in average). Primes were presented according to two frame durations (SOAs), 48 and 66ms to examine the time-course of priming. Three experimental lists were constructed using a Latin square in order to present each target once only.
Twenty-five students at the University of Toulouse participated in the experiment. All the participants were native speakers of French and their average age was 26 (7.23 sd). They were all right handed and had normal to corrected-to-normal vision. The experiment lasted around 40 minutes and in exchange for their time, participants received a 4 Giga USB key.

The results are presented in Table 1. As we didn't find any effect of the frame duration, we decided to present the averaged RTs.
Mistaken answers were not considered for the statistical analysis (2.8% of the data), neither were reaction times lower than 250ms and over 1500ms (1% of the data). Cut-offs for the rest of the data were set to 2.5 standard deviations from general average and outliers were removed (1.4%).

Table 1: mean reaction times across the three priming conditions and the two targets conditions. Net priming effects are expressed in ms.

		RTs	Net priming effects (U-M/O-M and U-O)
HighSF primes- LowSF targets	M primes	613	+45* / + 36*
	O primes	649	+9
	U primes	658	
LowSF primes- HighSF targets	M primes	572	+22 / +18
	O primes	590	+4
	U primes	594	

* : p < .05

The results show a clear pattern of a morphological facilitation effect (reaction times decreases when the prime-target relationship is morphological, compared to orthographic and unrelated control conditions).
A significant difference across conditions can be observed only when the target word has a lower frequency than the primes. Statistical analysis showed that the critical net priming effects (difference between the reaction times for morphological primes against orthographic and unrelated control ones) for HighSF primes - LowSF targets was of 45* and 36*ms (respectively).
When looking at the LowSF-primes and HighFs targets the RTs differences of the net priming effects previously described, where not statistically significant Morphological facilitation effects seem to be larger when the frequency of the prime is higher than the frequency of the target, regardless of the SOA used.

2. Conclusion

The results of the present study are in line with the previously found by Giraudo and Grainger (2000), showing differential priming effects when the surface frequency of the prime is manipulated. The absence of a morphological priming effect in the High frequency M-primes/Low frequency targets contrasted with the strong significant priming effects obtained with the Low frequency M-primes/High frequency targets, suggests competition effects to the detriment of an obligatory decomposition process. According to this view both low and high frequency targets should have benefit from the prior presentation of a morphologically related word, but the results revealed this was not the case. Only base targets having a surface frequency lower than the surface frequency of their derivation were significantly facilitated relative to both the orthographic and the unrelated conditions (+45 and +36ms). We interpret these data as an evidence of a competition process among the word forms at the word level.

References

Amenta, S. & Crepaldi, D. (2012). Morphological processing as we know it: An analytical review of morphological effects in visual word identification. *Frontiers in Language Sciences, 3*, 232.

Aronoff, M. (1994). *Morphology by itself.* Cambridge: MIT Press.

Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.

Boudelaa, S., & Marslen-Wilson, W. D. (2001). Morphological units in the Arabic mental lexicon. *Cognition*, 81, 65-92.

Duñabeitia, J.A., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition, 105*, 691-703.

Duñabeitia, J.A., Laka, I., Perea, M., & Carreiras, M. (2009). Is Milkman a superhero like Batman? Constituent morphological priming in compound words. *The European Journal of Cognitive Psychology,* 21(4), 615 – 640.

Drews, E. & Zwitserlood, P. (1995). Morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance,* 21, 1098-1116.

Forster, K.I., & Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 680-698.

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.

Giraudo, H. & Grainger, J. (2000). Effects of Prime word frequency in masked morphological and orthographic priming. *Language and Cognitive Processes,* 15, 421-444.

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin & Review,* 8(1), 127-131.

Giraudo, H. & Voga, M. (2007). Lexeme-Based Model vs. Morpheme-Based Model from Psycholinguistic Perspectives. In F. Montermini, G. Boyé, and N. Hathout (Eds.), *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, pp. 108-114. Somerville, MA: Cascadilla Proceedings Project.

Giraudo, H. & Voga, M. (2014). Measuring morphology: the tip of the iceberg? A retrospective on 10 years of morphological processing, *Carnets de Grammaire, 22*, xxx-xxx.

McCormick, S. F., Rastle, K. & Davis, M. H. (2009). Adore-able not adorable? Orthographic underspecification studied with masked repetition priming. *European Journal of Cognitive Psychology, 21,* 813-836.

McCormick, S. F., Brysbaert, M., & Rastle, K. (2009). Is morphological decomposition limited to low-frequency words?. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1706-1715.

Marantz, A. (2013) No escape from morphemes in morphological processing, *Language and Cognitive Processes, 28:7*, 905-916

Marslen-Wilson, W.D., Ford, M., Older, L., & Zhou, X. (1996). The combinatorial lexicon: Priming derivational affixes. In G. Cottrell (Ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, (pp. 223-227). Mahwah, NJ: Lawrence Erlbaum Associates.

New B., Pallier C., Ferrand L., Matos R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE, *L'Année Psychologique, 101*, 447-462.

Rastle, K., Davis, M.H., Marslen-Wilson, W.D., & Tyler, L.K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes 15 (4-5)*, 507-537.

Rastle, K., & Davis, M.H. (2003). Reading morphologically complex words: some thoughts from masked priming. In S. Kinoshita & S.J. Lupker (Eds.), *Masked priming: State of the art* (pp. 279-305). New York: Psychology Press.

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

Rastle, K. & Davis, M.H. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes, 23*, 942-971.

Seidenherg, M.S. (1987). Sublexical structures in visual word recogni- tion: Access units or orthographic redundancy? In M. Coltheart (Ed.), *Attention and performance XII: Reading* (pp. 245-263). Hills- dale, NJ: Erlbaum.

Voga, M. & Grainger, J. (2004). Masked Morphological Priming with Varying Levels of Form Overlap: Evidence from Greek Verbs. *Current Psychology Letters: Behaviour, Brain & Cognition,* 13(2).

Language proficiency moderates morphological priming in children and adults

Jana Hasenäcker
Max-Planck-Institute
for Human Development,
Berlin
hasenaecker@mpib-berlin.mpg.de

Elisabeth Beyersmann
Laboratoire de Psychologie
Cognitive, Aix-Marseille
Université and Centre de la
Recherche Scientifique,
Marseille
lisi.beyersmann@gmail.com

Sascha Schroeder
Max-Planck-Institute
for Human Development,
Berlin
sascha.schroeder@mpib-berlin.mpg.de

1 Introduction

A number of studies have shown that skilled readers decompose morphologically complex words upon encountering them (for a review, see Rastle & Davis, 2008). It has been proposed that this segmentation process is early and automatic and is driven by orthographic form, while being blind to semantic content, thus also called *morpho-orthographic* (Rastle, Davis, & New, 2004; Taft, 2003). One key finding in favor of this proposition comes from masked morphological priming: the recognition of a target word is facilitated when it is preceded by a morphologically related word prime (*teacher-TEACH*). Facilitation has also been found in a number of languages for targets preceded by pseudocomplex word primes that is words that appear to have a morphologically complex structure, but are simplex words (*corner-CORN*). Moreover, facilitation has as well been observed from complex pseudoword primes, that is a non-existing combination of a stem and affix (*flexify-FLEX*). For non-morphological nonword primes, that is a non-existing combination of a word and a non-morphemic ending (*flexint-FLEX*), mixed results have been obtained (Longtin & Meunier, 2005; Morris, Porter, Grainger & Holcomb, 2011). Recent evidence from French points to a moderating role of language proficiency: the magnitude to which morpho-orthographic information is used increases as a function of individual vocabulary and spelling skills in adults (Andrews & Lo, 2013; Beyersmann, Casalis, Ziegler & Grainger, 2014).

Only few studies have been concerned with morphological decomposition in beginning readers. The few studies from English and French used complex word primes, pseudocomplex word primes and non-morphological word primes. Quémart, Casalis and Colé (2011) found priming in French grade 3, 5, and 7 children from complex as well as pseudocomplex words, thus suggesting that children already use adult-like decomposition processes. In contrast, Beyersmann, Castles and Coltheart (2012) only found priming from truly complex words in grade 3 and 5 English-speaking children, indicating that morpho-orthographic priming is not automatized yet and decomposition relies more on semantics in developing readers. However, no studies with children have used complex pseudoword primes so far, although they provide the possibility to utilize the paradigm in languages that do not naturally have pseudocomplex words, such as German.

Morphological decomposition in German can be insightful to investigate, because of its language specific characteristics. German has a transparent orthography and is morphologically rich. As a consequence, morphological entities might present a very useful unit for effective word recognition, even for beginning readers. Nevertheless, for children being still in the process of reading acquisition and showing more variability in their lexical representations, language proficiency can be expected to play an even greater role than Beyersmann et al. (2014) found for adults.

The aim of the present study was therefore to test whether the moderating effect of lan-

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

guage proficiency, as indexed by vocabulary and spelling skills, on morphological priming can be replicated with German adults and whether it generalizes to readers at the lower end of the proficiency range, namely children. We expect to see evidence for a more automatized form of morpho-orthographic decomposition in highly proficient children (replicating the Quémart et al. pattern), whereas low-skilled children should show less priming (as in Beyersmann et al., 2012) or no robust priming at all. In our adult group, we expect robust priming in all three prime conditions (including the nonsuffixed condition) in high proficiency participants, but reduced non-suffixed priming in low proficiency participants

2 Method

2.1 Participants

Twenty-four university students (13 women, $M_{\text{age}} = 25.2$ years, age range: 20–29 years) and 24 elementary school children (13 girls, $M_{\text{age}} = 9.5$ years, age range: 8;6–10;9 years, grade 3-5) participated in the experiment.

Each participant's language proficiency was assessed, using a spelling and a vocabulary test. Adults performed a spelling recognition test, which was modelled after the one used by Andrews and Lo (2012). Participants were asked to classify 100 words as correctly or incorrectly spelled. Children performed the fill-in-the-gap dictation test of the SLRT-II (Moll & Landerl, 2010). For assessment of vocabulary, adults completed the German version of the LexTALE (Lemhöfer & Broersma, 2012), and children the vocabulary subtest of the CFT 20 (Weiß, 1998). A composite measure of spelling and vocabulary was calculated by standardizing and averaging the spelling and vocabulary scores for each participant.

2.2 Materials

We conducted a masked priming lexical decision experiment using real suffixed words (*kleidchen-KLEID*), suffixed pseudowords (*kleidtum-KLEID*), nonsuffixed pseudowords (*kleidekt-KLEID*) and unrelated controls (*träumerei-KLEID*) as primes. 50 word targets were selected from the childLex corpus (Schroeder, Würzner, Heister, Geyken, & Kliegl, 2014) and 50 pseudoword targets were created by changing one letter from a real word that was not in the target word set. Word and nonword targets were matched on length. For each target all four types of primes were created. Primes were matched on length, suffix length and non-morphemic ending length across conditions. Four counterbalanced lists of prime-target combinations were created, each containing a target only once, such that participants saw each target only with one of the four prime conditions.

2.3 Procedure

Stimuli were presented in white 20-point Courier New font in the center of a black screen on a 15″ laptop monitor with a refresh rate of 60 Hz. Each trial consisted of a 500-ms fixation cross, followed by a 500-ms forward mask of hash keys, then a prime in lowercase for 50 ms, followed by the target in uppercase. The target remained on screen until response. Participants were instructed to indicate whether the presented stimuli was an existing German word or not by pressing a key as quickly and as accurately as possible. They were not informed about the presentation of the prime.

2.4 Results

Reaction times were analyzed using linear mixed-effects modeling. Participants and items were included as random factors and lexical status of the target (word, pseudoword), prime type (suffixed word, suffixed pseudoword, nonsuffixed pseudoword, unrelated word), age group (adults, children) and language proficiency (continuous measure combined of the spelling and vocabulary scores), as well as all their interactions, were included as fixed effects. Where appropriate, one-sided post-hoc contrasts were applied comparing all related priming conditions with the unrelated condition. For contrasting readers with higher and lower proficiency, reaction times of participants scoring one standard deviation above or below the mean proficiency measure within their age group were used. Significance was evaluated using the normal distribution. Results are reported for word targets only. Descriptive statistics are provided in Table 1.

For adults, priming was observed from all three related conditions (suffixed word, suffixed pseudoword and nonsuffixed pseudoword) relative to the unrelated condition, z=5.04, z=4.43, z=2.07, all p<.05. However, language proficiency moderated priming effects. Priming in the nonsuffixed pseudoword condition was only significant for adults with higher language proficiency (+1SD), z=1.74, p<.05, but not for adults

Prime Type	RT Adults	Children	Stimulus Example
	All participants		
Suffixed Word	593 (12)	1024 (36)	*kleidchen* - *KLEID*
Suffixed Nonword	597 (12)	1051 (38)	*kleidtum* - *KLEID*
Nonsuffixed Nonword	614 (13)	1045 (38)	*kleidekt* - *KLEID*
Unrelated	629 (14)	1087 (41)	*träumerei* - *KLEID*
	Higher Language Proficiency (+1SD)		
Suffixed Word	588 (12)	900 (28)	*kleidchen* - *KLEID*
Suffixed Nonword	583 (12)	924 (30)	*kleidtum* - *KLEID*
Nonsuffixed Nonword	602 (12)	903 (28)	*kleidekt* - *KLEID*
Unrelated	620 (13)	974 (33)	*träumerei* - *KLEID*
	Lower Language Proficiency (-1SD)		
Suffixed Word	599 (12)	1189 (48)	*kleidchen* - *KLEID*
Suffixed Nonword	613 (13)	1218 (51)	*kleidtum* - *KLEID*
Nonsuffixed Nonword	626 (14)	1239 (52)	*kleidekt* - *KLEID*
Unrelated	638 (14)	1229 (51)	*träumerei* - *KLEID*

Table 1. Response times (in ms) for children and adults, averaged across items for each participant. Standard errors are presented in parentheses.

with lower language proficiency (-1SD), z=1.16, p=.25.

For children, proficiency played an even more pronounced role than for adults: higher proficiency children (+1SD) showed the same pattern as higher proficiency adults, namely priming from all related condition, z=3.03, z=2.02, z=2.96, all p<.05. In contrast, in lower proficiency children (-1SD) priming in none of the conditions reached significance, although there was a numerical advantage from suffixed word primes in the mean reaction times (40ms faster compared to the unrelated condition).

3 Conclusion

Our results confirm recent evidence for French adults (Beyersmann et al., 2014), showing that the extent to which morphological information is exploited depends on language proficiency also in German. Adults in the present study showed morphological priming effects from suffixed word primes (*kleidchen-KLEID*), suffixed pseudoword primes (*kleidtum-KLEID*) and also nonsuffixed pseudoword primes (*kleidekt-KLEID*) relative to unrelated words (*träumerei-KLEID*). Priming from the nonsuffixed pseudoword condition did not continue to be significant with decreasing language proficiency. Moreover, the pattern of priming generalizes to beginning readers with higher language proficiency: they show priming similar to that of higher proficient adults. For children with lower language proficiency, the effects did not reach significance, but were clearly most pronounced in the suffixed word condition.

We argue that there is a developmental gradient in the use of morphological information during reading acquisition, driven by language proficiency. Beginning readers with low language proficiency are only able to benefit from morpho-semantic information, if at all. More advanced lexical knowledge allows readers to extract morpho-orthographic information. Following Andrews and Davis (1999) and Grainger and Ziegler (2011), we assume that this happens through segmentation of the affix in lower proficiency adults, as indicated by the priming effects of both suffixed prime conditions. Crucially, higher proficiency adult and even child readers with sophisticated lexical knowledge are able to additionally use segmentation of the embedded stem, therefore showing facilitation also in the nonsuffixed prime condition. Our results highlight the importance of lexical knowledge as a further determinant of the ability to exploit morphological structure in the word recognition process, especially in children.

Reference

Andrews, S., & Davis, M. H. (1999). Interactive activation accounts of morphological decomposition: Finding the trap in mousetrap? *Brain and Language, 68*, 355-361.

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

Beyersmann, E., Casalis, S., Ziegler, J. C., & Grainger, J. (2014). Language proficiency and morpho-orthographic segmentation. *Psychonomic Bulletin Review.*

Beyersmann, E., Castles, A., & Coltheart, M. (2012). Morphological processing during visual word recognition in developing readers: Evidence from masked priming. *The Quarterly Journal of Experimental Psychology, 65*, 1306-1326.

Grainger, J., & Ziegler, J. C. (2011). A dual-route approach to orthographic processing. *Frontiers in Psychology, 45,*

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods, 44*, 325-343.

Longtin, C. M., & Meunier, F. (2015). Morphological decomposition in early visual word processing. *Journal of Memory and Language, 53*, 26-41.

Moll, K., & Landerl, K. (2010). *SLRT-II: Lese- und Rechtschreibtest: Weiterentwicklung der Salzburger Lese- und Rechtschreibtests (SLRT): Manual.* Bern: Verlag Hans Huber.

Morris, J., Porter, J. H., Grainger, J., & Holcomb, P. J. (2011). Effects of lexical status and morphological complexity in masked priming: An ERP study. *Language & Cognitive Processes, 26*, 558-599.

Quémart, P., Casalis, S., & Colé, P. (2011). The role of form and meaning in the processing of written morphology: A priming study in French developing readers. *Journal of Experimental Child Psychology, 109*, 478-496.

Rastle, K., & Davis, M. H. (2008). Morphological decomposition based on the analysis of orthography. *Language & Cognitive Processes, 23*, 942-971.

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

Schroeder, S., Würzner, K.-M., Heister, J., Geyken, A., & Kliegl, R. (2014). *childLex: A lexical database for German read by children*. Manuscript submitted for publication. Max Planck Institute for Human Development, Berlin.

Taft, M. (2003). Morphological representation as a correlation between from and meaning. In E. Assink & D. Sandra (Eds.) *Reading complex words* (pp. 113-137). Amsterdam, The Netherlands: Kluwer.

Weiß, R. H. (1998). *Grundintelligenztest Skala 2 (CFT 20) mit Wortschatztest (WS) und Zahlenfolgentest (ZF).* Göttingen: Hogrefe.

Grouping morphologically complex words in the mental lexicon: Evidence from Russian verbs and nouns

Natalia Slioussar
HSE Moscow & St. Petersburg State University
slioussar@gmail.com

Anastasia Chuprina
St. Petersburg State University
a.o.chuprina@gmail.com

1 Introduction

Frequency is known to play a crucial role in lexical access. The notions primarily discussed in the literature are form frequency, (whole) word frequency and morpheme frequency, e.g. root frequency. In numerous studies (Alegre & Gordon, 1999; Baayen & al. 2007, a.m.o.), these characteristics were manipulated to find out whether various word forms are decomposed during lexical access or are stored and can be accessed as a whole. Similar issues arise when we turn from inflection to derivation, at least with semantically transparent derivates (Niswander-Klement & Pollatsek, 2006; Taft 2004, a.m.o.).

2 Our study

Some morphologically complex words were shown to be accessed as a whole (then their own frequency played a crucial role), the others were demonstrated to be decomposed (then root frequency and the frequency of the word they are derived from was important). Both options are available in some models: the one that is more efficient in a particular case wins. However, the picture may be more complex in morphologically rich languages. If a word has many inflectional forms or derivates that are stored as a whole, they probably form groups, and lexical access to this word may depend on the properties of such groups. Our hypothesis is that if a word has a large group of morphologically complex derivates which are relatively semantically transparent, access to and storage of this word would depend on the properties of this group even though the derivates do not necessarily undergo the process of decomposition. We explored this question in our study on Russian.

2.1 Experiment 1

Method. We conducted a lexical decision experiment using *E-Prime* software. Participants were 27 speakers of Russian (age: 19-52 years, 20 female). Materials were 18 triplets of unprefixed imperfective verbs and 12 pairs of unprefixed deverbal nouns. Word frequency, length and CV structure were matched inside triplets and pairs, while the summed frequency of the corresponding prefixed verbs and nouns was different for every verb and noun inside a triplet/pair (as shown in Table 1). Word frequency information was taken from the *The Frequency Dictionary of the Modern Russian Language* (Lyashevskaya & Sharoff, 2009).

word	*letters (in Cyrillic)*	*word F (ipm)*	*summed F of prefixed words*	*group*
torčat' to stick out	7	86,3	2,0	1
dyšat' to breath	6	90,8	29,4	2
platit' to pay	7	89,0	86,3	3
roždenie birth	8	98,5	35,8	1
javlenie apparition	7	94,3	297,5	2

Table 1. An example of stimuli for Exp.1.

It is important to note that prefixed verbs are derived from unprefixed ones, while prefixed deverbal nouns are not (they are derived from prefixed verbs). For verbs, we also counted derivates with the reflexive postfix *-sja*. We made a simplification not taking suffixes into account because, firstly, suffixes change the inflectional class the word belongs to and often cause stress shifts and various alternations, and, secondly, most unprefixed verbs have dramatically more derivates created by prefixation than by suffixation.

In total, every participant saw 54 verbs in infinitive and 24 nouns in nominative singular form, and 78 nonce stimuli. They were shown on

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

the computer screen for 500 ms or until a response button was pressed. If no button was pressed, participants saw a blank screen for up to 2 s. After a response was given or after these 2,5 s were over, an interstimulus interval was initiated and then the next trial began.

Results and discussion. We analyzed participants' question-answering accuracy and reaction times. All participants gave at least 85% of correct answers (92,4% on average); trials with incorrect answers were excluded from further analysis. We also discarded all RTs that exceeded 1,5 s, as is customary in many such studies (e.g. Alegre & Gordon, 1999). In total, 0,3% of reactions to real stimuli were discarded.

We demonstrated that RTs for verbs differ significantly depending on the summed frequency of corresponding prefixed (and postfixed) verbs (repeated measures ANOVA, $F1(2,52) = 8,66$, $p = 0,001$, $F2(2,34) = 4,99$, $p = 0,013$), but RTs for nouns do not. Average RTs for different groups of verbs and nouns are given in Tables 2a and 2b.

group	*av. F (ipm)*	*av. summed F of prefixed words*	*av. RT (ms)*
1	40,1	11,0	643,4
2	41,1	43,5	632,3
3	41,1	139,0	607,6

Table 2a. Average RTs for verb stimuli in Exp.1.

group	*av. F (ipm)*	*av. summed F of prefixed words*	*av. RT (ms)*
1	33,1	60,3	637,8
2	31,9	220,6	635,4

Table 2b. Average RTs for noun stimuli in Exp.1.

We believe that these results can be explained as follows. The majority of Russian prefixed verbs and nouns are likely to be stored as a whole because even relatively transparent ones tend to have some aspects of meaning that cannot be predicted compositionally. Still, prefixed verbs have close connections with their unprefixed counterpart in the mental lexicon due to direct derivational links and therefore influence lexical access to it. Prefixed deverbal nouns are not connected to their unprefixed counterpart in a similar way due to the lack of derivational links, so the summed frequency of such nouns does not influence lexical access to it.

However, an alternative explanation can also be suggested: prefixed verbs are decomposed (and thus boost the frequency of their unprefixed counterpart), while the results for nouns are inconclusive. We chose deverbal nouns for our experiment to find enough relatively transparent prefixed and unprefixed ones, and, if prefixed ones are decomposed, the system should go to the prefixed verb by stripping the suffix rather than to the unprefixed noun by stripping the prefix (*rodit'(v) → porodit'(v) → poroždenie(n)*). To refute this alternative explanation, we designed a follow-up experiment.

2.2 Experiment 2

Method. The method was the same as in Experiment 1. Participants were 24 speakers of Russian (age: 18-55 years, 18 female). Materials included 60 prefixed verb and noun stimuli and 60 nonce words. Real words were chosen from the pool of prefixed verbs and nouns whose unprefixed counterparts were analyzed in Experiment 1. This time both verbs and nouns were grouped in pairs. They were matched in length, CV structure and the frequency of their unprefixed counterparts, but differed in whole word frequency. An example is given in Table 3.

word	*letters (in Cyrillic)*	*word F (ipm)*	*unprefixed word F*	*group*
podyšat' to breath a little	8	7,7	90,8	1
otplatit' to pay back	9	1,7	89,0	2
poroždenie production	10	5,1	98,5	1
projavlenie manifestation	10	45,3	94,3	2

Table 3. An example of stimuli for Exp.2.

Moreover, we took care of the following. If verbs like *podyšat'* 'to breath a little' and *otplatit'* 'to pay back' from Table 3 are accessed as a whole, their word frequency should matter, and *podyšat'* (group 1) will be accessed faster. Now let us assume that they are decomposed, and so are many other prefixed verbs. Then not the word frequency of *dyšat'* 'to breath' and *platit'* 'to pay' will predict the speed of the lexical access, but the frequency of these unprefixed verbs plus the summed frequency of their decomposed derivates. As Table 1 shows, this value is greater for *platit'* than for *dyšat'*, so *otplatit'* (group 2) will be accessed faster. This was true for all other prefixed verb pairs in Experiment 2, so the whole word access and decomposition scenarios always gave different predictions.

We could not find prefixed noun pairs with a similar distribution of frequencies in our materials. However, no approach would predict that they could be decomposed by stripping off their prefix first anyway. So noun stimuli were included mainly to make experimental materials more diverse, they will not let us tease apart different lexical access scenarios.

Results and discussion. We analyzed participants' question-answering accuracy and reaction times. All participants gave at least 85% of correct answers (92,0% on average); trials with incorrect answers were excluded from further analysis. We also discarded all RTs that exceeded 1,5 s. In total, 0,4% of reactions to real stimuli were discarded.

We demonstrated that this time, RTs for verbs and nouns differed depending on their whole word frequencies. The difference was statistically significant both for prefixed verbs (RM ANOVA, $F1(1,23) = 17,87$, $p < 0,001$, $F2(1,17) = 5,98$, $p = 0,026$) and for prefixed nouns ($F1(1,23) = 21,27$, $p < 0,001$, $F2(1,11) = 7,88$, $p = 0,017$). Average RTs for different groups are shown in Tables 4a and 4b.

group	*av. F*	*corresp. unpref. verb from Exp.1*	*av. RT*
1	16,3	low summed F	707,3
2	2,0	high summed F	746,0

Table 4a. Average RTs for verb stimuli in Exp.2.

group	*av. F*	*corresp. unpref. noun from Exp.1*	*av. RT*
1	12,4	low summed F	688,4
2	76,4	high summed F	657,5

Table 4b. Average RTs for noun stimuli in Exp.2.

The results are indicative of the whole word lexical access. We can conclude that prefixed verbs influence lexical access to their unprefixed counterpart not through decomposition, but because they are closely connected in the mental lexicon due to direct derivational links.

3 Conclusion

Using Russian prefixed and unprefixed verbs, we demonstrated that a group of semantically transparent derivates influence the recognition of the word they are derived from. The higher is summed frequency of derivates, the faster is the lexical access. One could argue that this is due to decomposition.We showed that this is not the case.

In two lexical decision experiments we conducted, reactions times to prefixed verbs and deverbal nouns depended on their own frequencies, which points to whole word storage. At the same time, reaction times to unprefixed verbs were influenced by the summed frequency of their derivates (created by prefixation and postfixation). We conclude that this effect is explained not by decomposition of the derivates during lexical access, but by their strong connection to the word they are derived from.

Our conclusion is confirmed by the data from deverbal nouns. On the surface (i.e. phonologically), the overlap between unprefixed and prefixed verbs on the one hand and unprefixed and prefixed nouns on the other hand is the same: as examples from Tables 1 and 3 show, they coincide once the prefix is stripped. If this factor played a role, the results for unprefixed verbs and nouns would be the same.

However, reactions times to unprefixed nouns are not influenced by the summed frequency of their prefixed counterparts. This proves that connections through derivational links matter. Prefixed deverbal nouns are derived from prefixed verbs, not from unprefixed nouns (*porodit'(v)* 'to give birth, to generate' → *poroždenie(n)* 'production', not *roždenie(n)* 'birth' → *poroždenie(n)* 'production'). Phonologically, prefixed nouns resemble unprefixed ones much more than prefixed verbs, but this does not play a role.

In total, our results can be taken as a piece of evidence for a new type of frequency information to be taken into account. Somewhat similar conclusions were reached by Moscoso del Prado Martín et al. (2004) who studied morphological family size effects in Finnish compared to Dutch and Hebrew.

Of course, many things remain to be explored. As we noted earlier, we did not look at suffixation. We did not specify the mechanisms by which derivationally related forms are connected in the mental lexicon and how these connections are formed. In the connectionist approach where no decomposition is assumed, regular connections between words' phonological forms and meanings should matter. In dual route models, it can be suggested that decomposition normally does not win in some cases like derived verbs and nouns we analyzed, but still takes place. Then only the existence of a direct derivational link and, probably, semantic transparency should really matter.

To solve these and other problems, many crucial questions need to be answered. Which derivates 'boost' the frequency of a base word and to what extent? What is the role of semantic transparency and phonological similarity between a derivate and its base form? How important is it for their connection whether they belong to one part of speech or to one inflectional class? Would stress shifts and alternations influence our results? We hope to address some of these questions in our further research.

References

Alegre, M., & Gordon, P. 1999. Frequency effects and the representational status of regular inflections. Journal of Memory and Language, 40, 41-61.

Baayen, H., Wurm, L.H., & Aycock, J. 2007. Lexical dynamics for low-frequency complex words: A regression study across tasks and modalities. The Mental Lexicon, 2, 419-463.

Lyashevskaya, O.N., & Sharoff, S.A. 2009. Chastotnyj Slovar' Sovremennogo Russkogo Jazyka ('The Frequency Dictionary of Modern Russian Language'). Moscow: Azbukovnik.

Moscoso del Prado Martín, F., Bertram, R., Häikiö, T., Schreuder, R., & Baayen, R.H. 2004. Morphological family size in a morphologically rich language: The case of Finnish compared to Dutch and Hebrew. Journal of Experimental Psychology: Learning, Memory and Cognition, 30, 1271–1278.

Niswander-Klement, E., & Pollatsek, A. 2006. The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. Memory and Cognition, 34, 685-702.

Taft, M. 2004. Morphological decomposition and the reverse base frequency effect. Quarterly Journal of Experimental Psychology, 57A, 745-765.

Extraction and Analysis of Proper Nouns in Slovak Texts

Radovan Garabík
Ľ. Štúr Institute of Linguistics
Slovak Academy of Sciences
Bratislava, Slovakia
garabik@kassiopeia.juls.savba.sk

Radoslav Brída
Ľ. Štúr Institute of Linguistics
Slovak Academy of Sciences
Bratislava, Slovakia
brida@korpus.sk

1 Introduction

Unknown named entity recognition in inflected languages faces several specific problems – the first and foremost is that the entities themselves are inflected[1] (Dvonč et al., 1966) leading to a problem of identifying word forms as belonging to the same lexeme, and also the problem of finding correct lemma. In this article we analyse the distribution of word forms for proper nouns in Slovak and describe an algorithm for their automatic extraction and lemmatisation.

The task of lemmatisation and morphological annotation of flective (and more specifically, Slavic) languages is reasonably researched and developed (Hajič, 2004). Since we cannot expect a morphological database (data relating lemmata to inflected word forms and their grammatical tags) to cover all or almost all the words present in the corpus (*especially* proper names that keep appearing depending on who or what has become a hot topic in mass media), using a well tuned guesser can improve the accuracy of lemmatisation and tagging.

Common sense says that named entities (proper names in particular) behave differently from common names, which translated into information theory terms means that the information about whether a word is a proper name is not independent from the information about its morphology paradigm. This means we can use the information about proper names to decrease the entropy of inflections, which is good because it helps the guesser choose between the possible lemmata and morphological tags.

2 Datasets

We denote Levenshtein distance (Левенштейн, 1965) between two words l and w by $\rho(l, w)$. Since a typical Slovak noun has up to 12 different word forms (two numbers, six cases – the vocative is rare), and the inflection is mostly realized by changing the suffix and root vowel alteration, we can expect the overall distance between lemma and its word forms to be not only bounded from above, but also have a regular distribution (roughly speaking, the less typical the suffix length, the less likely is such a word form to appear).

We used the morphological database of Slovak language (Garabík and Šimková, 2012; Karčová, 2008; Garabík, 2007), which contains (at the time of writing) complete morphological information of 35 009 nouns (lemmata), out of which 1031 are proper nouns. We randomly divided the database into two parts, the training set and the evaluation set, ensuring that about 90% of both common and proper nouns is present in the training set. The evaluation set contained 101 lemmata and 694 unique word forms for proper nouns.

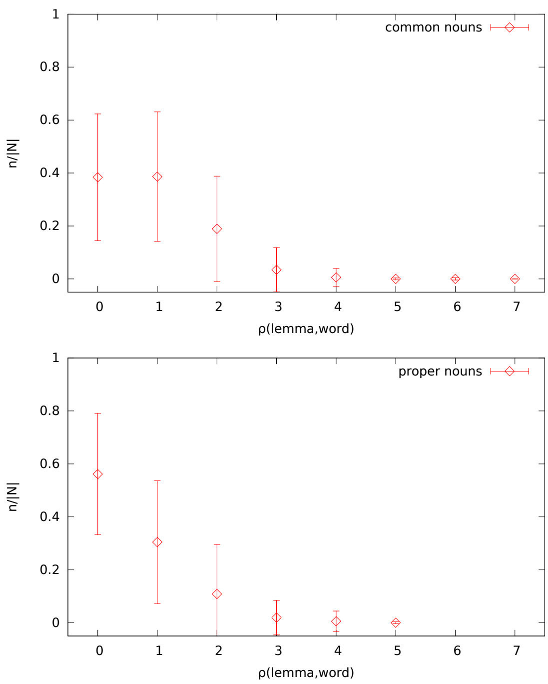


Figure 1: Distribution of word forms according to their Levenshtein distance from lemma.

[1] e.g. for the lemma *Galileo*, genitive would be *Galilea*, dative *Galileovi* etc.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

…	† Toska	Toskala	Toskalu	Toskánec	‡ Toskánska	Toskánske	‡ Toskánsko	‡ Toskánskom	‡ Toskánsku	† Toske	† Tosky	Toso	…
	10	33	28	20	221	11	110	20	304	15	26	16	

Table 1: Alphabetic list of proper noun candidates, with number of occurances in the corpus. Note the extracted lemmata/lexemes $Toska^{\dagger}$ (La Tosca), $Tosk\acute{a}nsko^{\ddagger}$ (Tuscany), as well as unrelated $Toskala$, $Toso$ and related $Tosk\acute{a}nec$ (inhabitant of Tuscany) and $Tosk\acute{a}nske$ (Tuscan, adjective).

Fig. 1 displays the distribution of known common (top) and proper (bottom) nouns, summed and normalized through all the nouns in the training set. Vertical error bars display the standard deviation for the given distance of word form from the lemma. From the graphs, we derive several conclusions – proper nouns are "less inflected", higher ratio of them is in the basic form (lemma), and the maximum distance is $\rho = 7$ for common nouns (nouns with greater distance are those with very irregular declension, e.g. $\check{c}lovek \rightarrow ľudia$ "human/humans") and $\rho = 5$ for proper nouns. Distributions of common and proper nouns from the evaluation set match those from the training set, so there appears to be some difference between common and proper nouns globally. However, categorising single nouns using these differencies between distributions is not reliable.

3 Extracting Candidates

Our algorithm extracts plausible candidates for proper nouns (those beginning with a capital letter but not at the beginning of a sentence, together with some additional filters) and for each candidate, it considers the set of words with $\rho \leq 5$. This would require calculating the Levenshtein distance between all pairs of words in the set and the complexity would be $O(n^2)$, which is unacceptable for corpus sized inputs. Unfortunately, Levenshtein distance is a metric but cannot be used to make an ordered set out of a list of words (in particular, it cannot be used to define an ordering binary relation $\leq$).

However, a trick can be applied – in a lexicographically ordered list of words (see Table 1) we need to look only at some interval around the word; word forms from beyond the interval are very unlikely to belong to the same lexeme. The complexity will be $O(Cn)$; where C is the (constant) size of the interval. This means that for some of the nouns not all word forms will be covered; especially for the shorter ones, where there is a higher probability that many unrelated words will be within the interval. Empirically we estimated the reasonable interval width to be 2000 words – increasing it above this number does not improve the accuracy anymore and the speed is acceptable. It should be noted that this interval is *not* a width of the context of the concordance – this is an interval in the lexicographically ordered set of proper noun candidates extracted from a given text, e.g. from a novel if we want to extract the whole inflectional paradigms of (new, unknown) proper nouns from the novel, or indeed from the whole corpus, if we aim to augment a morphological database.

We formally describe a Levenshtein edit operation $e = (o, i_s, i_d)$ – a triple of operation type o, position i_s in the source string s and position i_d in the destination string d, where operation type o is one of *replace*, *insert* or *delete*. For *replace* or *insert*, the replacement/new character is taken from the destination string d.

Sequence of edit operations $q = (e_1, e_2, e_3, ...)$, together with the destination string d, when applied to a string $s \in \mathbb{S}$ defines a mapping $f_{q,d} : \mathbb{S}_{q,d} \mapsto \mathbb{S}$, where $\mathbb{S}_{q,d}$ and $\mathbb{S}$ are sets of strings.[2]

If we denote by t a morphological tag for a given word form w, then for a lexeme with a lemma l a tuple (w, t) unambiguously refers to one inflected word form and its grammatical categories. We can then construct a sequence of edit operations leading from l to w, denoted by $q(l, t)$.

For each proper noun from the training set, we precompute the functions $f_{q(l,t),l}$ (this can be improved by dividing the nouns into categories based on their declension rules and using only one noun from each category), to get the sequence of operations leading from the lemma to the tuple (w, t) of the word form and morphological tag. Then, for each extracted word, we apply the functions $f_{q(l,t),l}$ to every word from the abovementioned interval and the word with greatest coverage (sum of the frequencies of generated word forms within the interval) is declared the lemma to the extracted word. Of course, this maximum can be attained by more than one word, especially if the lexeme is incom-

[2] It is not possible to define the function f for every source string, since some of the operations might not be applicable to the given strings.

plete. We assume that at least the most common inflectional paradigms (used for proper nouns) are present in the training set.

word forms	[%]	number of lemmata assigned per word form correct	all
100	18.9	0	1
4	0.8	0	2
418	79.2	1	1
6	1.1	1	2
Σ 528	100.0		

Table 2: Number of automatically assigned lemmata per word form.

4 Evaluation

We used the algorithm to extract proper nouns from the Slovak National Corpus, version *prim-6.1-public-all*[3], of the size 829 million tokens, and evaluated the results on the proper nouns from the evaluation set. The percentage of correctly automatically assigned lemmata is shown in Table 2 – we see that 79.2% word forms had been assigned a unique lemma, which was also the correct one, while 18.9% had been assigned a unique, but incorrect lemma[4].

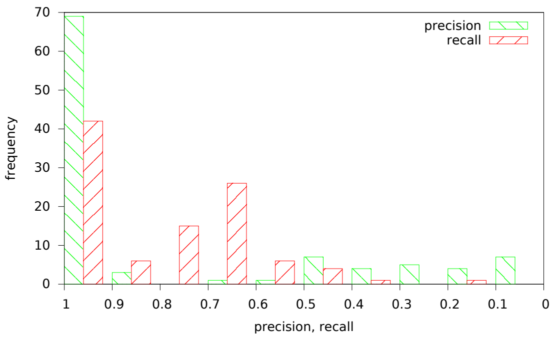


Figure 2: Histogram of precision and recall on automatically assigned word forms of the lexeme(s) for the evaluation data.

Figure 2 displays the precision and recall on word forms for proper nouns (i.e. how much of the lexeme has been extracted; the numbers are not weighted by the frequency of word forms in the corpus) from the evaluation set; we note that about 70 lexemes[5] have $precision \approx 1$; about 40 lexemes have $recall \approx 1$, and about 50 lexemes have $0.9 \gtrapprox recall \gtrapprox 0.6$, while only a small number of lexemes have lower precision. The lower recall is caused by insufficient data coverage – not all the word forms were present in the analysed corpus. The precision we obtained is excellent and the accuracy of automatic lemma assignment is good.

5 Augmenting Morphological Database

The abovementioned process was used to increase the number of proper nouns in Slovak morphological database. We used the extracted candidates from the *prim-6.1-public-all* corpus with a number of occurrences at least 100 (count of all possible word forms derived from a given lemma). We calculate the coverage of word forms for one lemma as $r = C(w,t)/C(g)$, where $C(w,t)$ is the number of generated tuples of word forms and their corresponding morphological tags, and $C(g)$ the number of grammar categories (usually 7 or 14; 7 cases including the vocative and one or two grammatical numbers, with many proper nouns present only in singular).

After removing generated word forms with no corpus evidence, the average coverage of word forms per lemma is $r = 0.84 \pm 0.23$, i.e. 84% of word forms is present in the corpus, 0.23 is the standard deviation of the coverage. Generated word forms still contain a lot of noise, therefore we also removed those word forms whose contribution to the number of occurrences of given lemma was less than 1% (it is rare for a grammatical case to have such a low percentage compared to other cases). After this, the coverage changed to $r = 0.75 \pm 0.24$, where again 0.24 is the standard deviation of the coverage. Then we manually proofread, corrected and filled in the word forms for the several hundred most frequent lemmata. After adding these words to the morphological database, we iterated the process, re-training the algorithm and generating another list of less frequent proper nouns.

6 Conclusion

The method has been used to improve the coverage of proper nouns in the Slovak morphological database and is used as a part of morphological guesser, providing candidate lemmata and morphological tags for unknown proper nouns, as part of the morphosyntactic analysis and part of speech tagging of the Slovak National Corpus.[6]

[3]http://korpus.juls.savba.sk/res.html

[4]For 13 word forms (2.5%) the correct lemma was not present in the interval of 2000 words.

[5]Since the number of proper nouns in our evaluation set was 101, these numbers are fortuitously almost identical to percentage.

[6]http://korpus.juls.savba.sk

References

[Левенштейн1965] Владимир Иосифович Левенштейн. 1965. Двоичные коды с исправлением выпадений, вставок и замещений символов. *Докл. АН СССР*, 4(163):845–848.

[Dvonč et al.1966] Ladislav Dvonč, Gejza Horák, František Miko, Jozef Mistrík, Ján Oravec, Jozef Ružička, and Milan Urbančok. 1966. *Morfológia slovenského jazyka*. Vydavateľstvo SAV, Bratislava, Slovakia, 1st edition. 895 p.

[Garabík and Šimková2012] Radovan Garabík and Mária Šimková. 2012. Slovak Morphosyntactic Tagset. *Journal of Language Modelling*, 0(1):41–63.

[Garabík2007] Radovan Garabík. 2007. Slovak morphology analyzer based on Levenshtein edit operations. In M. Laclavík, I. Budinská, and L. Hluchý, editors, *Proceedings of the WIKT'06 conference*, pages 2–5, Bratislava. Institute of Informatics SAS.

[Hajič2004] Jan Hajič. 2004. *Disambiguation of Rich Inflection (Computational Morphology of Czech)*. Karolinum, Charles Univeristy Press, Prague, Czech Republic.

[Karčová2008] Agáta Karčová. 2008. Príprava a uskutočňovanie projektu morfologického analyzátora. In Anna Gálisová and Alexandra Chomová, editors, *Varia. 15. Zborník materiálov z XV. kolokvia mladých jazykovedcov*, pages 286–292, Bratislava. Slovenská jazykovedná spoločnosť pri SAV – Katedra slovenského jazyka a literatúry FHV UMB v Banskej Bystrici.

Mapping the Constructicon with SYMPAThy. Italian Word Combinations between fixedness and productivity

Alessandro Lenci
Gianluca E. Lebani, Marco S. G. Senaldi
University of Pisa
alessandro.lenci@ling.unipi.it
gianluca.lebani@for.unipi.it
marco.senaldi@sns.it

Sara Castagnoli
Francesca Masini
University of Bologna
s.castagnoli@unibo.it
francesca.masini@unibo.it

Malvina Nissim
University of Groningen
m.nissim@rug.nl

Abstract

This work introduces SYMPAThy, a data representation model in which the combinatorial properties of a lexical item are described by merging surface and deeper linguistic information. The proposed approach is then evaluated by comparing, for a sample list of verbal idioms, a set of SYMPAThy-based fixedness indexes against the relevant speaker-elicited indexes available in the descriptive norms collected by Tabossi et al. (2011).

1 Word combinatorics and constructions

By "Word Combinations" (WoCs) we broadly refer to the range of constructions typically associated with a lexical item. In Construction Grammar, constructions (Cxn) are conventionalized form-meaning pairings that can vary in both complexity and schematicity (Fillmore et al., 1988; Goldberg, 2006; Hoffmann and Trousdale, 2013). The Constructicon spans from fully specified structures (*kick the bucket*) to complex, productive abstract structures such as argument patterns (e.g., the Ditransitive Cxn "Subj V Obj1 Obj2", *she baked him a cake*), passing through "intermediate" Cxns with different degrees of schematicity, complexity and productivity (e.g., *take* Obj *for granted*), in what is known as the lexicon-syntax continuum. WoCs thus comprise so-called Multiword Expressions (MWEs), i.e. a variety of recurrent expressions acting as a single unit at some level of linguistic analysis, like phrasal lexemes, idioms, collocations (Calzolari et al., 2002; Sag et al., 2002; Gries, 2008), as well as the preferred distributional properties of a word at a more abstract level, i.e. argument structures and selectional preferences (Goldberg, 1995).

Each lexeme can thus be described as having a *combinatory potential* to be defined and observed at a more constrained, surface POS-pattern level (P-level) *and* at the more abstract level of syntactic structure (S-level). These two levels are often kept separate, not only theoretically, but also computationally, as their performance varies according to the different types of combinations that we want to track (Sag et al., 2002; Evert and Krenn, 2005).

We advocate a unified and integrated view of a lexeme's combinatory potential, in order to capture both fixed combinations (MWEs of various types) and more productive aspects of the lexeme's distributional behaviour. The theoretical premises lie in the constructionist view of the mental lexicon outlined above, whereas a proposal for a computational implementation is illustrated here. Specifically, we i) present SYMPAThy, *a model of data representation* that takes into account both surface and deeper linguistic information; ii) develop and test an *index of productivity* for Italian WoCs based on SYMPAThy.

2 SYMPAThy: a joint approach to WoCs

We argue that to obtain a comprehensive picture of the combinatory potential of a word and enhance extracting efficacy for WoCs, the P-based approach (which exploits sequences of POS-patterns and association measures) and the S-based approach (which exploits syntactic dependencies and association measures) should be combined. We illustrate this point with an example based on the Target Lexeme (TL) *gettare* 'throw' (V).[1]

We want to use S-based methods to capture the fact that V occurs typically within some syntactic Frames and not others, that for each Frame we have typical Fillers (lexical items) instantiating Frame slots, and that each slot is associated with certain semantic (ontological) classes:[2]

[1]All data is from a version of the "la Repubblica" corpus (Baroni et al., 2004) POS tagged with the Part-Of-Speech tagger described in Dell'Orletta (2009) and dependency parsed with DeSR (Attardi and Dell'Orletta, 2009).

[2]Data extracted by LexIt (Lenci, 2014). The list is partial: only the first three Frames are included; Frames with the re-

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

- subj#obj#comp-su
  - OBJ Filler: {*acqua, ombra, benzina, ...*}; {Substance, Natural_Phenomenon, ...}
  - COMP-su Filler: {*fuoco, tavolo, bilancia, lastrico, istituzione, ...*}; {Artifact, Substance, ...}
- subj#obj#comp-in
  - OBJ Filler: {*scompiglio, sasso, corpo, fumo, cadavere, ...*}; {Natural_Object, Substance, ...}
  - COMP-in Filler: {*panico, caos, sconforto, mare, stagno, cestino, ...*}; {Feeling, State, ...}
- subj#obj
  - OBJ Filler: {*spugna, base, ombra, acqua, luce, ponte, ...*}; {Substance, Artifact, ...}

At this point, we observe that all these words are typically associated with our TL, but we don't know in which way they are all linked to one another. For instance, we have no elements for thinking that subj#*gettare*#*acqua*#*su_fuoco* is any different from subj#*gettare*#*acqua*#*su_tavolo* or subj#*gettare*#*ombra*#*su_istituzione*. However, while *gettare acqua sul fuoco* 'defuse' is an idiom in Italian, *gettare acqua sul tavolo* only has a literal meaning ('throw water on the table'); subj#*gettare*#*fango*#*su_istituzione* is yet different, since *gettare fango su* 'defame' is a fixed expression, but the Filler *istituzione* 'institution' is just one of many possibilities, so the expression is partially fixed, resulting in something like [*gettare fango su* PERSON/INSTITUTION]. The significance of *gettare acqua sul fuoco* with respect to *gettare acqua sul tavolo* emerges much more clearly if we use a P-based method. Extracting surface material, the former expression will be ranked higher than the latter (given the pattern "V N PREPART N") as the association between all words is stronger.

So, fine-grained differences do not emerge with the S-method, while the P-based method fails to capture the higher-level generalizations we get with the S-method. In order to get the best of both worlds, we extracted corpus data into SYMPAThy (SYntactically Marked PATterns), a database where information on both levels is stored and accessible jointly:

- syntactic frames with argument slots and fillers;
- linear order of all elements for each TL;
- POS tag for each element (simple preposition vs. preposition with article, definite vs. indefinite article, modal vs. full verb, etc.);
- morphosyntactic features: gender, number, finiteness, tense, etc.

flexive form *gettarsi* 'throw oneself' and objectless forms are excluded.

3 WoC fixedness with SYMPAThy

Since constructions span along a continuum between fixedness and productivity, there have been various attempts at measuring how fixed a given WoC is, mostly based on surface features. Nissim and Zaninello (2011) assess the fixedness of a subset of complex nominals by comparing inflected and lemmatized forms, and taking into account the proportion of elements that undergo variation in a given MWE. Inflection is also used by Squillante (2014) on noun-adjective expressions, and is combined with two other measures, interruptibility and substitutability. Zeldes (2013) extends Baayen's morphological productivity approach to argument structure and estimates the productivity of a syntactic slot from the number of its hapax noun fillers. Wulff (2009) uses a set of morphosyntactic indexes of variations and a collocation-based index of compositionality as variables in a regression study to determine fixedness.

We extend the state of the art of the quantitative approach to construction fixedness by exploiting the potentialities of SYMPAThy to develop a series of corpus-based indexes able to describe the fixedness of some idiomatic expressions. Our approach is then evaluated by comparing, for a sample list of expressions, a composition of our indexes against the behavioral judgments of syntactic flexibility collected by Tabossi et al. (2011).

3.1 The combinatory behaviour of a TL

In the SYMPAThy model, the combinatory space of a Target Lexeme is assumed to be formed by a network of Cxns, varying for their degree of fixedness/productivity. For any given TL such a representation is built by means of the following four-step procedure:

1. its SYMPAThy patterns are extracted from a reference corpus;
2. the set of single and multiple slot Cxns that TL combines with are semi-automatically identified. An example for the verb *gettare* is reported and explained in Appendix 1;
3. each construction is associated with a *variational profile* formed by a number of statistics extracted from the SYMPAThy pattern to estimate: i) the variability of the fillers that instantiate the syntactic slots of constructions; ii) the

morphological variability of the constructions' components; iii) the variability with respect to determiners; iv) the variability with respect to adjectival and adverbial modifications; v) the variability in the linear order.

4. variational profiles are then used to measure the lexical, morphological and syntactic *degrees of freedom* of Cxns, providing a multidimensional quantitative characterization of their level of fixedness.

3.2 Entropy-based Cxn fixedness modeling

In what follows, we devise a way to encode the variation possibilities shown by Cxns, as well as a meaningful way to combine them. Specifically, we distinguish a series of dimensions of variation and propose to exploit Entropy (Shannon, 1948) to measure how fixed is the behavior of a Cxn in a given dimension.

Entropy is a measure of randomness, calculated as the average uncertainty of a single variable:

$$H(X) = -\sum_{x \in X} p(x) \log_2(p(x)) \tag{1}$$

This measure of randomness can be adapted to our needs by taking the variable X as being a Cxn of interest, and the states of the system x as its values on one dimension of variation. Lower entropy values are to be understood as evidence of fixedness, while higher values suggest a more variable distribution of the states of a given variable, i.e. the target construction tends to be freer.

Observed entropy values, however, can span from 0 to the logarithm of the number of values that X can assume. As a consequence, entropy values related to different dimensions of variation are not comparable, and cannot be combined into a single fixedness index. We overcome this limitation by following Wulff (2008) and describing the randomness of each variability dimension in terms of relative entropy, computed as the ratio between the observed entropy from eq.1 and the maximum entropy H_{max} for the variable X:

$$H_{rel}(X) = \frac{H(X)}{H_{max}(X)} = \frac{H(X)}{\log_2(|X|)} \tag{2}$$

This measure, that ranges from 0 to 1, has been employed as a flexibility measure to describe the flexibility of a given set of target Cxns along the following dimensions of variation:

LEXICAL VARIABILITY. The entropy of the lexical instantiation of the slot positions of a Frame is calculated by assuming that the states x of the random variable X are all the possible fillers that can instantiate a given slot in Cxn (e.g. in subj#*gettare*#obj:*luce*#su_X, X can be filled by *vicenda* 'matter', *mistero* 'mystery', etc.).

MORPHOLOGICAL VARIABILITY. It is calculated as the entropy of the morphological features manifested by the fillers of a Cxn (e.g., *gettare*#*ombra*-<u>*fs*</u> 'cast shadow-singular'; *gettare*#*ombra*-<u>*fp*</u> 'cast shadow-plural').

ARTICLES VARIABILITY. This index encodes how variable is the presence or absence of articles determining the available slots in a Cxn, and, if appropriate, their type (DEFinite vs. INDefinite): for instance, *gettare*#<u>∅</u>+*acqua*#su_<u>*DEF*</u>+*fuoco*.

PRESENCE OF MODIFIERS. This index encodes how variable is the presence or absence of adjectives, adverbs or prepositional phrases modifying the available slots. In this way, it is possible to account for patterns like:*gettare*#<u>*molta*</u>+*acqua*#su_<u>∅</u>+*fuoco*.

DISTANCE VARIABILITY. This index exploits information on linear order available in SYMPAThy to estimate how variable is the distance in tokens between a TL and the other constituents of a given lexically specified Cxn.

In the experiment reported in the next section, we have combined the single variability measures $H_{rel}(X)$ into an overall flexibility index $F(X)$ corresponding to four possible combinations:

- SUM: $F(X)$ is obtained by summing over all the single $H_{rel}(X)$ values;
- AVERAGE: $F(X)$ is the mean of the single $H_{rel}(X)$ values;
- AVERAGE$_{POS}$: $F(X)$ is the mean of the positive $H_{rel}(X)$ values;
- MAX: $F(X)$ is the highest $H_{rel}(X)$ value.

We leave to future research the investigation of further ways to combine the variability indexes.

4 Evaluation

In order to evaluate our approach, we set out to test if our indexes can mimic the intuitive judgments of native speakers about the fixedness of fully lexically specified constructions. To do so, we selected a subset of the idioms in the norms collected by Tabossi et al. (2011), and tested to what degree the speaker-elicited flexibility judgments available in this repository can be modeled by a composition of our variability indexes.

4.1 The descriptive norms by Tabossi et al.

Tabossi et al. (2011) collected several normative measures for 245 Italian verbal idiomatic expressions. Using a group of 740 Italian speakers, they collected a minimum of 40 elicited judgments for each idiom on several psycholinguically relevant variables.

Among the different kinds of ratings, those concerning syntactic flexibility have been collected by inserting each idiomatic expression in a sentence in which one of the following five syntactic modifications occurred: adverb insertion, adjective insertion, left dislocation, passive and movement. Participants were asked to evaluate, on a 7-point scale, how much the meaning of the idiomatic expression in the syntactically modified sentence was similar to its unmarked meaning as expressed in a paraphrase prepared by the authors.

4.2 Data extraction

Out the 245 expressions in Tabossi et al., we selected the 23 target idioms reported in Appendix 2. Each such idiom can be represented, in our approach, as a fully lexically specified transitive Cxn headed by a given verbal TL, for which the subject slot is underspecified (e.g. *gettare*#obj:*maschera*). We built the variational profiles of our target idioms by adopting an adapted version of the procedure described in Section 3:

1. for each TL, we extracted the SYMPAThy patterns from the "la Repubblica" corpus;
2. the patterns involving one of our target idioms were identified and selected;
3. for each idiom, the variability indexes described in Section 3.2 were calculated. Note that, given the nature of our experimental stimuli, the lexical variability index is not relevant;
4. we built a fixedness index for each idiom, according to the four composition methods in the previous section.

4.3 Results and discussion

In order to test the cognitive plausibility of the fixedness indexes extracted from SYMPAThy, we calculated the Pearson's Product-Moment Correlation strength between them and the syntactic flexibility ratings in Tabossi et al. (2011). Correlation values are reported in Table 1. In all cases, there is a significant ($p < .05$) positive correlation, ranging between .44 and .47, thus supporting the psycholinguistic plausibility of our corpus-based variability indexes.

Combination	***r***
SUM	.44
AVERAGE	.44
AVERAGE $_{POS}$	.46
MAX	.47

Table 1: Pearson's Correlation strength between different combination methods of the SYMPAThy-based fixedness indexes and the syntactic flexibility judgments in Tabossi et al. (2011). All reported values are associated with $p < .05$, $N = 23$.

These results, albeit preliminary, look promising especially given the different nature of the behavioral and corpus-based indexes. On the one hand, the speakers' ratings are semantically driven, since they are thought to model how much the figurative meaning of a given idiom is sensitive to its syntactic form. On the other hand, the automatically corpus-derived information exploited by our indexes does not take meaning into account. SUch indexes describe a lexically specified Cxn that can in principle have an idiomatic as well as a compositional, literal meaning (even if, presumably, the latter case is rare in the corpus).

5 Conclusion

In this study we presented a procedure for characterizing the combinatorial potential of a lexical item and the degree of fixedness of the Cxns it occurs in. Such a procedure has been preliminary tested on a small sample of idiomatic expressions and the resulting representation has been evaluated against the subject-elicited judgments collected by Tabossi et al. (2011). In the future, we are planning to extend the inventory of variability dimensions (addressing also the question of the semantic compositionality of Cxns), to study their relative weight and their interactions, and to develop more sophisticated ways to combine them.

Acknowledgments

This research was carried out within the CombiNet project (PRIN 2010-2011 *Word Combinations in Italian: theoretical and descriptive analysis, computational models, lexicographic layout and creation of a dictionary*, n. 20105B3HE8) funded by the Italian Ministry of Education, University and Research (MIUR).

References

[Attardi and Dell'Orletta2009] Giuseppe Attardi and Felice Dell'Orletta. 2009. Reverse revision and linear tree combination for dependency parsing. In *Proceedings of NAACL 2009*, pages 261–264.

[Baroni et al.2004] Marco Baroni, Silvia Bernardini, Federica Comastri, Lorenzo Piccioni, Alessandra Volpi, Guy Aston, and Marco Mazzoleni. 2004. Introducing the La Repubblica Corpus: A Large, Annotated, TEI(XML)-Compliant Corpus of Newspaper Italian. In *Proceedings of LREC 2004*, pages 1771–1774.

[Calzolari et al.2002] Nicoletta Calzolari, Charles J. Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod, and Antonio Zampolli. 2002. Towards best practice for multiword expressions in computational lexicons. In *Proceedings of LREC 2002*, pages 1934–1940.

[Dell'Orletta2009] Felice Dell'Orletta. 2009. Ensemble system for Part-of-Speech tagging. In *Proceedings of EVALITA 2009*.

[Evert and Krenn2005] Stefan Evert and Brigitte Krenn. 2005. Using small random samples for the manual evaluation of statistical association measures. *Computer Speech & Language*, 19(4):450–466. Special issue on Multiword Expression.

[Fillmore et al.1988] Charles J. Fillmore, Paul Kay, and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language*, 64(3):501–538.

[Goldberg1995] Adele Goldberg. 1995. *Constructions. A Construction Grammar Approach to Argument Structures*. The University of Chicago Press, Chicago.

[Goldberg2006] Adele Goldberg. 2006. *Constructions at work*. Oxford University Press, Oxford.

[Gries2008] Stefan Th. Gries. 2008. Phraseology and linguistic theory: a brief survey. In Sylviane Granger and Fanny Meunier, editors, *Phraseology: an interdisciplinary perspective*, pages 3–25. John Benjamins, Amsterdam & Philadelphia.

[Hoffmann and Trousdale2013] Thomas Hoffmann and Graeme Trousdale, editors. 2013. *The Oxford Handbook of Construction Grammar*. Oxford University Press, Oxford.

[Lenci2014] Alessandro Lenci. 2014. Carving verb classes from corpora. In Raffaele Simone and Francesca Masini, editors, *Word Classes. Nature, typology and representations*, Current Issues in Linguistic Theory, pages 17–36. John Benjamins.

[Nissim and Zaninello2011] Malvina Nissim and Andrea Zaninello. 2011. A quantitative study on the morphology of Italian multiword expressions. *Lingue e Linguaggio*, X:283–300.

[Sag et al.2002] Ivan A. Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and D. Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of CICLing 2002*, pages 1–15.

[Shannon1948] Claude E. Shannon. 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379 – 423.

[Squillante2014] Luigi Squillante. 2014. Towards an empirical subcategorization of multiword expressions. In *Proceedings of the 10th Workshop on Multiword Expressions (MWE)*, pages 77–81, Gothenburg, Sweden, April. Association for Computational Linguistics.

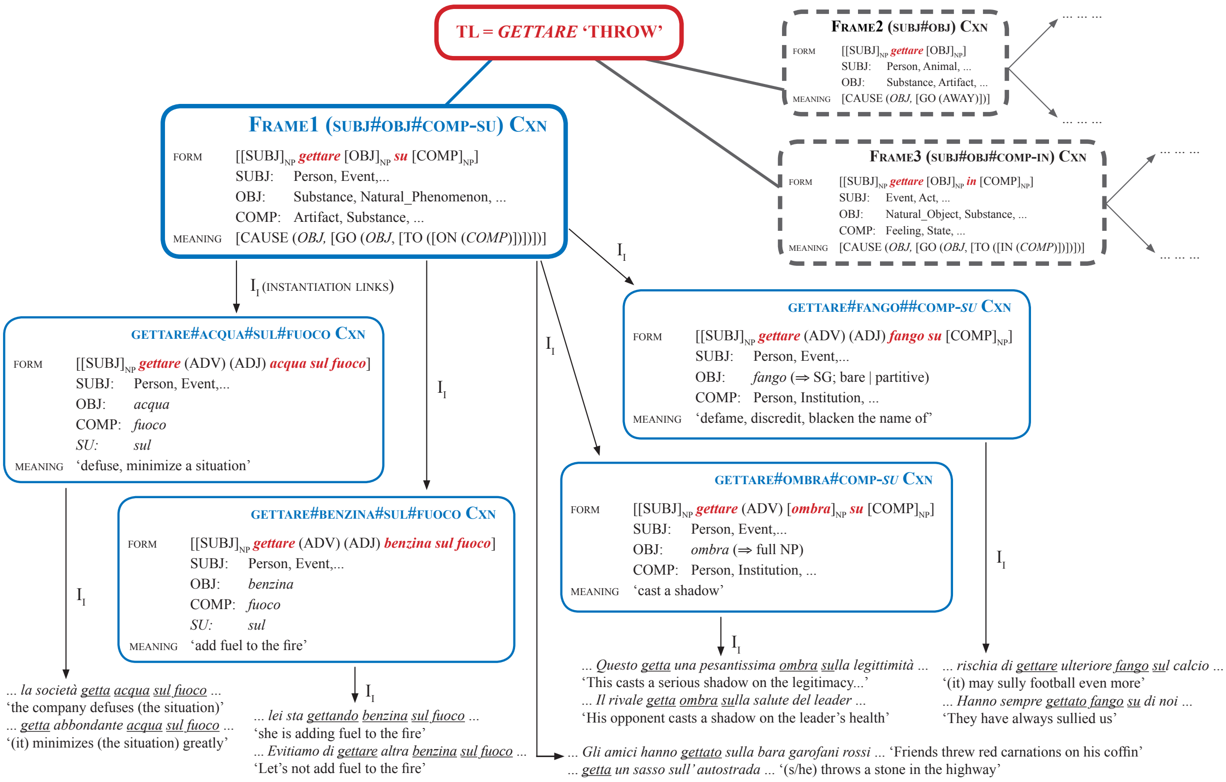
[Tabossi et al.2011] Patrizia Tabossi, Lisa Arduino, and Rachele Fanari. 2011. Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, 43:110–123.

[Wulff2008] Stefanie Wulff. 2008. *Rethinking Idiomaticity: A Usage-based Approach*. Continuum.

[Wulff2009] Stefanie Wulff. 2009. Converging evidence from corpus and experimental data to capture idiomaticity. *Corpus Linguistics and Linguistic Theory*, 5(1):131–159.

[Zeldes2013] Amir Zeldes. 2013. Productive argument selection: Is lexical semantics enough? *Corpus Linguistics and Linguistic Theory*, 9(2):263–291.

Appendix 1: A SYMPAThy-based view of the network of Cxns with the verb *gettare*



The verb *gettare* 'to throw' combines with the highly schematic subj#obj#comp-su Cxn, whose slots can freely vary with respect to linear order, presence of determiners, modifiers, etc. A semi-productive instance of this construction is the subj#obj:*ombra*#comp-su Cxn, with a fixed object slot and a partially variable oblique slot, which can appear with a semantically limited range of arguments. A fully lexically specified instance of the same construction is instead the subj#obj:*acqua*#comp-su:*sul-fuoco* Cxn, which has both slots instantiated and limited degree of variability.

Appendix 2: List of idioms used as experimental stimuli

Gettare la maschera ('*to reveal oneself*')
Gettare la spugna ('*to give up*')
Gettare acqua sul fuoco ('*to defuse a situation*')
Gettare olio sul fuoco ('*to inflame a situation*')
Mettere la mano sul fuoco ('*to stake one's life on sth*')
Mettere il carro davanti ai buoi ('*to put the cart before the horse*')
Mettere le carte in tavola ('*to lay one's cards on the table*')
Mettersi il cuore in pace ('*to resign oneself to sth*')
Mettere nero su bianco ('*to put sth down in black and white*')
Mettere il dito sulla piaga ('*to hit someone where it hurts*')
Mettere i puntini sulle i ('*to be nitpicking*')
Mettere zizzania ('*to sow discord*')
Perdere la testa ('*to lose one's head*')
Perdere il treno ('*to miss an opportunity*')
Perdere il filo ('*to lose the thread*')
Perdere la bussola ('*to lose one's bearings*')
Prendere il toro per le corna ('*to take the bull by the horns*')
Prendere una cotta ('*to get a crush on somebody*')
Prendere un granchio ('*to make a blunder*')
Tirare i remi in barca ('*to rest on one's oars*')
Tirare la cinghia ('*to tighten one's belt*')
Tirare le cuoia ('*to die*')
Tirare la corda ('*to take sth too far*')

On the use of antonyms and synonyms from a domain perspective

Debela Tesfaye
IT PhD Program
Addis Ababa University
Addis Ababa, Ethiopia
dabookoo@gmail.com

Carita Paradis
Centre for Languages and Literature
Lund University
Lund, Sweden
carita.paradis@englund.lu.se

Abstract

This corpus study addresses the question of the nature and the structure of antonymy and synonymy in language use, following automatic methods to identify their behavioral patterns in texts. We examine the conceptual closeness/distance of synonyms and antonyms through the lens of their DOMAIN instantiations.

1 Introduction

Using data from Wikipedia, this corpus study addresses the question of the nature and the structure of antonym and synonymy in language use. While quite a lot of empirical research using different observational techniques has been carried on antonymy (e.g. Roehm et al. 2007, Lobanova 2013, Paradis et al. 2009, Jones et al. 2012), not as much has been devoted to synonymy (e.g. Divjak 2010) and very little has been carried out on both of them using the same methodologies (Gries & Otani 2010). The goal of this study is to bring antonyms and synonyms together, using the same automatic methods to identify their behavioral patterns in texts. We examine the conceptual closeness/distance of synonyms and antonyms through the lens of their domain instantiations. For instance, *strong* used in the context of wind or taste (of tea) as compared to *light* and *weak* respectively, and *light* as compared to *heavy* when talking about rain or weight.

The basic assumption underlying this study is that the strength of co-occurrence of antonyms and synonyms is dependent on the domain in which they are instantiated and co-occur. In order to test the hypothesis we mine the co-occurrence information of the antonyms and the synonyms relative to the domains using a dependency grammar method. [1]

[1] http://nlp.stanford.edu/software/lexparser.shtml

The rationale is that the dependency parsing produces the relational information among the constituent words of a given sentence, which allows us to (i) extract co-occurrences specific to a given domain/context, and (ii) capture long distance co-occurrences between the word pairs. Consider (1).

1. Winters are cold and dry, summers are cool in the hills and quite hot in the plains.

In (1), the antonyms *cold*: *hot* modify winters and summers respectively. Those forms express the lexical concepts winter and summer in the domain temperature. The antonyms *cold*: *hot* co-occur but at a distance in the sentence. Thanks to the dependency information, it is possible to extract such long distance co-occurrences together with the concepts modified.

The article is organized as follows. In section 2, we describe the procedure and the two methods used: co-occurrence extraction of lexical items in the same sentence and a variant domain dependent co-occurrence extraction method. The latter method extracts patterns of co-occurrence information of the synonyms and antonyms in different sentences. In section 3 we present the results and discussions followed by a discussion of our results in comparison with related previous works in section 4. The conclusions are presented in section 5.

2 Procedure

Using an algorithm similar to the one proposed by Tesfaye & Zock (2012) and Zock & Tesfaye (2012), we extracted the co-occurrence information of the pairs in different domains separately, measuring the strength of their relation in the different domains with the aim of (i) making principled comparisons between antonyms and synonyms from a domain perspective, and (ii) determining the structure of antonymy and synonymy as categories in language and cognition.

Our algorithm is similar to the standard n-gram co-occurrences extraction algorithms, but

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

instead of using the linear ordering of the words in the text, it generates co-occurrences frequencies along paths in the dependency tree of the sentence as presented in the sections 2.2–2.5.

2.1 Training and testing data

The antonyms and synonyms employed for training and testing were extracted from the data used by Paradis et al. (2009) where the antonyms are presented according to their underlying dimensions and synonyms were provided for all the individual antonyms (for a description of the principles see Paradis et al. 2009). That set of antonyms and synonyms were used to extract their co-occurrence patterns from the Wikipedia texts in this study.

Dimensions	Antonyms	The associated synonyms of the antonyms
Size	Large	huge, vast, massive ,big ,bulky, giant ,gross, heavy, significant ,wide
	Small	little, low, minor, minute, petite, slim, tiny
Speed	Fast	quick, hurried, prompt, accelerating, rapid
	Slow	sudden, dull, gradual, lazy
Strength	Strong	forceful, hard, heavy, muscular, powerful, substantial, tough
	Weak	light, soft, thin, wimpy
Merit	Bad	crappy, defective, evil ,harmful, poor ,shitty ,spoiled ,unhappy
	Good	awful ,genuine ,great, honorable ,hot, neat, nice, reputable, right ,safe ,well

Table 1. The antonym pairs in their meaning dimensions and the associated synonyms.

2.2 Extracting the co-occurrences of the antonyms and synonyms in the respective domains

In order to extract the co-occurrences of the antonyms/synonyms in the respective domains we produced the relational information among the constituent words of a given sentence. To this end, we extracted the patterns linking the synonyms/antonyms and the concepts they modify and used this same pattern to extract more lexical concepts. The procedure was as follows.

- Start with the selected set of synonym/antonym pairs
- Extract sentences containing the pairs
- Identify the dependency information of the sentences
- Mine the dependency patterns linking the pairs with the concepts they modify
- Use these learned patters to extract further relations (synonym/antonym pairs and the associated concepts)

2.3 Extracting the domains

We created a matrix of antonym and synonym pairs matching every antonym and synonym from the list in Table 1. Using the patterns learned in section 2.2 we identified as many domains as possible for the pairs of synonyms and antonyms and calculated their frequency of co-occurrence in the respective domains.

When the lexical concepts were considered too specific, we referred them to more inclusive, superordinate domains. Frequency of occurrence was used as a criterion for conflation of concepts into superordinate ones as follows.

- Extract term co-occurrence frequencies within a window of sentences constituting both the antonyms/synonyms and the potential domain concepts. For instance:
  - Antonyms: *cold*: *hot*, domain concepts: winter, summer
  - Synonyms: *strong*: *heavy*, domain concepts: wind, rain
- Create a matrix of the potential domain concepts and the co-occurring terms with their frequencies
- Cluster them using the k-means algorithm
- Take the term with the maximal frequency (centroid) in each cluster and consider it the domain term
- Test the result using expert judgment running the algorithm on the test set.

Antonym/Synonym	Potential Domain concept	Words co-occurring with possible domain concepts	Frequency
hot cold	summer winter	temperature	50
		climate	43
		Wind	30

strong heavy	wind rain	wind rain	86
	winds snow-fall	winds snow-fall	3
	winds rainfall	winds rainfall	34
	waves rain-fall	waves rainfall	4

Table 2. The matrix of the frequencies of terms co-occurring with sample antonyms and the associated potential domain concepts

2.4 Extracting co-occurrences frequency specific to a given Domain/Context

The algorithm calculated the co-occurrence frequency of the antonyms/synonyms with the different concepts they refer to (or modify) as presented in table 3 by combining the information obtained in section 2.3 and section 2.4.

Antonyms	Concept 1	Concept 2	Frequency	Domain
hot cold	sum-mer	winter	10 5	temper-ature
strong heavy	wind	rain	11 2	winds rain
	winds	snowfall		
	winds	rainfall		
	waves	rainfall		

Table 3. The frequency of sample antonym specific to the underlying domains

2.5 Variant Domain Dependent Co-occurrence Extraction

In the previous algorithm, the co-occurrence information was extracted from the same sentence. However, unlike the antonyms, synonyms rarely occurred together in the same context (the same sentence and domain). It is natural to assume that in most cases synonyms are used in different contexts since they evoke similar but not identical meanings. This is however not the case for antonyms, which were always used to evoke properties of the same meanings when these antonymic words were used to express opposition (Paradis & Willners 2011), and in fact also when they are not used to express opposition (Paradis,et al., 2015). Because of this we decided to extract a variant domain dependent co-occurrence algorithm for the synonyms and antonyms, which instead extracts patterns of co-occurrence information of the synonyms and antonyms in different sentences, because we expected synonyms to be applicable to different, rather than the same contexts, since complete overlap of meanings of words are rare or even non-existent. This way we were able to gain information indirectly about their use by extracting their co-occurrence when they appear separately in different sentences while still being instantiated in the same domain. We mined the co-occurrence information of the synonym/antonym pairs separately in all possible domains and check if they co-occurred in the same sorts of domains:

- X(y, f)
- Z(y, f)

Where,

X and Z are a pair of a given antonym/synonym, Y is the domain within which the pairs of the antonym/synonym co-occur and f the frequency of the x-y or z-y co-occurrence.

The frequency of a pair of the antonyms/synonyms in the Y domain was counted and the same applies to the other pair. This made it possible to measure the degree of co-occurrence of the antonym/synonym pairs from the domain perspective indirectly.

3. Results and discussion

3.1 Co-occurrences in the same sentence

Based on the results of the experiment the strength of the antonyms/synonyms varies in relation to the domains of instantiation. Hence, the strength of the co-occurrence of antonyms and synonyms is a function of the domains. For instance, the antonyms: *slow: fast*, *slow: quick* and *slow: rapid* were used in completely different domains with little or no overlap. *Slow: fast* is used in the domains of motion, movement, speed; *slow: quick* is used for time, march, steps domains. The synonyms *powerful: strong* are used in the domains of voices, links, meaning; *strong: muscular* in the domains of legs, neck; *strong: heavy* are used in the domains of wind rain, waves rainfall, winds snow respectively; *intense: strong* in the domains of battle resistance, radiation gravity, updrafts clouds respectively.

We observed some unique patterns among the antonyms and synonyms as described below:

The antonyms:

- Co-occurred frequently in the same domain in the same sentence.

- The strength of the co-occurrence depends on the domain: *slow: fast* in the domains of growth, lines , motion, movement, speed ,trains, music, pitch*; slow: quick* in the domains of time, march, steps*; slow: gradual* in the domains of process, change, transition*; small: big* in the domains of screen, band*; small: large* in the domains of intestine, companies, businesses*; week: strong* in the domains of force, interaction, team, ties, points, sides, wind.

The Synonyms:

- Co-occurred in the same sentence but mainly in different domains. For instance, *fast: quick*, *strong: heavy*. Few co-occurrences in the same sentences in the same domains as exhibited by the pairs *gradual: slow* in the domains of process, change, development.
- The strength of the synonym co-occurrence depends on the domains. For instance, the synonyms *strong: heavy* in wind *and* rain domains respectively to express intensity; the synonyms *large: wide* in the domains of population and distribution domains respectively; *gradual: slow* in the domains of process, change, development*; small: low* in the domain of size cost, range, size weight, area, size price, amount density*; micro: small* in the domains of enterprises, businesses, entrepreneurs..

3.2 The variant domain dependent co-occurrence method

As mentioned before, the variant domain dependent co-occurrence extraction algorithm mines the patterns of co-occurrence information of the synonyms and antonyms in different sentences. The result from the variant co-occurrence experiment showed hardly any differences in the domains with which the synonyms and antonyms are associated. *Strong* in the domains of influence, force, wind, interactions, evidence, ties*; Heavy* in the domains of loss, rain, industry, traffic; *gradual: slow* in the domains of process, change, transition. However, we observed that the frequency of co-occurrence differed significantly. For instance, the frequency of the pair *gradual: slow* was 76 in same sentences experiment but 1436 in the variant co-occurrence experiment.

4 Comparison with related works

Previous research has shown that there are antonyms that are strongly opposing (canonical antonyms) (Paradis et al. 2009, Jones et al. 2012). Such antonyms are very frequent in terms of co-occurrence as compared to other antonyms: *small: large* as compared with *small: big*. In this experiment we found that the canonical antonyms are the set of antonyms the domains in which they function were numerous and productive. For instance the number of domains for *small*: *large* (11704) is by far greater than for *small*: *big* (120). However this doesn't make the antonym *small*: *large* more felicitous in all the domains. *Small: big* are the most felicitous antonyms for the domains such as screen, band as compared to *small: large*.

Measuring the strength of antonyms without taking domains into account provided higher values for the canonicals as they tended to be used in several domains. If domains were taken in to account, as we did in this experiment, all the antonyms were strong in their specific domains. The antonym pair *small*: l*arge* had higher value without considering domain in to account yet had 0.29 value in the domain of screen where *small*: *big* has much higher value (0.71). The values were calculated taking the frequency of co-occurrence of the domain term (*screen* in this case*)* with each antonyms and dividing it by the summation of the frequency of co-occurrence of the domain term (again *screen* in this case*)* with both antonyms (small big and small large).

5 Conclusion

The strength of the antonyms/synonyms varied in relation to the domains of instantiation. The use of antonyms and synonyms was very consistent with few overlaps across the domains. Similar results were observed in both experiments from the domain perspective although with significant differences in frequency. Antonyms frequently co-occurred in the same domains in the same sentences and synonyms co-occurred in different domains in the same sentences (with less frequency) and more frequently in different sentences in the same domains.

Acknowledgments

We acknowledge European Science Foundation (ESF) for providing us the funding to undertake this work.

References

Dagmar Divjak. 2010. Structuring the lexicon: a clustered model for near-synonymy. Berlin: de Gruyter.

Gries Stefan Th. & N. Otani. 2010. Behavioral profiles: a corpus-based perspective on synonymy and antonymy. *ICAME Journal, 34:121–150.*

Jones Steven, M.L. Murphy, Carita Paradis & Caroline Willners. 2012. Antonyms in English: Construals, constructions and canonicity. *Cambridge University Press*, Cambridge, UK.

Anna Lobanova. 2012. The Anatomy of Antonymy: A Corpus-Driven Approach. *Dissertation*, University of Groningen.

Carita Paradis. 2005. Ontologies and construals in lexical semantics. *Axiomathes,*15:541–573.

Carita Paradis, Caroline Willners & Jones Steven. 2009. Good and bad opposites: using textual and psycholinguistic techniques to measure antonym canonicity. *The Mental Lexicon, 4(3): 380–429.*

Carita Paradis, Simon Löhndorf , Joost van de Weijer & Caroline Willners. 2015. Semantic profiles of antonymic adjectives in discourse. *Linguistics, 53.1: 153 – 191.*

Roehm, D., I. Bornkessel-Schlesewsky, F. Rösler & M. Schlesewsky. 2007. To predict or not to predict: Influences of task and strategy on the processing o f semantic relations. *Journal of Cognitive Neuroscience, 19 (8):1259–1274.*

Debela Tesfaye. & Michael Zock. 2012. Automatic Extraction of Part-whole Relations. *In Proceedings of the 9th International Workshop on Natural Language Processing and Cognitive Science*.

Michael Zock. & Debela Tesfaye. 2012. Automatic index creation to support navigation in lexical graphs encoding part of relations. *Proceedings of the 3rd Workshop on Cognitive Aspects of the Lexicon (CogALex-III), COLING 2012.*

From physical to metaphorical motion: A cross-genre approach

Rosario Caballero
Universidad de Castilla-La Mancha (Spain)
MRrio.Caballero@uclm.es

Iraide Ibarretxe-Antuñano
Universidad de Zaragoza (Spain)
iraide@unizar.es

1 Introduction

Talmy (1991, 2000) classifies languages into verb-framed and satellite-framed types according to whether the Path of a motion event is lexicalized as a satellite of the main verb in the clause or as the verb itself. Thus, in English (and other S-languages like Dutch or Danish) verbs often encode rich information concerning Manner, Cause and/or Movement but need a so-called *satellite* to convey the Path of motion. In contrast, in Spanish and Romance languages in general, verbs are mainly concerned with trajectory or Path, and any other additional information (Manner or Cause of motion) is expressed by means of sentence constituents playing an adverbial role. As a result, speakers of verb-framed and satellite-framed languages appear to exhibit different rhetorical styles when describing the same motion event (Slobin, 1996, 2004).

Together with dealing with real motion, Talmy's work has provided the starting point in research on (a) fictive motion, i.e. the dynamic predication of physical yet static entities such as roads or cables, as in *The road climbs over the hill* (Langacker, 1986; Matsumoto, 1996a; Talmy, 2000; Matlock and Bergmann, in press), and (b) metaphorical motion, i.e. the dynamic predication of abstract entities such as the economy, emotions, and the like as in *Jealousy snaked its way into our relationship* (Özçalışkan, 2003, 2004, 2005, 2007; Morris et al., 2007). In general, research on fictive and metaphorical motion has focused on the way the speakers of different languages typically describe motion events in everyday, general contexts. Although yielding interesting results for the overall characterization of languages, this may result in a degree of overgeneralization towards the phenomenon at issue. This is reinforced by the way in which the data illustrating the research claims are often presented: the examples often appear in a decontextualized manner, with scarce or no mention to the characteristics of the discourse context or event where they are used, typically, the discourse genre where they occur). This is unfortunate since the inclusion and description of the context of use is critical for the study of motion patterns –whether real, fictive or metaphorical– and, above all, their correct interpretation.

2 Research questions

In this talk, we discuss the lexicalization patterns of metaphorical events in genre-specific texts in English, a satellite-framed language, and Spanish, a verb-framed language. More concretely, we explore whether (a) the lexicalization and rhetorical differences between Spanish and English discussed in the motion literature are sustained in genres other than narratives, and (b) the idiosyncrasy of those genres has any typological implications and, at the same time, affects the expressions' creativity and expressiveness.

3 Methodology

We use a 600.000-word corpus comprising tennis, wine and architecture reviews written in these two languages. These genres (or genre colony (Bhatia, 2000)) fall within reviewing practices: their main goal is to describe and evaluate an event (a tennis match) or an entity (wine and buildings) for an audience that may or may not have any previous knowledge about them, yet is interested in having an assessment written by a knowledgeable source. The texts were searched by hand in order to identify the motion constructions used in them. The unit of analysis was any instance concerned with motion –figurative or otherwise. A second step involved cleaning the texts and converting them into machine readable in order to run a concordance program and count the verb types and number of instances (tokens) in the three sub-corpora. After identifying the verbs used in the three genres, they were classified into two main groups in agreement with two criteria. First, the semantic information of verb involved (motion1-when the verb includes motion information in its semantic description and motion2-when the verb, despite not being a motion verb per se, can be reinterpreted as such due to the construction it is used in) and, second, the motion elements (Path-the trajectory or course followed by the moving object, Manner-the way

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

in which motion is performed) present in the examples. Figure 1 illustrates the coding.

WEC1925_SPECTATOR673

Like a gymnastMANNER, this lithe white glidesMOTION1 across the palatePATH, its underlying strength and tautness merely supporting the dried apricot, guava, honey and candied citrus. Finish just meltsMOTION2 awayPATH. Delicious. Drink now through 2004.

Figure 1: Example of corpus coding.

4 Results

As far as our first goal is concerned, our results show that the lexicalization and rhetorical patterns described for Spanish and English are maintained in the specific contexts explored, and therefore, results are congruent with research done on metaphorical motion events in general contexts. However, the data also yield interesting insights: metaphorical motion instances found in specific contexts are more expressive and abundant with regard to Manner than what is the case in general uses of language. This is particularly noteworthy in the Spanish data, whose expressivity contrasts with the general tendency to omit Manner and other details of motion events in other contexts. For instance, examples such as those in (1) are frequently used in our corpus:

(1a) architecture
La senda de exhibiciones de arte nurágico se ***desliza*** *entre ambas pieles del edificio permitiendo una visualización más íntima de las obras*
'The exhibition path of nuragic art slides between the two skins of the building allowing a more intimate visualization of the works'

(1b) wine
En boca tiene una magnifica entrada, suave, sabroso y equilibrado […], aunque en el paso sobresalen rasgos vegetales y ***se precipita*** *hacia un final en el que predominan notas tostadas y amargas*
'Smooth, tasty and balanced, it enters the mouth powerfully […] although some vegetal notes peek mid journey and it plunges towards a finish where toasty and bitter notes predominate'

(1c) tennis
Murray ***se pasea*** *en el ágora de Valencia*
'Murray strolls in the agora in Valencia'

This expressivity is more outstanding in the case of English: the data from the specific corpus not only reinforce the high expressivity and richness of this language with regard of Manner, but add novel verbs to those susceptible to being used in the description of motion events in other contexts (e.g. *hobble*, *sally forth*, *waltz*…), hence showing the creativity and –almost– endless possibilities of this language in this respect.

With respect to our second goal, we found that knowledge of the genre where the expressions are used is critical to correctly understand and explain metaphorical motion instances. This is particularly salient when comparing the use of the same verb in three different genres: indeed, a single verb may foreground aspects of a given situation irrelevant in a different context. For instance, the verb *tumble* in (2):

(2a) architecture
A stair tumbles down from this first floor incision onto the man-made island.

(2b) wine
The fruit shows well-ripened apples and peaches all the way into pineapples and mangoes, offering up a cascade of flavors that tumble across the palate.

(2c) tennis
Andy Murray has been sent tumbling out of AO 2008 by Frenchman Tsonga

The property of *tumble* shared by all these examples is 'uncontrolled', but this lack of control has a different interpretation in each genre. Thus, although in (2a) *tumble* suggests a certain lack of order, the main concern of the verb is to convey the visual force of the stair thus described, which somehow overwhelms those gazing at it. In (2b), the 'uncontrolled' property does not suggest a certain disorder or chaos of a wine's gustatory properties; rather, it expresses a sensory overflow or gustatory richness perceived by this critic as a positive trait of a complex wine. Finally, in (2c) the verb not only conveys Tsonga's convincing win, but Murray's pain and shame when losing to an inferior player ranking-wise.

Examples like these are interesting in three respects. First, although the information conveyed by motion verbs may be perfectly obvious for architects, tennis fans and wine aficionados and critics, this may not be the case for people outside these communities. Hence, the need to underline the importance of bringing the notion of *acculturation* to the centre of metaphor research, i.e. the relevance of taking into account all the factors that shape a given culture and its characteristic genres within a broader cultural panorama. Second, they problematize some of the views on both fictive and metaphorical motion discussed by cognitive scholars: (a) the trajectors and verbs involved depart from those typically described in fictive motion, and (b) the constructions dealing with buildings and wines do not comply with the unidirectional concrete-onto-abstract quality of the metaphorical mappings described in, for instance, the expression of financial issues or emotions, but involve concrete sources and targets. This suggests that fictiveness as opposed to metaphoricity may be a question of degree, yet this can only be ascertained by considering all the factors underlying the use of motion constructions in communication — from the trajectors involved to the reasons motivating their use. Third, while English and Spanish differ in the expression of real motion events, their differences are less dramatic in the expression of figurative motion which, again, points to the impact of culture and genre in the language use.

Acknowledgements

This research has been supported by ESF Short Visit Grants to both authors (NetWords 09-RNO-089, European Science Foundation) and by the Spanish Government (MovEsII, FFI2013-45553-C3-2-P; FFI2013-45553-C3-1-P).

References

Vijay K. Bhatia. 1999. Integrating products, processes, purposes and participants in professional writing. In C. Candlin and K. Hyland (eds.), *Writing: Texts, Processes and Practices*, 21-40. London: Longman.

Ronald W Langacker. 1986. Abstract motion. *Proceedings of the Twelfth Annual Meeting of the Berkeley Linguistics Society*, 12: 455-471.

Teenie Matlock and Till Bergmann. In press. Fictive Motion. In E. Dabrowska and D. Divjak (eds.), *Mouton Handbook of Cognitive Linguistics*. Berlin: Mouton de Gruyter.

Yo Matsumoto. 1996a. Subjective motion and English and Japanese verbs. *Cognitive Linguistics* 7, 183-226.

Yo Matsumoto. 1996b. How abstract is subjective motion? A comparison of access path expressions and coverage path expressions. In A. Goldberg (ed.), *Conceptual Structure, Discourse and Language*, 359-373. Stanford: CSLI Publications.

Michael W. Morris, Oliver J. Sheldon, Daniel R. Ames and Maia J. Young. 2007. Metaphors and the market: Consequences and preconditions of agent and object metaphors in stock market commentary. *Organizational Behavior and Human Decision Processes*, 102: 174-192.

Şeyda Özçalışkan. 2003. Metaphorical motion in crosslinguistic perspective: A comparison of English and Turkish. *Metaphor and Symbol*, 18 (3): 189-228.

Şeyda Özçalışkan. 2004. Encoding the manner, path, ground components of a metaphorical motion event. *Annual Review of Cognitive Linguistics*, 2: 73-102.

Şeyda Özçalışkan. 2005. Metaphor meets typology: Ways of moving metaphorically in English and Turkish. *Cognitive Linguistics*, 16 (1): 207-246.

Şeyda Özçalışkan. 2007. Metaphors we move by: Children's developing understanding of metaphorical motion in typologically distinct languages. *Metaphor and Symbol*, 22 (2): 147-168.

Dan I. Slobin. 1996. Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani and S. A. Thompson (eds.), *Grammatical Constructions: Their Form and Meaning*, 195-219. New York: Oxford University Press.

Dan I. Slobin. 2004. The many ways to search for a frog. In S. Strömqvist and L. Verhoeven (eds.), *Relating Events in Narrative. Typological and Contextual Perspectives*, 219-257. Hillsdale, NJ: Lawrence Erlbaum.

Leonard Talmy. 1991. Path to realization: A typology of event conflation. *Proceedings of the Seventeenth Annual Meeting of the Berkeley Linguistics Society*, 17: 480-519.

Leonard Talmy. 2000. *Toward a Cognitive Semantics*. Cambridge, MA: MIT Press.

'Taste' and its conceptual extensions: the example of Croatian root *kus/kuš* and Turkish root *tat*

Ida Raffaelli
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
iraffael@ffzg.hr

Barbara Kerovec
Faculty of Humanities and Social Sciences
Ivana Lučića 3, 10000 Zagreb, Croatia
bkerovec@ffzg.hr

This paper deals with the concept of 'taste' and its importance in the formation of Croatian and Turkish lexicon. 'Taste' as one of five basic sensory concepts serves as a source domain in conceptualizing various abstract domains, mostly related to human internal sensations (Sweetser, 1990). However, within the research of perception vocabulary, lexical structures related to the concept of 'taste' have been among the least investigated areas, especially according to different parts of speech and their correlation in building of vocabulary. A comparative analysis of the taste vocabulary in two typologically different and genetically unrelated languages like Croatian and Turkish could reveal the differences and similarities in processes that come into play in building their vocabulary. This is the reason why these two languages are chosen for the analysis. According to the embodiment hypothesis within Cognitive Linguistic theoretical framework, it can be expected that Croatian and Turkish share conceptual extensions towards the same abstract domains. However, since the two languages are typologically different and immersed in different cultures, some differences in conceptual mappings are also expected. Thus, one of the main goals of the present research is to provide a more fine grained analysis of semantic extensions of the taste vocabulary in the two languages. Besides examining similarities and differences in conceptual mappings, the aim of the paper is also to see to what extent the two languages differ with respect to lexicalization patterns that influence formation of the 'taste' vocabulary.

Croatian and Turkish taste vocabularies are described with respect to the morphosemantic structures of Croatian root *kus/kuš* "taste" and Turkish root *tat* "taste". The model of morphosemantic patterns (MP model) as developed by Raffaelli and Kerovec (2008) and Raffaelli (2013) regards the lexicon as morphologically and semantically related, i.e. each motivated lexeme is related to a root with respect to the word-formation processes and to the semantic (cognitive) processes. Moreover, the MP model regards the lexicon as a constructional continuum with no clear-cut boundaries between grammatical and lexical structures (cf. Langacker, 1987; Goldberg, 1995; Booij, 2010). It means that constructions such as *okušati se* "to try; to give it a go", *okušati se u* "to try out (a certain activity)" and *okušati se kao* "to try (out) as" are regarded as separate lexical units since they differ with respect to their usage, and exhibit differences in their meanings and their syntactic realizations. The MP model is a usage based model, thus conclusions about lexical structures and meanings are based upon a detailed analysis of lexical realizations in different contexts.

Meanings and contextual realizations of all analyzed lexical units in Croatian and in Turkish have been checked in the Croatian National Corpus, Croatian Web Corpus and METU Turkish Corpus.

As pointed out by Viberg (1984), concept of 'taste' is in general extended towards domains 'like'/'dislike'. Moreover, some cross-linguistic evidence (cf. Viberg, 1984; Evans and Wilkins 2000) shows a regular and frequent extension of taste verbs towards the meanings "to try", "to experience", "to enjoy". Although some cross-linguistic regularities of conceptual extensions of the concept 'taste' have already been established, the

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

comparative study of Croatian and Turkish taste vocabulary shows that there are some other abstract domains conceptualized by the domain of taste. Such domains are for example ‘ambience’, ‘mood’, ‘atmosphere’, ‘charm’, ‘enchantment’, that are all conceptualized by the domain of taste in Turkish, but not in Croatian, as showed in some examples below.

In Croatian the root *kus/kuš* is a basis of the verb *kušati* “to taste” that, by the process of prefixation, enabled formation of various verbs and constructions such as *pokušati* “to try; to attempt”, *iskušati/iskušavati/iskusiti* “to try; to experience”, *prokušati se* “to try; to try out”, *okušati se* (*u/kao*) “to try (out) (as)” that differ with respect to prefixes (and prepositions) and thus with respect to their usage and meanings. The perfective verb *okusiti* “to taste” differs from the verb *kušati* primarily in aspect, however all the others verbs cannot be used in relation with tasting food. They exclusively have abstract meanings like nouns *kušnja* and *iskušenje* “temptation”. Croatian is somehow specific with respect to the existence of two morphologically closely related nouns: *okus* “taste” and *ukus* “system of aesthetic judgement”, differing significantly according to their semantic structures. A distinction in usage and meanings of the two nouns will be analyzed and some specificities will be pointed out.

Morphosemantic field of the Turkish root *tat* exhibits some similarities and some differences in comparison to the morphosemantic field of the Croatian root *kuš/kus*. *Tat* “taste” is a noun used as a basis in the formation of the verb *tatmak* “to taste” and of the phrasal verbs *tadını görmek* “to taste” (lit. “to see the taste of”) and *tadına bakmak* “to taste” (lit. “to look at the taste of”). This means that, unlike in Croatian, verbs for visual perception are used for lexicalization of taste experience and taste activity. Similarly to Croatian, all three verbs relate to the domain of food as well as to the abstract domain of experience (e.g. *hayat tadını görmek* “to taste/experience life”, (lit. “to see the taste of life”). Turkish verbs do not extend their meanings to all abstract domains Croatian prefixed verbs do: they do not share meanings with Croatian verbs *pokušati* “to try; to attempt”, *okušati se, okušati se u, okušati se kao* “to try (out) (as)”, nor can they be related to the abstract domain of temptation (as with Croatian *iskušavati* ”to tempt; to test”, *iskušenje* ‘temptation’, *kušnja* ‘temptation; crucible”). Similarly, Turkish root *tat* cannot relate to the domain of aesthetic judgement (Croatian *ukus*), but when morphologically extended by suffixes *–li* “with” or *–siz* “without”, it extends to some domains Croatian root does not: *tatlı* (lit. „with taste“) does not mean “tasty”, but “sweet”. Accordingly, *tatlı* relates to a variety of pleasant experiences (feelings, climate, activities), while *tatsız* means “untasty”, but also “unpleasant”, “irritating”, “disturbing”, “annoying” etc. In addition, Croatian root *kus/kuš* cannot be used to express “enjoying” as Turkish root *tat* can (e.g. *tatilin tadını çıkarmak* “to enjoy holidays”, lit. “to extract the taste of holidays”). As far as contextual realizations are concerned, one of the most prominent differences between Croatian and Turkish is that Turkish root *tat*, besides verbs for visual perception, combines with verbs expressing motion (*Paris’in tadına varmak* “to experience the spirit/charm of Paris“, lit. “to come to the taste of Paris”), taking (*tadını almak* “to taste”, “to experience”, “to enjoy”, lit. “to take the taste of”; *tadını çıkarmak* “to enjoy”, lit. “to extract the taste of”), and cognitive activity (*tadını bilmek* “to experience”, lit. “to learn/to know the taste of”; *tadını tanımak* “to experience”, lit. “to get to know the taste of”), which is not the case in Croatian. Combining nouns and verbs derived from the same root is also characteristic for Turkish but not for Croatian (*tadını tatmak* “to taste the taste of”).

Thus, it could be claimed that Croatian verbs with extended abstract meanings are mostly realized in constructions such as [pref – $V_{kus/kuš}$ – prep] as *okušati se u* “to try out”, whereas Turkish verbs with extended meanings mostly appear in construction such as [N_{tat} – V] in which verbs within a construction often refer to concrete domains based in human experience, like for example motion.

The aim of this paper is: a) to provide an exhaustive description of the structure of the taste vocabulary related to the roots *kuš/kus* in Croatian and *tat* in Turkish, b) to point to some similarities and differences in the conceptual extensions of the concept 'taste' in the two languages and thus in the organization of their vocabularies, c) to implement the MP model in the description of lexical structures of non IE languages, and thus demonstrate its applicability in the lexical analysis of typologically different languages, pointing to regular and specific lexicalization patterns in the two languages.

References:

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

Nicholas Evans and David Wilkins. 2000. In the Mind's Ear: The Semantic Extension of Perception Verbs in Australian Languages. *Language* 76/(3): 546-592.

Adele E. Goldberg,. 1995. *Constructions. A construction Grammar Approach to Argument Structure.* Chicago and London: The University of Chicago Press.

Ronald W. Langacker, 1987. *Foundations of Cognitive Grammar*: Volume 1, *Theoretical Prerequisites*. Standford: Standford University Press.

Ida Raffaelli and Barbara Kerovec. 2008. Morphosemantic fields in the analysis of Croatian vocabulary. *Jezikoslovlje* (9.1-2): 141-169.

Ida Raffaelli. 2013. The model of morphosemantic patterns in the description of lexical architecture. *Linuge e linguaggio* 1: 47-72.

Eve Sweetser. 1990. *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantics.* Cambridge: Cambridge University Press.

Åke Viberg 1984. The verbs of perception: A typological study. In: Brian Butterworth, Bernard Comrie, and Osten Dahl (Eds.) *Explanations for language universals*. Berlin: Mouton de Gruyter. 123-62.

Love in the time of the corpora. Preferential conceptualizations of love in world Englishes

Javier E. Díaz-Vera
Universidad de Castilla-La Mancha
Departamento de Filología Moderna
13071 Ciudad Real, Spain
JavierEnrique.Diaz@uclm.es

1 Introduction

According to Gibbs (2006) "there is still insufficient attention paid to the exact ways that cultural beliefs shape both people's understandings of their embodied experiences and the conceptual metaphors which arise from these experiences." For example, the conceptual metaphor EMOTIONS ARE FLUIDS WITHIN THE BODY seems to underlie a wide variety of metaphorical expressions used by speakers from different linguistic and cultural areas all around the world. The geographical distribution of these metaphorical expressions is so general that numerous researchers have proclaimed their universal character, in so far as they are based on our common, embodied experience (Kövecses, 2000). However, the apparent ubiquity of this metaphorical mapping in contemporary emotional expressions does not necessarily imply that speakers from different linguistic or dialectal areas understand (or, of course, experience) emotions in the same identical way (Díaz-Vera and Caballero, 2013).

In this paper, I deal with the analysis of conceptual variation in the metaphorical construction of love in a group of dialectal varieties of contemporary English. Differently to earlier studies of love metaphors in English (Quinn 1987; Baxter, 1992; Kövecses, 1998), my main aim here is to analyze the socio-cultural dynamics of conceptual metaphor through the reconstruction of the preferential conceptualizations of love by speakers of a series of dialectal varieties of the same language, as spoken in culturally diverse regions. Through the analysis of the socio-cultural dynamics of conceptual metaphor, I intend to contribute to the field of Cognitive Dialectology by addressing the question whether cultural and conceptual differences can be detected language-internally, not just across languages.

Based on textual data extracted from the *Corpus of Global Web-Based English* (GloWbE; Davies, 2013), I will demonstrate here that the varieties of world English under scrutiny show significant differences in the conventional use of figurative expressions. Thereafter, these findings will be related to the cultural background of each speech community.

2 Research questions

Through the fine-grained analysis of the data described below, in this paper I will address the following research questions: (a) How do speakers from different parts of the English-speaking world conceptualize love? (b) What do these conceptual preferences tell us about these English varieties from a sociolinguistic perspective? (c) To what extent can social and cultural factors account for these processes of conceptual variation?

3 Methodology

As indicated above, the data used for this analysis has been collected using the GloWbE, which contains 1,9 billion words. This corpus is illustrative of the different ways English is used by speakers living in 20 different countries. The texts included in this corpus represent the genre 'personal blog' (Miller and Shepherd, 2009); these texts come from 1,8 million web-pages compiled in December 2012 using a highly automated production process.

The present study is limited to the analysis of data extracted from four different national sections within the GloWbE, illustrating two very different sociolinguistic contexts: the *inner circle* (i.e. countries where English is the primary language) and the *outer circle* (i.e. countries where English plays an important 'second language' role in a multilingual setting; Kachru, 1988). The four sub-corpora under scrutiny here are UK (inner circle), India, Pakistan and Nigeria (outer circle). In doing so, I will try to describe the different ways speakers from radically different cultural, social and religious regions conceptualize

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

love. I am especially interested in determining whether, and to what extent, these extra-linguistic factors can account for the conceptual differences illustrated in my quantitative analysis of love expressions.

In order to identify the metaphors for love used in the corpus, I have adopted the metaphorical pattern analysis (MPA) as proposed by Stefanowitsch (2004, 2006). This method, which takes the target domains of the figurative expressions as the starting-point of the analysis, consists in choosing one or more lexical ítems referring to the target domain under scrutiny and extracting a significative sample of their occurrences in the corpus. To start with, I have located all the instances of the noun *love* in the four corpus sections (GB, IN, PK and NG). As can be seen in Table 1, the absolute and relative distributions of this noun are highly irregular. For example, whereas only the GB section of the corpus scores a *per mil* frequency for this noun below the general GloWbE corpus average (217.98 ‰), the IN and the NG sections show much higher frequency rates.

SECTION	FREQ	PER MIL
GB	69392	179.02
IN	26355	273.30
PK	13114	255.30
NG	12179	285.58
GloWbE	410815	217.98

Table 1: Absolute and relative frequencies of the noun 'love' in four corpus sections.

In order to be able to compare the four corpus sections with each other, I have selected and analyzed only a random sample of 1,000 love expressions in each sub-corpus (4,000 expressions in all). After collecting 1,000 instances incorporating the key term *love* in each corpus section, I extracted the expressions where the emotion was discussed in metaphoric terms, and sorted them according to the general source domains motivating the figurative expression (e.g., NUTRIENT, JOURNEY, UNITY OF PARTS, FIRE, etc.). These were then further tagged paying attention to the more specific source and target domains involved in the metaphors (e.g., LOVE IS MADNESS within the more general metaphor LOVE IS INSANITY scenario). Thereafter, the resulting conceptual metaphors were further classified into three broad classes on the basis of their source-domain orientation (Kövecses, 2000: 110):

- *Space*-related source domains: The first category includes very general spatial metaphors, such as LOVE IS A BOUNDED REGION and LOVE IS A CONTAINER.
- *Force*-related source domains: The second category includes most of the source domains typically used in the conceptualization of emotions in English, such as EMOTION IS A NATURAL FORCE, EMOTIONS IS INSANITY or EMOTION IS FIRE.
- *Relationship*-related source domains: The third category includes a set of specific source domains for human relationships in English, such as HUMAN RELATIONSHIP IS A PLANT, HUMAN RELATIONSHIP IS A JOURNEY or HUMAN RELATIONSHIP IS ECONOMIC EXCHANGE.

Based on the above classification of specific source domains, I will assume here that speakers from different parts of the English-speaking world construe love via conceptual metaphor in different ways. Through the quantitative and qualitative analysis of the set of figurative love expressions collected in the GloWbE corpus, it is possible to determine the speakers' relative preferences to talk about love as a state, as an emotion or as a relationship. Through the comparative analysis of the figurative expressions used in the four corpus sections under scrutiny, I will try to illustrate how these conceptual preferences might be embedded in different cultural backgrounds. The results from each corpus section are discussed in turn in the following sections.

4 Findings and discussion

As indicated above, the data used for this analysis has been collected using the GloWbE. The texts included in this corpus illustrate the genre 'personal blog'; furthermore, as indicated above, these texts where compiled during a relatively short period of time (December 2012). Consequently, they are highly homogeneous not only in terms of their genre, but also in terms of their date of production.

As described above, in the first stage of this research I have located all the instances of the noun *love* in four corpus sections (GB, IN, PK and NG). Thereafter, I have classified these expressions into two large groups: literal and figurative expressions. According to this part of my analysis (see Table 2), the four corpus sections analyzed here show relatively similar rates of literal and non-literal love expressions. Whereas the highest amount of figurative expressions is found in the GB section (43.6%), the lowest number of metaphors corresponds to the PK section (34.7%).

SECTION	LITERAL	FIGURATIVE
GB	564	436
IN	568	432
PK	653	347
NG	596	404
TOTAL	2,381	1,619

Table 2: Distribution of literal and figurative 'love' expressions in four corpus sections.

However, as can be seen in Table 3, major differences arise if we compare the relative frequencies of the three broad categories of source domain described above (i.e. space, force and relationship). In spite of the very similar total number of instances of each category, the geographical distribution of these occurrences clearly points towards a preference for *force*-related source domains in the PK (42.0%) and in the IN (37.0%) sections, in clear contrast with the neat preference for *space*-related source domains in GB and NG (41.0%).

SECTION	SPACE	FORCE	RELATION
GB	177	129	130
IN	140	159	133
PK	108	146	93
NG	164	142	98
TOTAL	589	576	454

Table 3: Distribution of *space*-, *force*- and *relationship*-related source domains in four corpus sections.

Furthermore, according to the data presented above, whereas *relationship*-related source domains occupy a secondary position in the four corpus sections, their relative frequency is especially low in the PK (27.0%) and in the NG (24.0%) sections.

4.1 *Space*-related metaphorical patterns

Space-related metaphorical patterns represent the most general and neutral option as regards the expression of states and emotions. According to these EVENT STRUCTURE metaphors, states in general are conceptualized as physical locations or bounded regions in space. Speakers use sentences such as 'I am in love' to indicate, in a very neutral way, their emotional state. The adverb *deeply* is frequently used in these examples in order to indicate intensity of the emotion. The notion of change is viewed as motion into (as in 'I am falling in love') or out of (as in 'I am falling out of love') this emotional state, conceptualized as a container. Within this group, I have found several expressions where love is conceptualized as a nest, and lovers are birds in the nest.

According to the GB data, there is a strong preference among British speakers to use the noun *love* in expressions conveying the metaphors LOVE IS A BOUNDED REGION (83 instances) and LOVE IS A CONTAINER (94 occurrences). The relative frequency of these metaphors is much lower in the other three corpus sections. As can be seen in Table 4, only in the NG section we find a similar relative frequency of the metaphor LOVE IS A CONTAINER.

SECTION	REGION	CONTAINER	TOTAL
GB	83	94	177
IN	65	75	140
PK	42	66	108
NG	68	96	164
TOTAL	258	331	589

Table 4: Distribution of *space*-related source domains in four corpus sections.

4.2 *Force*-related metaphorical patterns

Force-related metaphors are frequently used by English speakers in order to express their emotions. According to this view, love can be conceptualized as a NATURAL/PHYSICAL FORCE, as an OPPONENT IN A STRUGGLE, or as FIRE/LIGHT, among others. Broadly speaking, these conceptual mappings indicate that the person in love is passively affected by a force (either external or, less frequently, internal), which produces either resistance or loss of control (or both). Preference for these metaphorical expressions points towards a stronger presence of the passionate ideal of love that characterizes the earliest stages of the relationship (Luhmann 1996; Schröder 2009: 105).

Within this group, I have analyzed the distribution of 17 love metaphors in the four corpus sections. The results of this part of the analysis can be seen in Table 5.

SOURCE	GB	IN	PK	NG
FLUID/CONTAINER	32	8	45	11
INSANITY	25	18	21	8
NATURAL FORCE	15	15	14	16
OPPONENT	14	12	4	3
WAR	14	10	4	4
FIRE/LIGHT	10	12	20	12
NUTRIENT	7	18	8	15
HIGH/RAPTURE	7	9	4	6
HEALING	3	2	3	1
SPORT/GAME	2	6	2	8
BOND	-	8	5	5
DEITY	-	19	3	22
ART/SKILL	-	5	3	6
CAPTIVE ANIMAL	-	8	9	8
WARMTH	-	-	-	4
MAGIC	-	-	-	5
AIR	-	-	-	1
TOTAL	129	159	146	142

Table 5: Distribution of *force*-related source domains in four corpus sections.

According to the data described in Table 3 and in Table 5, the GB section yields the lowest number of instances in which love is portrayed as a force (129 instances in all). The largest number of examples in this corpus section portray love either as a SUBSTANCE INSIDE THE EXPERIENCER (32 instances) or as INSANITY (25 instances) and, hence, are compatible with views of other emotions (such as anger or happiness; Kövecses 2000). The other three sections yield not only a higher frequency rate of *force*-related metaphors (IN: 259; PK: 146; NG: 142), but also a more varied articulation in terms of source domains within this category. In fact, many of the expressions analysed here instantiate the metaphors LOVE IS A DEITY, LOVE IS WARMTH and LOVE IS MAGIC, all of which are completely absent from the part of the GB section analysed here.

4.3 *Relationship*-related metaphorical patterns

This category includes those metaphorical expressions where love is portrayed by speakers as a romantic relationship between two individuals, who cooperate with each other in order to reach a common goal. These metaphors are frequently found in reference to other types of human relationship (such as friendship), and are normally related either to the handling of complex physical objects (such as plants, buildings or machines) or to interactive cooperation (as in, for example, economic exchange, hidden object or journey). The overall distribution of the 7 conceptual mappings included within this category in each corpus section (Table 3 above) indicates that *relationship*-related source domains motivate a relative low number of metaphorical expressions in the four sections. This is especially true in the case of the PK (27.7%) and the NG (24.0%) sections, both of which yield a high number of examples of *force*-related mappings.

SOURCE	GB	IN	PK	NG
VALUABLE OBJECT	43	36	19	11
LIVING ORGANISM	25	9	9	8
HIDDEN OBJECT	24	26	13	27
ECON. EXCHANGE	20	36	34	32
UNION OF PARTS	9	8	4	5
JOURNEY	6	11	9	9
BUILDING	3	7	5	6
TOTAL	130	133	93	98

Table 6: Distribution of *relationship*-related source domains in four corpus sections.

5 Conclusion

The findings of my research of love expressions in a variety of world Englishes shows that there exist important differences in the conceptualization of love, from the more passional *force*-related expressions to the more rational *relationship*-related ones. Based on this distinction, I have analyzed the distribution of each set of metaphors in four GloWbE sections. Whereas overseas Englishes show a preference for *force*-based mappings, GB English is relatively neutral (as in the general LOVE IS A STATE metaphor). Further, whereas the idea of romantic love (emphasis on the collaborative relationship between two partners, typically Western love ideal; Novak 2013) is more frequent in the GB section, the other corpus sections show a greater tendency to talk about love as an emotion, accentuating the moment rather than the future.

Acknowledgments

This research has been supported by the Regional Government of Castilla-La Mancha (“The Expression of basic Emotions in English: Diachronic and Sociolinguistic Variation” PPII-2014-015-A).

References

Leslie A. Baxter. 1992. Root metaphors in accounts of developing romantic relationships. *Journal of Social and Personal Relationships*, 9: 253-275.

Mark Davies. 2013. *Corpus of Global Web-Based English: 1.9 billion Words from Speakers in 20 Countries*. Available online at http://corpus.byu.edu/glowbe/.

Javier E. Díaz-Vera and Rosario Caballero. 2013. Exploring the feeling-emotions continuum across cultures: Jealousy in English and in Spanish. *Intercultural Pragmatics*, 10(2): 265-294.

Raymond W. Gibbs. 2006. Cognitive linguistics and metaphor research: Past successes, skeptical questions, future challenges. *DELTA* 22: 1–20.

Braj B. Kachru. 1988. The spread of English and sacred linguistic cows. In Peter H. Lowenberg (ed.), *Language Spread and Language Policy: Issues, Implications and Case Studies*, 207-228. Washington, D. C.: Georgetown University Press.

Zoltan Köveces. 1998. *The Language of Love: The Semantics of Passion in Conversational English.* Toronto: Bucknell University Press.

Zoltan Köveces. 2000. *Metaphor and Emotion. Language, Culture, and Body in Human Feeling.* Cambridge: Cambridge University Press.

Niklas Luhmann, 1996. *Liebe als Passion. Zur Codierung vor Intimität.* Frankfurt: Suhrkamp.

Carolyn Miller and Dawn Shepherd. 2009. Questions for genre theory from the blogosphere. In Janet Giltrow and Dieter Stein (eds.), *Genres in the Internet. Issues in the Theory of Genre*. 263-290. Amsterdam: John Benjamins.

Michael Novak. 2013. The Myth of Romantic Love and other Essays. Piscataway: Transaction.

Naomi Quinn. 1987. Convergent evidence for a cultural model of American marriage. In Dorothy Holland and Naomi Quinn (eds.), *Cultural Models in Language and Thought*. 173-192. Stanford: Stanford University Press.

Ulrike A. Schröder. 2009. Preferencial metaphorical conceptualizations in everyday discourse about love in the Brazilian and German speech communities. *Metaphor and Symbol* 24: 105-120.

Anatol Stefanowitsch. 2004. HAPPINESS in English and German: A metaphorical-pattern analysis. In Kemmer Achard and Susanne Kemmer (eds.), *Language, Culture, and Mind*, 137–149. Stanford: CSLI.

Anatol Stefanowitsch. 2006. Words and their metaphors: A corpus-based approach. In Anatol Stefanowitsch and Stephan Th. Gries (eds.), *Corpus-based Approaches to Metaphor and Metonymy*, 63-105. Berlin and New York: Mouton de Gruyter.

Is black always the opposite of white? The comprehension of antonyms in schizophrenia and in healthy participants

Cristina Cacciari
Univ. of Modena, Italy
cacciari.cristina@unimore.it

Francesca Pesciarelli
Univ. of Modena, Italy
francesca.pesciarelli@unimore.it

Tania Gamberoni
AUSL, Modena, Italy
t.gamberoni@ausl.mo.it

Fabio Ferlazzo
Sapienza Univ. Rome, Italy
fabio.ferlazzo@uniroma1.it

Abstract

In this study, we tested the online comprehension of antonyms in 39 Italian patients with paranoid schizophrenia and in an equal number of pairwise-matched healthy controls. Patients were rather accurate in identifying antonyms, but compared to controls, they showed longer response times and higher priming scores, suggesting an exaggerated contextual facilitation. Presumably, this reflects a deficient controlled semantic processing and an overreliance on stored semantic representations.

1 Introduction

In this study we investigated the recognition of antonym word pairs in patients with paranoid schizophrenia and in pairwise matched healthy participants.

Conceptual knowledge stored in semantic memory includes representations of many different types of lexico-semantic relationship, among which antonymy. Antonymy is thought to be the most robust of the lexico-semantic relations, relevant to both the mental organization of the lexicon and the organization of coherent discourse (Fellbaum, 1998; Willners, 2001; Jones, 2002; Murphy, 2003; Paradis and Wilners, 2006; van de Weijr et al., 2014). Antonymy is the label generically used to refer to any of two words that are semantically opposed and incompatible for at least one of their senses (e.g., *black/white*, *dead/alive*). Antonyms are recognized faster than any other words or non-words in word recognition, elicit each other in word association tests and are often mistaken in speech errors. Antonyms occur very frequently in written and oral language, presumably because binary contrast is a powerful organizing principle in perception and cognition (Bianchi et al., 2011). In sum, antonym word pairs represent an important phenomenon for elucidating the nature of the semantic dysfunction that characterizes schizophrenia (henceforth, SZ) and, on more general grounds, for establishing the neural and cognitive prerequisites of word comprehension. Studying the types of semantic relationship that patients with SZ can or cannot correctly understand may also yield further insights into the ways in which semantic knowledge is represented in the human brain, and into the mechanisms underlying its use.

SZ is a neurobiological disorder associated to several cognitive deficits that include mild to severe language comprehension and production abnormalities (at word and sentence levels) as well as attentional and information processing impairments (Harvey, 2010; Kuperberg, 2010ab, Kiang, 2010; Levy et al., 2010). The literature has shown that language comprehension impairment in SZ are not global and generalized but selectively involve abnormalities at a word and/or sentence level (Kuperberg, 2010ab). Studies on word processing in SZ have predominantly used the semantic priming paradigm obtaining mixed results (for overviews, Minzenberg et al., 2002; Pomarol-Clotet et al., 2008; Pesciarelli et al., 2014). Typically, studies have obtained greater than normal semantic priming (*hyper-priming*) at short intervals between the presentations of prime and target (SOA, stimulus onset asynchrony) especially, but not only, in thought-disordered patients. Hyper-priming is often accompanied by reduced or absent priming at long SOAs (more than 300 msec). These distorted priming effects have been interpreted in terms of abnormal neural processing of the relationships between concepts in long-term semantic memory and of functional abnormalities of semantic memory neural networks that produce abnormally fast and/or far-reaching spreading of activation among concepts (Kiang, 2010). Patients with SZ would also fail in suppressing or deactivating contextually inappropriate semantic associations because of the distorted use of context that characterize SZ. This deficit has been attributed to a more general deficit in constructing and maintaining an internal representation of context for control of action (Cohen et al., 1999), due to working memory deficit (Barch and Ceaser,

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

2012). But, according to some authors, patients would fail in inhibiting contextually-irrelevant information, especially at long SOAs (Minzenberg et al., 2002), rather than in encoding contextually-relevant information. This impairment would be linked to a more global deficit in controlled semantic processing (Titone et al., 2000; Titone et al., 2002).

The importance of antonyms for elucidating the organization and retrieval of semantic knowledge is documented by the recent resurgence of interest on antonyms in normal comprehension (e.g., de Weijers et al., 2014; Paradis et al., 2009). In contrast, the vast literature on semantic processing deficit in SZ has almost ignored antonyms with the exception of a few paper-and-pencil studies of the 1960s (Blumberg and Giller, 1965; Burstein, 1961) that have documented impairment of SZ patients on antonyms. This underestimation of antonyms as a relevant test case of semantic organization can also be attributed to the fact that most neuropsychological studies on conceptual representations have primarily investigated semantic similarity rather than opposition (Crutch et al., 2012) at variance with the fact that semantic opposition, rather than similarity, is thought to be the axis around which the adjectival lexicon clusters (Murphy, 2003; Paradis and Willners, 2011).

2 Aims of the study

Shedding light on whether or not antonym identification is spared in a neurobiological disorder typically associated to semantic deficit may improve our understanding of the organization of word storage in the human brain (Jeon et al., 2009). Our general aim was therefore to expand the knowledge about the cognitive processes underlying the recognition of antonyms, and to evaluate whether these processes differed in SZ and in normal language comprehension. We tested whether the semantic dysfunction that often characterizes people with SZ necessarily leads to a loss of the capacity to recognize antonyms when antonyms are presented alone, rather than with homonyms and/or synonyms (Blumberg and Giller, 1965; Burstein, 1961), and when they are tested with a real-time task (for a more detailed version of this study, see Cacciari et al., 2015).

SZ patients tend to be less accurate and slower than healthy controls on most cognitive measures (Harvey, 2010; Vinogradov et al., 1998). Since response slowing is related to the disease, rather than necessarily reflecting semantic dysfunction (Niznikiewicz et al., 2010), this may lead to an artificial increase of the reaction time difference with healthy participants. To avoid this confound, often semantic priming studies have used a priming score (PRI; Spitzer et al., 1993), rather than the mere response times to the targets. The PRI reflects the amount of facilitation of prior context on the response time to a target and is calculated as follows: $(RT_{\text{unrelated targets}} - RT_{\text{related targets}})/ RT_{\text{unrelated targets}})*100$ (Spitzer et al., 1993). Here, we compared the individual PRI of patients to those of pairwise matched healthy controls.

Subjects read a definitional sentence fragment (*The opposite of word is..*) that, upon pressing the space bar, was followed by the antonym or an unrelated control word. This self-paced target verification task is suited to obtain information on real-time comprehension while placing little demand on the need to maintain and update information in working memory. We did not use similar, fixed time durations for patients and controls because SZ patients typically need longer presentation durations than healthy subjects to perceive a stimulus.

Healthy subjects should respond in a fast and accurate way, in line with the literature. Semantic priming studies often observed an exaggerated priming score of patients compared to controls (for an overview, see Kuperberg, 2010b; Pomerol-Clotet et al., 2008). This, as we mentioned, has been mostly attributed to faster than normal and far-reaching spread of activation in semantic memory. This larger semantic priming effect has been observed under the ‘automatic’ condition of word priming at short SOAs (Minzenberg et al., 2002). In this study, the priming effect elicited by the definitional sentence fragment on the target word, if any, would occur under strategically controlled conditions since the target presentation is self-paced, and the definitional sentence fragment strategically guides the semantic search toward the item that fulfills the antonymy definition. Notwithstanding, if indeed patients are characterized by an abnormal spread of activation, we should obtain larger priming scores in patients than in controls. This result would contribute to clarify the conditions under which hyper-priming effect can occur. The easy nature of the task, the high written frequency and bound lexical couplings of the antonym pairs of this study can minimize semantic processing demands. However, it is unlikely that an even *intact* ability to identify antonyms may eliminate any group difference, given the general cognitive deficits of people with SZ. To limit this potential confound, we carried out analyses of covariance on mean response times and accuracy to partial out the contribution of covariates (i.e., Verbal fluencies, Vocabulary, and Digit Span). Although we did not necessarily expect accuracy to be compromised in patients, given their mild-to-moderate form of SZ, the low demanding nature of the task and the high familiarity of the stimuli, we expect accuracy to be modulated by the severity of thought disorder and the clinical state of patients, as found in prior studies on semantic processing in SZ.

3 Method

3.1 Participants

Participants included 39 Italian chronic outpatients with paranoid SZ (14 female; mean age 31 years, age range 20-45, SD 6.2) and 39 healthy volunteers as control participants (see Table 1 for a characterization of patients and controls). The diagnosis of paranoid SZ is based on the *Positive and Negative Syndrome Scale* (PANSS; mean score: 46.69, range: 34-68) and was confirmed by the clinical consensus of staff psychiatrists. Participants gave their informed consent for inclusion before they participated in the study (approved by the Ethics Committee of Modena).

Table 1
Demographic characteristics of the study sample, and clinical characteristics of the schizophrenic patients

	Patients		Controls		
	Mean	SD	Mean	SD	p
Sex	M=25; F=14		M=25; F=14		
Age (years)	31.41	6.22	31.28	6.31	.93
Education (years)	12.56	1.33	12.51	1.48	.88
Drug	SG=33; FG=2; FSG=4				
Years of illness	8.97	5.94			
WAIS-R (Verbal Scale)	91.05	15.41			
WAIS-R (Performance Scale)	86.31	19.42			
WAIS-R (Total Score)	87.82	18.31			
Vocabulary (WAIS-R)	8.23	3.24	10.77	2.38	.0001
Phonemic Fluency	28.51	8.25	37.28	7.68	.0001
Semantic Fluency	38.44	8.44	44.10	7.74	.003
BADA (errors)	1.15	1.18	0.03	.16	.0001
Digit SPAN (Forward)	5.44	.74	5.85	.83	.04
Digit SPAN (Backward)	3.75	1.07	4.28	.97	.05
Digit SPAN (Total Score)	9.18	1.51	10.13	1.57	.02
BPRS			2	0	
PANSS (Positive Scale)	11.64	3.12			
PANSS (Negative Scale)	11.21	4.02			
PANSS (Gen Psyc Scale)	23.84	3.43			
PANSS (Total Score)	46.69	8.13			

M = male; F = female; FG = first-generation antipsychotics; SG = second-generation antipsychotics; FSG = combination of first- and second–generation antipsychotics.

3.2 Materials and Procedure

Participants were presented with a definitional sentence fragment containing the first word of the antonym pair (e.g., *The opposite of black is*) followed by the correct antonym (*WHITE*) or by a semantically unrelated word (*NICE*). Subjects had to decide whether or not the target was correct. We used 40 very familiar antonym word pairs (W1-W2; e.g., *black/white*, *dead*/*alive*; *long/short; optimistic/pessimistic*) in which the antonym had a cloze probability value of 0.98.

Each W1 was also paired with a semantically unrelated non-antonym target word (W3). Two lists were created each containing 40 sentences with the same format. The target word was an antonym in 20 sentences and a semantically unrelated, non-antonym word in the other 20 sentences. A spacebar press initiated the presentation of the definitional sentence fragments as *The opposite of word is*; a second spacebar press initiated the presentation of the target word that remained on the screen until response. Participants pressed a *YES* button to respond to correct targets and a *NO* button for incorrect targets.

4 Results

Significant group differences emerged in all the neuropsychological tests (see Table 1) administered to patients and controls. The priming scores revealed a statistically significant, enhanced contextual priming in patients compared to controls (16.04% vs. 9.6%). The ANCOVA on response times showed significant main effects of Group, with patients overall slower than controls (Ant.: 1273 ms; Unrel.: 1645 ms; Ant.: 984 ms; Unrel.: 1108 ms, for patients and controls respectively), and of Vocabulary. The ANCOVA on accuracy (Ant.: 96%; Unrel.: 98%; Ant.: 98%; Unrel.: 99%; for patients and controls respectively) showed a main effect of Vocabulary. In addition, the accuracy and response times of patients significantly correlated with Vocabulary scores (WAIS-R) in that patients scoring higher in the Vocabulary test also were overall faster in responding to antonyms and non-antonyms and more accurate in rejecting non-antonyms. Patients scoring higher on the Verbal Scale (WAIS-R) also had faster response times to antonyms, and patients scoring higher on the Positive Scale (PANSS) a lower accuracy on antonyms.

5 Conclusions

While antonym recognition was fast and accurate in heathy controls, the picture emerging for patients is more complex. Specifically, the preceding definitional fragment facilitated antonym recognition in both patients and healthy controls but the amount of facilitation indeed differed. In fact patients were helped more than controls by the previous definitional context, as shown by the larger reduction of response times to antonyms than to non-antonyms (on average, patients were 25.4% faster in responding to antonyms than to non-antonyms compared to 11.8% of controls), and by the exaggerated priming effect of patients (close to twice the effect of controls). This enhanced semantic priming was not associated to the clinical state and/or the thought disorder of patients. In sum, the patients group encoded contextually relevant target words (Titone et al., 2000; Titone et al., 2002) but to a much higher degree than controls. Interestingly, this larger semantic effect occurred under strategically controlled conditions rather than under the automatic condition typical of word priming at short SOAs (Minzenberg et al., 2002). This suggests a compromised ability of patients with SZ to engage in the controlled processing operations necessary to flexibly use semantic memory representations. At the same time the relatively high level of accuracy of patients (96.6% vs. 98.5% of healthy subjects) suggests a preserved semantic storage and access to semantic representations (Titone et al., 2002; Titone et al., 2007). High accuracy may reflect a ceiling effect as well as the fact that polarity information processing can be less demanding on executive resources than other types of semantic relationships (Crutch et al., 2012). Consistently with the reported effects of thought disorder on semantic processing (for overviews, see Kuperberg, 2010b; Pomarol-Clotet et al., 2008), patients with higher scores of positive thought disorder were also less accurate in identifying antonyms. Accuracy instead improved in patients scoring higher in both the Vocabulary sub-test and the Verbal scale of WAIS-R (these patients also had faster response times). These results are consistent with prior studies indicating that in SZ high Vocabulary scores are protective of semantic deterioration (Brébion et al., 2010) reflecting premorbid intelligence (Lezak et al., 2004). On more general grounds, these results provide further evidence of the already documented association of verbal intelligence to efficient language comprehension (Hunt, 1977). Overall, our results indicate that the state of residual SZ contributed to slower antonym recognition above and beyond the cognitive deficits that characterize SZ patients. In sum, it is not the case that patients comprehended antonyms as controls, but simply at a slower pace. In fact, compared to controls, patients not only had longer response times but also enhanced priming scores that presumably reflect deficient controlled semantic processing and overreliance on stored semantic representations. In conclusion, all other things being equal, antonym identification requires a preserved ability to appreciate the difference between *maximally similar* and *maxi-*

mally dissimilar meanings (Paradis and Willners, 2011). This ability to a large extent relies on preserved executive resources, integrity of the semantic processing system and size of the lexicon.

References

Deanna, M. Barch and Alan Ceaser. 2012. Cognition in schizophrenia: core psychological and neural mechanisms. *Trends in Cognitive Sciences*, 1: 27-34.

Ivana Bianchi, Ugo Savardi and Michael Kubovy. 2011. Dimensions and their poles: a metric and topological theory of opposites. *Language and Cognitive Processes*, 26: 1232-1265.

Stephen Blumberg and Donald W. Giller. 1965. Some verbal aspects of primary-process thought: a partial replication. *Journal of Personality and Social Psychology*, 95: 517-520.

Gildas Brébion, Rodrigo A. Bressan, Ruth I. Ohlsen, Lyn S. Pilowsky and Anthony, S. David. 2010. Production of atypical category exemplars in patients with schizophrenia. *Journal of the International Neuropsychological Society*, (16), 5: 822-828.

Alvin G. Burstein. 1961. Some verbal aspects of primary process thought in schizophrenia. *The Journal of Abnormal and Social Psychology,* 62: 155-157.

Cristina Cacciari, Francesca Pesciarelli, Tania Gamberoni, Fabio Ferlazzo, Leo Lo Russo, Francesca Pedrazzi, and Ermanno Melati. 2015. Is black always the opposite of white? An investigation on the comprehension of antonyms in people with schizophrenia and in healthy participants. *Behavioral Sciences,* 5:93-112.

Jonathan, D. Cohen, Deanna, M. Barch, Cameron Carter and David Servan-Schreiber. 1999. Context-processing deficits in schizophrenia: converging evidence from three theoretically motivated cognitive tasks. *Journal of Abnormal Psychology*, 108: 120-133.

Sebastian J. Crutch, Paul Williams, Gerard R. Ridgway and Laura Borgenicht. 2012. The role of polarity in antonym and synonym conceptual knowledge: evidence from stroke aphasia and multidimensional ratings of abstract words. *Neuropsychologia*, 50: 2636-2644.

Christiane Fellbaum. 1998. (ed.) *WordNet: an electronic lexical database*. Cambridge, MA: MIT Press.

Philip D. Harvey. 2010. Cognitive functioning and disability in schizophrenia. *Current Directions in Psychological Science*, 19: 249-254.

Earl Hunt. 1977. We know who knows, but why? In R. C. Anderson, R. J. Spiro & W. E. Montague (eds) *Schooling and the acquisition of knowledge*, 327-333. Hillsdale NJ: Laurence Erlbaum.

Hyeon-AE Jeon, Kyoung-Min Lee, Young-Bo Kim and Zang-Hee Cho. 2009. Neural substrates of semantic relationships: common and distinct left-frontal activities for generation of synonyms vs. antonyms. *NeuroImage*, *48*: 449-457.

Steven Jones. 2002. *Antonymy: A corpus-based perspective*. London: Routledge.

Michael Kiang. 2010. Schizotypy and language: a review. *Journal of Neurolinguistics*, 23: 193-203.

Gina R. Kuperberg. 2010a. Language in Schizophrenia. Part 1. An Introduction. *Language and Linguistic Compass* (4), 8: 576-589.

Gina R. Kuperberg. 2010b. Language in Schizophrenia. Part 2. What psycholinguistics bring to the study of schizophrenia… and vice versa? *Language and Linguistic Compass* (4), 8: 590-604.

Deborah L. Levy, Micheal J. Coleman, Heejong Sung, Fei Ji, Steven Matthysse, Nancy R. Mendell and Debra Titone. 2010. The genetic basis of thought disorder and language and communication disturbances in schizophrenia. *Journal of Neurolinguistics*, 23: 176-192.

Muriel D. Lezak, Diane B. Howieson and David W. Loring. 2004. *Neuropsychological assessment* (4th Edition). Oxford: Oxford University Press.

Michael J. Minzenberg, Beth A. Ober and Sophia Vinogradov. 2002. Semantic priming in schizophrenia: a review and synthesis. *Journal of the International Neuropsychological Society,* 8: 699-720.

Lynne, M. Murphy. 2003. *Semantic relations and the lexicon: antonymy, synonymy and other paradigms*. Cambridge, UK: Cambridge University Press.

Margaret A. Niznikiewicz, Michelle Friedman, Marta E. Shenton, Martina Voglmaier, Paul G. Nestor, Melissa Frumin, Larry Seidman, John Sutton and Robert W. McCarley. 2004. Processing sentence context in women with schizotypal personality disorder: an ERP study. *Psychophysiology*, 41: 367-371.

Carita Paradis, Caroline Willners and Steven Jones. 2009. Good and bad opposites: using textual and psycholinguistic techniques to measure antonym canonicity. *The Mental Lexicon*, (4), 3: 380-429.

Carita Paradis and Caroline Willners. 2011. Antonymy: from convention to meaning-making. *Review of Cognitive Linguistics*, 9: 367-391.

Francesca Pesciarelli, Tania Gamberoni, Fabio Ferlazzo, Leo Lo Russo, Francesca Pedrazzi, Ermanno Melati and Cristina Cacciari. 2014. Is the comprehension of idiomatic sentences indeed impaired in paranoid Schizophrenia? A window into semantic processing deficits. *Frontiers in Human Neuroscience*, 9 October 2014, 8:799, doi:10.3389/fnhum.2014.00799.

Edith Pomarol-Clotet, Tomasina, M. Oh, Keith R. Laws and Peter J. McKenna. 2008. Semantic priming in schizophrenia: systematic review and meta-analysis. *British Journal of Psychiatry*, 192: 92–97.

Manfred Spitzer, Ursula Braun, Leo Hermle and Sabine Maier. 1993. Associative semantic network dysfunction in thought-disordered schizophrenic patients: direct evidence from indirect semantic priming. *Biological Psychiatry*, 34: 864-877.

Debra Titone, Deborah L. Levy and Philip S. Holzman. 2000. Contextual insensitivity in schizophrenic language processing: evidence from lexical ambiguity. *Journal of Abnormal Psychology*, 109: 761-767.

Debra Titone, Philip S. Holzman and Deborah L. Levy. 2002. Idiom processing in schizophrenia: literal implausibility saves the day for idiom priming. *Journal of Abnormal Psychology*, 111: 313-320.

Debra Titone, Maya Libben, Meg Niman, Larissa Ranbom and Deborah L. Levy. 2007. Conceptual combination in schizophrenia: contrasting property and relational interpretations. *Journal of Neurolinguistics*, 20: 92–110.

Joost van de Weijr, Carita Paradis, Caroline Willners and Magnus Lindgren. 2014. Antonym canonicity: temporal and contextual manipulations. *Brain and Language*, 128: 1-8.

Sophia Vinogradov, John H. Poole, Jason Willis-Shore, Beth A. Ober and Gregory K. Shenaut. 1998. Slower and more variable reaction times in schizophrenia: what do they signify? *Schizophrenia Research*, 32: 183-190.

Caroline Willners. 2001. Antonyms in Context. A Corpus-based Semantic Analysis of Swedish Descriptive Adjective. *Travaux de l'Institut de Linguistique de Lund*, 40, Dept. of Linguistics, Lund University.

Using network clustering to uncover the taxonomic and thematic structure of the mental lexicon

Simon De Deyne
University of Adelaide
School of Psychology
5005 Adelaide, Australia
simon.dedeyne@adelaide.edu.au

Steven Verheyen
University of Leuven
Department of Psychology
Tiensestraat 102, 3000 Leuven, Belgium
steven.verheyen@ppw.kuleuven.be

While still influential, the view that concepts are organized as a hierarchical taxonomy as proposed by Rosch (1973) has been challenged on several occasions. For example, some studies have attributed a larger role to thematic relations (Gentner and Kurtz, 2005; Lin and Murphy, 2001), whereas others have stressed the role of affect in structuring word meaning (Niedenthal et al., 1999). A comprehensive account of how these different principles shape and structure meaning in the lexicon is missing, and most studies continue to be biased towards concrete noun categories that fit into hierarchical taxonomies (Medin and Rips, 2005). To capture mental or psychological properties that organize the lexicon for a wide range of concepts and semantic relations, we propose a large-scale semantic network derived from word associations as the basis to uncover what the structural principles are.

1 Network Clustering

Since this is one of the first times the mental lexicon is mapped in its entirety using an extremely extensive word association corpus, an exploratory approach is warranted. To achieve this, network clustering was used as a way to study how the mental lexicon can be structured at different scales and what type of semantic relations dominate its structure. At the basis lies a semantic network derived from a large scale word association corpus including over 12,000 cues and 3.77 million responses (De Deyne et al., 2013). For the purpose of this study, non-dominant word forms were removed (e.g., *apples* was removed if *apple* was also present) resulting in a network of 11,000 words. Next, the recent *Order Statistics Local Optimization Method* (OSLOM) was applied to identify statistically reliable clusters in a directed weighted word associations network (Lancichinetti et al., 2011). This method includes words in the final cluster solution on the basis of statistical criteria and allows for overlapping clusters. Similar to taxonomic theories of knowledge representation, words are grouped in progressively larger clusters, which allows us to evaluate structural properties of the lexicon at different scales. This hierarchical structure is also derived from the data by using a statistical criterion that involves a comparison with an appropriate null-model for the weighted directed graph.

Applying OSLOM to the semantic network resulted in a solution with five hierarchical levels. An overview of this solution is shown in Table 1. There was a large degree of variability in the number of clusters across the five hierarchical levels ranging between 2 and 506 clusters. On average, the p-value of the extracted clusters was low across all levels, indicating that the obtained clusters were unlikely to arise in a comparable random network[1]. There were few homeless nodes at any level, indicating that most words were reliably attributed to a specific cluster. There was also a considerable degree of overlap at all levels relative to the size of the clusters; clusters were more distinct at the more precise levels, where more clusters were obtained. For instance, at the lowest level 1,676 words appeared in multiple clusters, compared to 5,943 at the highest level.

Figure 1 illustrates the obtained clusters with the most prototypical examples of each cluster at various levels. At the most general level, Figure 1 shows two distinct clusters, with one of them containing highly central words with a negative connotation. In order to verify whether this interpretation is supported statistically, we used the valence judgments reported by Moors et al. (2012), which

[1] Default parameters were used in the OSLOM algorithm, except for the p cut-off value. Setting this value depends on the task as it affects the size of the clusters (Lancichinetti et al., 2011). In this application, the cutoff was set at 0.25, because the few clusters in the final solution with high p-values were easy to interpret. Other values of p did not alter the general pattern of results we report here.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

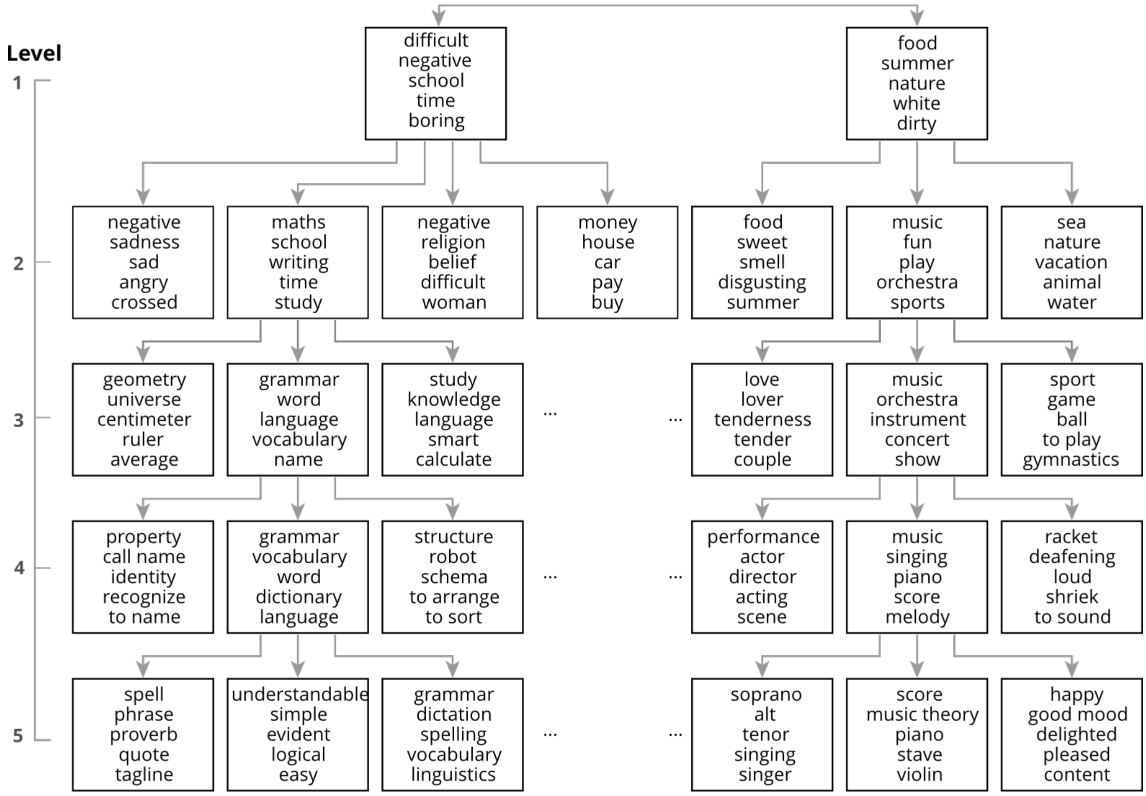


Figure 1: Hierarchical tree visualization of clusters in the lexicon with five most central members in terms of cluster in-strength.

Table 1: Overview of the hierarchical cluster structure showing five levels (Level 1 is broadest, Level 5 is most precise). The statistics include total number of clusters N, average cluster size $\langle N_c \rangle$ and its standard deviation, number of homeless nodes $N_{homeless}$, number of nodes member of multiple clusters $N_{overlapping}$, and the average p-value $\langle p \rangle$.

	1	2	3	4	5
N	2	7	37	161	506
$\langle N_c \rangle$	8588	3049	515	112	25
$sd(N_c)$	2112	973	364	66	12
$N_{homeless}$	18	18	39	86	380
$N_{overlapping}$	5943	6956	5263	4717	1676
$\langle p \rangle$	0	0.062	0.04	0.035	0.051

are applicable to 3,642 non-overlapping words in our clusters. The valence judgments differed significantly between our two clusters according to an independent t-test ($t(3640) = 7.367$, $CI = [0.190, 0.327]$). This post-hoc test confirmed our interpretation of a valence difference between the clusters, which brings further support to studies that indicated valence is the most important dimension in semantic space (De Deyne et al., 2014; Samsonovic and Ascoli, 2010) and empirical findings highlighting affect-based category structure (Niedenthal et al., 1999).

At Levels 2 to 4, the meaning clusters become increasingly more concrete. For instance, Level 2 shows that the "negative" cluster in Level 1 includes clusters with abstract words or words related to human culture (*school*, *money*, *religion*, *time*,...) which are now differentiated from a purely negative cluster with central members like *negative*, *sad*, and *crossed*. The subdivisions of the "positive" cluster involve the central nodes *nature*, *music*, *sports*, and *food*, which might be interpreted as covering sensorial information and natural kinds.

At the lowest level, 506 clusters were identified, with an average size of 25 words. A total of 1,676 words occurred in multiple clusters; at least a part of them because of homonymy (e.g., *bank*) or polysemy (e.g., *language*, assigned to clusters about nationality, speech, language education, and communication). Most importantly, inspection of the content of all clusters exhibited a widespread thematic structure: the clusters were often composed of both nouns (*racket*), adjectives (*loud*), and verbs (*to sound*), which does not reflect a pure taxonomy of entitities, but also includes properties and actions.

2 Evaluating Taxonomic Structure

To test whether the clusters provide evidence for a hierarchical taxonomic view along the lines of Rosch and colleagues (Rosch, 1973) or support an alternative view based on thematic relations identified in the previous section, data from an exemplar generation task from Ruts et al. (2004) was used. In this task, 100 participants generated as many exemplars they could think of for six artifact categories (CLOTHING, KITCHEN UTENSILS, MUSICAL INSTRUMENTS, TOOLS, VEHICLES, and WEAPONS) and seven natural kinds categories (FRUIT, VEGETABLES, BIRDS, INSECTS, FISH, MAMMALS, and REPTILES). If the clusters in the word association network group together different types of birds, vehicles, fruits, and so on, this would indicate a taxonomic organization of semantic memory. For each category, we investigated the size of the best matching cluster and calculated precision and recall in terms of the F-measure for clustering performance.

A taxonomic-like organization would be evident in clusters with high precision and recall, resulting from many true positives and few false positives and false negatives. For instance, if the cluster corresponding to the category BIRDS contained *robin* (a true positive) and did not contain *spoon* (a true negative), that would increase the F-score. Conversely, if it contained *guitar* (a false positive) or did not contain *ostrich* (a false negative), that would decrease the F-score. This way, high F-scores should reflect categories that are not overly specific (many false negatives) or general (many false positives).

On average, the best matching clusters were found at Level 5. The results for each category are shown in Table 2. The average number of members in the exemplar generation task was on average 41 for the seven natural kinds categories, which is in the same range as the average best matching cluster size of 42. For artifacts the generated categories included on average 55 members, which was somewhat larger than the obtained average cluster size of 37.

The resulting F-values were on average 0.48 for the natural categories and 0.28 for the artifacts, indicating only limited support for the presence of taxonomic categories. The highest values were obtained for FISH (F = .57) and REPTILES (F = .65) where most items in the clusters were true category members.

Table 2: F-values and cluster sizes for items generated for 13 concrete noun categories. N_{human} is the category size based on the exemplar generation task; N_c is the size of the best-matching cluster; F captures precision and recall according to the human categories for the full network. F' is calculated from a network that excluded potential thematic information. F-values are fairly low, indicating lack of correspondence between the clusters and the taxonomic categories. Excluding thematic information results in F' values that do capture taxonomic information.

Category	N_{human}	N_c	F	F'
FRUIT	40	50	0.47	0.84
VEGETABLES	35	58	0.50	0.90
BIRDS	53	63	0.53	0.90
INSECTS	40	34	0.46	0.68
FISH	37	48	0.57	0.91
MAMMALS	61	21	0.20	0.76
REPTILES	21	22	0.65	0.51
Mean	41	42	0.48	0.79
CLOTHING	46	70	0.35	0.80
KITCHEN UT.	71	18	0.20	0.66
MUSIC INSTR.	46	24	0.37	0.89
TOOLS	73	56	0.25	0.76
VEHICLES	46	28	0.16	0.73
WEAPONS	46	25	0.37	0.88
Mean	55	37	0.28	0.79

Inspecting the false positives for each of the clusters in Table 3 confirms the validity of the approach as in the majority of the cases the superordinate label (e.g., *fruit*, *tools*, etc.) was the most central member of each cluster. The remaining intrusions were thematic in nature (e.g., FRUIT: pick, BIRDS: nest), thus confirming our earlier exploratory findings.

One potential response to the previous analyses relates to the nature of the data upon which they are based. Perhaps the word association task simply fails to capture taxonomic information, and if so, the results of these analyses are simply an artifact of the choice of task. Alternatively, perhaps the "failure" arises because the word association task is more general than the tasks typically used to study taxonomic categories.

There is some evidence that a different choice of task would produce different results. For instance, much of the work on taxonomic organization relies on tasks in which participants are asked to list features of entities (McRae et al., 2005; Ruts et al., 2004). One could argue that feature generation is

Table 3: Top 5 false positives ordered by cluster in-strength per category. Most of the false positives are thematic in nature. For instance, false positives for BIRDS include *beak*, *egg*, *nest*, and *whistle*.

Category	**1**	**2**	**3**	**4**	**5**
FRUIT	fruit	juicy	pit	pick	summer
VEGETABLES	vegetable	healthy	puree	sausage	hotchpotch
BIRDS	bird	beak	nest	whistle	egg
INSECTS	insect	vermin	beast	crawl	animal
FISH	fish	fishing	rod	slippery	water
MAMMALS	rodent	gnaw	tail	pen	marten
REPTILES	reptile	scales	animal	tail	amphibian
CLOTHING	clothing	fashion	blouse	collar	zipper
KITCHEN UT.	cooking	kitchen	stove	cooker hood	burning
MUSICAL INSTR.	wind instrument	to blow	fanfare	orchestra	harmony
TOOLS	tools	carpenter	carpentry	wood	drill
VEHICLES	speed	drive	vehicle	motor	circuit
WEAPONS	sharp	stab	blade	point	stake

a constrained version of the word association task, and the key difference is the number of thematic responses one gets in both procedures. Similarly, feature generation stimuli are usually restricted to concrete nouns, which places restrictions on what words *can* be grouped together. In other words, the tendency to find taxonomic categories may be a result of restricting the task.

To test this idea, we used the word association data to construct a network that included *only* those 588 words that belonged to one of the taxonomic categories. Moreover, in order to approximate the "shared features" measure that is more typical of feature generation tasks, we computed the cosine similarity between pairs of words. That is, words that have the same associates are deemed more similar, and this similarity was used to weight the edges in the restricted network.[2] We then applied the clustering procedure to this restricted network and repeated the analysis from the previous section. The F-statistics from this analysis are reported as the F'-values in Table 2. This time, the results of the clustering show a high degree of agreement with the taxonomic organization, with an average F-value of 0.79. The only exception was REPTILES, which upon inspection appears to reflect a failure to distinguish REPTILES from INSECTS.

The success of this analysis suggests two things. First, the word association task *does* encode taxonomic information, as evidenced by the fact that we are able to reconstruct taxonomic categories. However, the fact that the only way to do so is to mimic all the restrictive characteristics of a feature generation task (e.g., limited word set) is revealing. Taxonomic information is not the primary means by which the mental lexicon is organized: if it were, we should not have to resort to such drastic restrictions in order to uncover taxonomic categories.

In summary, even at the most detailed level of the hierarchy, only limited evidence for a taxonomic view along the lines of Rosch was found, even for typical taxonomic domains like animals. These results suggest that in much of the previous work the pervasive contribution of affective and thematic or relational knowledge structuring might be overlooked by a selection bias in terms of the concepts (nouns, mostly concrete) and semantic relations (predominantly taxonomic). This finding is in line with previous results indicating that network derived similarity estimates account better for human thematic relatedness judgments than for taxonomic relatedness judgments (De Deyne et al., in press). In priming studies, the dominance of thematic over taxonomic structure can also explain facilitation when thematic but not coordinate prime-target pairs are used (Hutchison, 2003). Finally, our findings converge with recent evidence that highlights the role of thematic representations even in domains such as animals (Gentner and Kurtz, 2006; Lin and Murphy, 2001; Wisniewski and Bassok, 1999) whereas previous reports that have stressed taxonomic organization might be more exceptional as they are heavily culturally defined (Lopez et al., 1997), a consequence of formal education (Sharp et al., 1979), or reflect different levels of expertise (Medin et al., 1997).

[2]Note that one could also derive such a similarity-based network for the complete lexicon, which would reflect the similarity between cues rather than their weighted associative strength. We did in fact do this. It produced similar results to the original analysis.

Acknowledgments

This research has been supported by an ARC grant DE140101749 awarded to SDD. SV is a postdoctoral fellow at the Research Foundation - Flanders. A longer version of this work was also submitted to the 37th Annual meeting of the Cognitive Science Society, Pasadena, 2015. We wish to express our gratitude to Dan Navarro and Amy Perfors, who contributed to the longer version of this work.

References

[De Deyne et al.2013] Simon De Deyne, Daniel J. Navarro, and Gert Storms. 2013. Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods*, 45:480–498.

[De Deyne et al.2014] Simon De Deyne, Wouter Voorspoels, Steven Verheyen, Daniel J. Navarro, and Gert Storms. 2014. Accounting for graded structure in adjective categories with valence-based opposition relationships. *Language and Cognitive Processes*, 29(5):568–583.

[De Deyne et al.in press] Simon De Deyne, Steven Verheyen, and Gert Storms. in press. The role of corpus-size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*.

[Gentner and Kurtz2005] Dedre Gentner and Kenneth J. Kurtz. 2005. Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, and P. W. Wolff, editors, *Categorization inside and outside the lab.*, pages 151–175. American Psychology Association.

[Gentner and Kurtz2006] Dedre Gentner and Kenneth J. Kurtz. 2006. Relations, objects, and the composition of analogies. *Cognitive Science*, 30:609–642.

[Hutchison2003] Keith A. Hutchison. 2003. Is semantic priming due to association strength or feature overlap? *Psychonomic Bulletin and Review*, 10:785–813.

[Lancichinetti et al.2011] Andrea Lancichinetti, Filippo Radicchi, José J Ramasco, and Santo Fortunato. 2011. Finding statistically significant communities in networks. *PloS one*, 6(4):e18961.

[Lin and Murphy2001] Emilie L. Lin and Gregory L. Murphy. 2001. Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 1:3–28.

[Lopez et al.1997] Alejandro Lopez, Scott Atran, John D Coley, Douglas L Medin, and Edward E Smith. 1997. The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive psychology*, 32(3):251–295.

[McRae et al.2005] Ken McRae, George S Cree, Mark S Seidenberg, and Chris McNorgan. 2005. Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37:547–559.

[Medin and Rips2005] Douglas L. Medin and Lance J. Rips. 2005. Concepts and categories: memory, meaning, and metaphysics. In K. Holyoak and R. Morrison, editors, *The Cambridge Handbook of Thinking and Reasoning*, pages 37–72. Cambridge University Press, Cambridge, UK.

[Medin et al.1997] Douglas L. Medin, Elizabeth B. Lynch, John D. Coley, and Scott Atran. 1997. Categorization and reasoning among tree experts: Do all roads lead to rome? *Cognitive psychology*, 32(1):49–96.

[Moors et al.2012] Agnes Moors, Jan De Houwer, Dirk Hermans, Sabine Wanmaker, Kevin van Schie, Anne-Laura Van Harmelen, Maarten De Schryver, Jeffrey De Winne, and Marc Brysbaert. 2012. Norms of valence, arousal, dominance, and age of acquisition for 4,300 dutch words. *Behavior research methods*, pages 1–9.

[Niedenthal et al.1999] Paula M. Niedenthal, Jamin B. Halberstadt, and Åse H. Innes-Ker. 1999. Emotional response categorization. *Psychological Review*, 106(2):337.

[Rosch1973] Eleanor Rosch. 1973. Natural categories. *Cognitive Psychology*, 4:328–350.

[Ruts et al.2004] Wim Ruts, Simon De Deyne, Eef Ameel, Wolf Vanpaemel, Timothy Verbeemen, and Gert Storms. 2004. Dutch norm data for 13 semantic categories and 338 exemplars. *Behaviour Research Methods, Instruments, and Computers*, 36:506–515.

[Samsonovic and Ascoli2010] Alexei V. Samsonovic and Giorgio A Ascoli. 2010. Principal semantic components of language and the measurement of meaning. *PloS one*, 5(6):e10921.

[Sharp et al.1979] Donald Sharp, Michael Cole, Charles Lave, Herbert P Ginsburg, Ann L Brown, and Lucia A French. 1979. Education and cognitive development: The evidence from experimental research. *Monographs of the society for research in child development*, pages 1–112.

[Wisniewski and Bassok1999] Edward J. Wisniewski and M. Bassok. 1999. What makes a man similar to a tie? *Cognitive Psychology*, 39:208–238.

Classification of German verbs using nouns in argument positions and aspectual features

Michael Richter
Radboud University Nijmegen
mprrichter@t-online.de

Jürgen Hermes
University of Cologne
hermesj@uni-koeln.de

Abstract

This paper provides evidence that aspectual verb classes (Vendler, 1967) can be induced from nominal fillers in argument positions and aspectual features. We classified 35 German verbs in a supervised learning procedure using a support vector machine classifier and a classification into five aspectual classes (Richter and van Hout, 2015) as gold standard and observed excellent and substantial agreements.

1 Introduction

This study aims to empirically validate aspectual verb classes in German using large corpus data. Siegel (1997) and Siegel and McKeown (2000) induced the two aspectual classes *states* and *events* in the frame of a vector space model from corpora, however an induction of the complete Vendlerian typology has not yet been undertaken. We hypothesize that aspectual verb classes can be automatically induced from the classified nominal fillers in the argument position of verbs. Our hypothesis refers to the *Distributional Hypothesis* (Rubenstein and Goodenough, 1965; Schütze and Pedersen, 1995; Landauer and Dumais, 1997; Pantel, 2005) which says that semantically related linguistic elements appear in semantically related contexts. The present study in the framework of a vector space model is also driven by the *Statistical Semantics Hypothesis* (Weaver, 1955; Furnas et al., 1983; Turney and Pantel, 2010) which states that linguistic meaning can be derived from statistic linguistic patterns. In order to test our hypothesis, we took a test set of verbs from Schumacher (1986) and determined the nominal fillers and their classes in argument positions. That is, in subject, direct, indirect, and prepositional object positions by parsing a very large German corpus. As gold standard we used the aspect-based classification of Richter and van Hout (2015) into five classes which extends the typology of Vendler (1967), i.e. a*ccomplishments,* a*chievements, states* and *activities* by the class *accomplishments with an affected subject*.

This classification into five aspectual verb classes was derived by combining two user based classifications induced by cluster analyses from raters' judgments and associations with stimulus verbs and two usage based classifications induced from corpus data (Richter and van Hout, 2015). We took this classification as gold standard as we were interested in the correlation of the semantics of the nominal fillers in argument positions of verbs and the aspectual properties of verbs thereby following Klein (2009) who defines aspect as a grammatical category of verbs.

In the present study we represent verbs as vectors that consist of nouns in argument positions separated into areas according to their noun classes, which were induced by cluster analyses from similarity data. In addition, we added aspectual features as defined by Vendler (1967) to the vectors in order to compare the predictive power of the noun classes in argument positions against the predictive power of the aspectual features, respectively. The test set of verbs was classified in a supervised learning procedure using a support vector machine (SVM) classifier. In order to compare the results with aspectual verbs classes as gold standard with a gold standard-classification based on concrete semantic categories compatible with Schumacher's typology (1986) of German verbs, we trained the SVM classifier with a classification based on ten verb classes which comprises classes such as *verbs of consumption* and *verbs of handicraft working* (Richter and van Hout, 2015). This classification was induced

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

from the co-occurrence data bank (CCDB) of the Institut für Deutsche Sprache (IDS).[1]

2 Method

We classified 35 common German verbs used by Schumacher (1986), who defines seven lexical semantic macrofields and 30 subfields. We chose the verbs from all subfields, the only criterion being the representation of every subfield in order to cover the total semantic range of Schumacher's typology (1986). We checked the frequency of the verbs in the first one million sentences containing at least one of our selected verbs of the web based 880 million word SDEWAC corpus[2]. The verbs of our test set occurred in more than one million sentences with a mean frequency of approximately 30,000 occurrences per verb. 66 percent of the verbs was in the interval between 5,000 and 40,000 occurrences, the more frequent outliers being *müssen* 'to must' with 500,965 and *halten für* 'to take so./sth. for so./sth.' with 123,595 occurrences. We added five verbs; *hämmern* 'to hammer', *schneiden* 'to cut', *aufessen* 'to eat up', *laufen* 'to walk/to run', and *zersägen* 'to saw into pieces' since these verbs since a previous study (Richter and van Hout 2015) showed (i) that *laufen* 'to walk/to run' and *zersägen* 'to saw into pieces' are typical activity and accomplishment verbs respectively and (ii) that *aufessen* 'to eat up' is a typical accomplishment with an affected subject verb. *Schneiden* 'to cut' and *hämmern* 'to hammer' were ambiguous (Richter and van Hout, 2015), but we decided to classify in this study the former as accomplishment and the latter as a process verb.

In order to determine the verbs' arguments we parsed at most 30.000 sentences per verb using the Mate-Tools dependency-parser (Bohnet, 2010)[3]. The whole code we used for filtering and parsing the sentences, and aggregating the actants and aspectual features (see below) is available at GitHub.[4] The 35 verbs of our test set (Richter and van Hout, 2015) are represented as 139 dimensional vectors containing the 30 most frequent nouns in the verbs argument positions: subjects, direct objects, indirect objects and prepositional objects. The nouns were weighted by the TF-IDF measure and classified by cluster analyses carried out on a matrix with similarity values taken from the co-occurrence data bank (CCDB) of the Institut für Deutsche Sprache (IDS).[5] On the matrix of the similarity values, a cluster analysis with Ward's method and Euclidean distance was carried out. According to the *Bayesian Information Criterion* there are two optimal noun classes for all arguments. We interpreted the resulting noun classes using our intuition thereby applying the criterion of animacy (Croft, 2003; Aissen, 2003): The resulting two noun classes can be interpreted as denoting predominantly animate and inanimate things, respectively class $1_{[+animate]}$ for instance, contains nouns such as *Arzt* 'doctor' *Lehrkraft* 'teacher' and class $2_{[-animate]}$ contains nouns such as *Entwicklung* 'development', *Organisation* 'organization' and *Wahrnehmung* 'perception'. The verbs' vectors consist of areas for each argument type. There are four areas in total and each area is split into areas for each noun class as is depicted in (1):

$$\vec{v} = \begin{pmatrix} wn_1c_1 \\ wn_2c_1 \\ . \\ . \\ . \\ wn_n\ c_1 \\ wn_1\ c_2 \\ wn_2\ c_2 \\ . \\ . \\ . \\ wn_nv_n \end{pmatrix}$$

$(wn_i\ c_j: Weight\ of\ noun\ n_i\ in\ noun\ class\ \ c_j\ )$

Figure 1. *Dimensions of verb vectors*: *Weighted verbs in noun class areas.*

In addition, the vectors were completed by aspectual features that Vendler (1967) suggested in order to distinguish aspectual verb classes. The aspectual features indicate, for instance, whether the verbs occur in sentences with temporal specifications of duration or a limited time span with prepositions *in* and *for*, respectively, as in *he wrote the letter in an hour* versus *he wrote the letter for an hour*, whether the verbs can be embedded by matrix verbs such as *persuade* or whether they occur in imperative forms. In order to classify the 35 verbs we used a

[1] http://corpora.ids-mannheim.de/ccdb/. The similarity values were provided by Cyril Belica.
[2] The SdeWaC Corpus is available at the WaCky Corporadownload page at http://wacky.sslmit.unibo.it/doku.php?id=corpora
[3] See https://code.google.com/p/mate-tools/
[4] https://github.com/spinfo/verbclass

[5] The similarity values were provided by Cyril Belica.

SVM classifier with a non-linear kernel which achieved the best results.
We first trained the SVM using the classification of Richter and van Hout (2014) as a gold standard and tested it with a 10-fold cross-validation. The gold standard classification in detail:

1. accomplishments:
aufbauen auf 'to build on/to be based on', *herstellen* 'to produce', schneiden 'to cut', *zersägen* 'to saw into pieces', *verlängern* 'to extend', *mitteilen* 'to tell/to inform', *übermitteln* 'to communicate/to forward', *verhindern* 'to prevent', *abgrenzen* 'mark off/to define'
2. accomplishments with affected subject:
untersuchen 'to examine', *bedenken* 'to consider', *erörtern* 'to debate', *nachprüfen* 'to ascertain/to check', *aufessen* 'to eat up', *essen* 'to eat'
3. activities:
laufen 'to walk/to run', *eingehen auf* 'to respond to so./sth.', *hämmern* 'to hammer', *ansteigen* 'to increase'
4. achievements:
einschlafen 'to fall asleep', *vergehen* 'to go (by)/to pass/to diasappear', *übersehen* 'to overlook', *verlieren'to* loose', *anfangen* 'to begin', *abweichen* 'to deviate', *sich orientieren* an 'to be geared to', richten auf 'to direct towards/to focus'
5. states:
existieren 'to exist', *fehlen* 'to lack', *müssen* 'to must', *halten für* 'to take so./sth.for so./sth.', *folgen aus* 'to follow from', *angehören* 'to belong to', *übereinstimmen* 'to agree', *betreffen* 'to concern', *abweichen* 'to deviate' , *verhindern* 'to prevent'

The classification into classes of concrete lexical properties which we induced from the co-occurrence data bank (see above) is given below (the class labels are compatible with Schumacher's labels and are assigned using our linguistic intuitions; class 10 is incoherent and could not be labelled):

1. verbs of activities manipulating a substance (normally with a tool):
hämmern 'to hammer', *schneiden* 'to cut', *zersägen* 'to saw into pieces'
2. verbs of consumption:
aufessen 'to eat up', *essen* 'to eat'
3. verbs of difference, 'negative' processes, non-existence:
*müssen** 'to must', *einschlafen** 'to fall asleep' , *vergehen* 'to go (by)/ to pass/to disappear', *übersehen* 'to overlook', *fehlen* 'to lack', *verlieren* 'to loose', *verhindern* 'to prevent', *abgrenzen* 'to mark off/ to define', *abweichen* 'to deviate
4. verbs of transfer (of information):
mitteilen 'to inform', *übermitteln* 'to communicate/to forward'
5. verbs of examination (by mental activity):
nachprüfen 'to ascertain/to check', *erörtern* 'to debate', *untersuchen* 'to examine'
6. verbs of production:
aufbauen auf 'to build on'/acc to be based on', *herstellen* 'to produce'
7. verbs of beginning and rising processes:
anfangen 'to begin', *ansteigen* 'to rise/ to increase'
8. verbs of discussion and consideration:
betreffen 'to concern', *bedenken* 'to consider', *eingehen auf* 'to respond to so./sth.', *halten für* 'to take, *richten auf* 'to direct towards', *orientieren an* 'to be geared to'
9. verbs of membership and agreement:
angehören 'to belong to', *übereinstimmen mit* 'to agree with'
10. *folgen aus* 'to follow from', *laufen* 'to walk/to run', *existieren* 'to exist', *verlängern* 'to extend'

2.1 Results

In order to evaluate the consistency of the comparisons of the classifications against the gold standards we calculated both accuracy and Cohen's kappa. The latter measure considers the number of classes which differ in the two gold standards and, in addition, gives the significance levels.
Taking the classification with five aspectual verbs classes as gold standard the subject feature clearly outperforms the remaining features with .857 accuracy (which means that 30 of 35 verbs were classified correctly) and κ = .812. Kappa values above .61 are characterized as substantial, above .81 as almost perfect agreement and therefore highly significant. The combinations subject-direct object-prepositional object-aspectual features and subject-direct object-aspectual features yield .828 accuracy, κ = .775 and κ = .773, respectively. The combinations subject-prepositional object-aspectual features, subject-direct object-prepositional object and subject- aspectual features yield .8 accuracy each with κ = .741, κ = .739 and κ = .71, respectively.

In contrast the remaining features, including the aspectual feature which yields .514 % accuracy, with $\kappa = .317$ (fair agreement), perform poorly.
Taking the classification according to concrete semantic properties into ten classes as the gold standard we observed that the hierarchy remains almost the same, the subject feature outperforms the remaining features. However, the accuracy is considerably lower compared to the classification with 5 aspectual verb classes. The subject achieves .657 accuracy, $\kappa = .573$. The combinations subject-direct object-aspectual features and subject-direct object-prepositional object yield .628 accuracy with $\kappa = .458$, followed by the combinations subject-direct object and subject-aspectual features with .6 accuracy each and $\kappa = .495$. These combinations exhibit a moderate agreement. Again, the aspectual feature performs poorly with .428 accuracy, $\kappa = .266$ which is a fair agreement. In figure 2 the accuracy of the argument and aspectual features for the comparisons against both gold standard classifications are given.

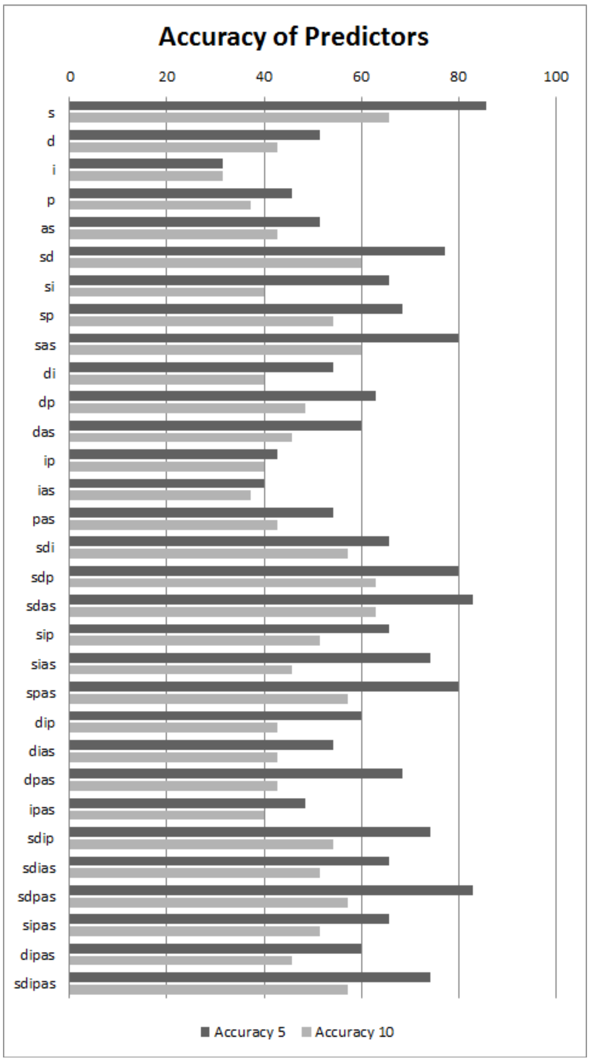


Figure 2. *Accuracy of the argument and aspectual features using five aspectual verb classes vs. ten classes with concrete lexical properties as gold standard.*

Note: s: subject, d: direct object, i: indirect object, p: prepositional object, as: aspectual features, and combinations of predictors, for instance, das: direct object and aspect, sp: subject and prepositional object.

3 Conclusion

The study provides evidence for the hypothesis that aspectual verb classes can be induced from classified nominal fillers in argument positions. For the five aspectual verb classes used as the gold standard (Richter and van Hout, 2014) it turned out that noun classes in subject positions have the highest predictive power compared to the nouns in the remaining argument positions and the aspectual features derived from Vendler (1967). This result is surprising since the Vendlerian aspectual categories were formulated in order to distinguish aspectual classes. Future research should explore a comparison of the predictive power of nominal and aspectual features.
Using a classification into concrete lexical fields as the gold standard of the predictive values we observed a considerable decrease in the predictive values indicated by the lower kappa values. We explain this result by the difference in information provided by the argument structures of the verbs in the 5 class-gold standard classification in contrast to the information provided by co-occurrences that is, lexical information of any type in the context of verb in the 10-class gold standard classification.
The results of this study show that: 1. Aspectual verb classes can be empirically validated, 2. Classified nouns in subject argument positions are reliable predictors of aspectual verb classes, i.e. the meaning of nouns in combination with their noun classes correlates with aspectual parts of the verbal meaning. In order to confirm these results further research with an extended test set of verbs is needed.

Acknowledgments

Roeland van Hout suggested to evaluate the classifications with Cohen‘s Kappa and was very helpful in the calculations of the κ-values.

References

Judith Aissen 2003. Differential Object Marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, 21: 435 – 483.

Bernd Bohnet 2010. Very High Accuracy and Fast Dependency Parsing is not a Contradiction. The 23rd International Conference on Computational Linguistics (COLING 2010), Beijing, China.

William Croft 2003. *Typology and Universals*. Cambridge University Press (2nd edition).

George W. Furnas, Thomas K. Landauer, Louis M. Gomez, and Susan T. Dumais 1983. Statistical semantics: Analysis of the potential performance of keyword information systems. *Bell System Technical Journal*, 62 (6): 1753 – 1806.

Wolfgang Klein 2009. How time is encoded. In W. Klein and P. Li (eds.), *The expression of time*, 39 – 82. Berlin: Mouton de Gruyter.

Thomas K. Landauer and Susan T. Dumais 1997. A solution to Plato's problem: The Latent Semanctic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104: 211 – 140.

Patrick Pantel 2005. Inducing ontological co-occurrence vectors. In *Proceedings of Association for Computational Linguistics* (ACL-05): 125 – 132.

Michael Richter and Roeland van Hout 2015. A classification of German verbs using empirical language data and concepts of Vendler and Dowty. To appear 2015 in *Sprache und Datenverarbeitung – International Journal for Language Data Processing.*

Herbert Rubenstein and John B. Goodenough 1965. Contextual correlates of synonomy. In *Communications of the ACM,* Vol 8 (10): 627 – 633.

Helmut Schumacher 1986. *Verben in Feldern. Valenzwörterbuch zur Syntax und Semantik deutscher Verben.* Berlin & New York: De Gruyter.

Hinrich Schütze and Jan Pedersen 1993. A vector model for syntagmatic and paradigmatic relatedness. In *Making Sense of Words:* 104 – 113. Ninth Annual Conference of the UW Centre for the New OED and Text Research, Oxford.

Peter D. Turney and Patrick Pantel 2010. From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, 37: 141 – 188.

Zeno Vendler 1967. *Linguistics in Philosophy*. Ithaka /New York: Cornell University Press.

Warren Weaver 1955. Translation. In W. N. Locke, D.A. Booth, *Machine Translation of Languages*: 15 – 23. Cambridge Mass.: MIT Press.

Processing of cognates in Croatian as L1 and German as L2

Maja Anđel
University of Zagreb
mandel@ffzg.hr

Jelena Radanović
Laboratory for Experimental Psychology
University of Novi Sad
jelena.radanovic@uns.ac.rs

Laurie Beth Feldman
SUNY, The University at Albany
Haskins Laboratories
lfeldman@albany.edu

Petar Milin
University of Novi Sad
Eberhard Karls University Tübingen
pmilin@uni-tuebingen.de

1 Introduction

Cognates are defined as words similar in form and meaning across two languages. Similarity in form may range from full orthographic overlap, as in English film – German Film, to partial overlap, as in English chapel – German Kapelle. Some pairs of cognate words developed historically from a common ancestor word, whereas others emerge when languages come into contact and loan each other words. Language users are typically unaware of such diachronic pressures. When acquiring a second language (L2) they can only perceive shared elements between L1 and L2.

Cognates help explain the nature of lexical processing and the manner in which elements from the two languages interact. Different measures have been used to explore cognate processing and representation, including ERP (Midgley et al., 2011; Peeters et al., 2013; Strijkers et al., 2009), latencies in single word (Dijkstra et al., 2010; Lemhöfer and Dijkstra, 2004), and primed lexical decision (De Groot and Nas, 1991), eye-movements (Mulder et al., 2011; Rosselli et al., 2012), and scores on standardized tests (Kelley and Kohnert, 2012; Pérez et al., 2010). Taken together, empirical findings support the claim that cognates are processed differently from noncognate words. Despite the fact that the aforementioned experimental measures and techniques diverge, the conclusion is similar both in language production and in language comprehension (Dijkstra et al., 2010, for an overview). Nevertheless, results do differ with respect to a range of details, including the direction as well as the magnitude of the cognate effect. Specifically, most studies find facilitation in the processing of cognates in L2 (Dijkstra et al., 1999; Lemhöfer and Dijkstra, 2004; Van Hell and De Groot, 2008), but results are less clear when it comes to the effect of cognates in L1. For example, Van Hell and Dijkstra (2002) and Duyck (2005) reported cognates facilitation in the dominant language, while Kroll et al. (2002) reported small cognate inhibition in an L1 naming task, and Caramazza and Brones (1979) failed to find such an effect at all.

In the present study we sought to examine the influence of cognates on lexical processing in a visual lexical decision task, using L1/L2 language pairs that belong to different subgroups of Indo-European languages: Slavic L1 and Germanic L2. The aim was to carefully replicate recent findings from a study by Radanović, Feldman, and Milin (2014). Crucially, their study showed quite a complex pattern of effects that included a three-way interaction of language (Serbian L1 vs. English L2) by cognate status (cognate vs. noncognate) by word frequency (as a numerical predictor – covariate). Cognates were processed faster than noncognates in L2, but, surprisingly, significantly slower than noncognates in L1. Furthermore, the size of the effect was greater when word frequency was low.

Because this pattern of effects differs from what is typically reported in the literature, we designed a replication of the Radanović et al. study and followed their method and design, this time using another contrasting pair of languages: Croatian (L1) and German (L2).

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

2 Experiment

Late bilinguals of German ($N = 69$) – students of German with Croatian as their L1, participated in a visual lexical decision experiment. There were two forms of the experiment (in Croatian and in German), and students were randomly assigned to one version. The entire experiment (materials and instruction) was in one language and presentation sequence was randomized for each participant.

In preparation for their study, Radanović et al. (2014) also conducted a normative survey with 1000 Serbian – English translation equivalents ranging from pairs consisting of completely different words (e.g., priča – story) to the identical cognates (e.g., drama – drama). They then selected 400 noun pairs covering a wider range of ortho-phonological similarity between L1/L2 words, using both subjective similarity ratings as well as Levenshtein distance. In the present study we made use of 344 of the previously rated word pairs, and constructed the same number of pseudowords. All of the selected 344 pairs fitted nicely for the present purposes of studying Croatian – German cognates, consistently ranging from perfect cognates to orthographically different words. We reused the same noun pairs to allow for strict comparisons of the experimental data.

2.1 Results

We calculated normalized Levenshtein distance measure for pairs of nouns used in two forms of the present experiment. Similarly to the study of Radanović et al., the distribution of the Levenstein distance measure was strictly bimodal, and, as before, the modes matched cognate vs. noncognate distinction. That allowed us to further use a dummy-coded variable *cognate* (TRUE/FALSE), same as in the original study (Radanović et al., 2014).

Furthermore, we transformed the measures to ensure a better approximation to a Gaussian distribution. Word frequencies and word length were log-transformed, while an inverse transformation was applied to response latencies, following Baayen and Milin (2010).

As a last step, we excluded a small number of the extreme outliers (0.07%) from further analysis based on the visual inspection of the reaction time distribution.

The data were analyzed with Linear Mixed Effect Modeling (LMM), in the `R` software environment for statistical computing (R Core Team, 2014), with the `lme4` and the `lmerTest` packages (Bates et al., 2014; Kuznetsova et al., 2014). The refitted model (after removing residual values greater than 2.5 of absolute standardized units), revealed a significant effects of the control predictors, in the expected direction: facilitation from order of a presentation ($\beta = -.044$; $SE_\beta = .007$; $t = -6.42$; $Pr(> |t|) < .0001$), and inhibition from the word length ($\beta = .211$; $SE_\beta = .023$; $t = 9.33$; $Pr(> |t|) < .0001$). Also, there was a significant effect of the lexicallity of the previous word, where stimuli preceded by a word were recognized faster than those preceded by a pseudoword ($\beta = -.077$; $SE_\beta = .005$; $t = -14.36$; $Pr(> |t|) < .0001$).

Most interestingly, the model revealed a significant three-way interaction between word frequency, language and cognate status ($\beta = .053$; $SE_\beta = .012$; $t = 4.44$; $Pr(> |t|) < .0001$). The observed interaction is an almost exact replication of the three-way interaction reported by Radanović et al. (2014): cognates are processed faster than noncognates in German (L2), but slower than noncognates in Croatian (L1), and the size of the effect is attenuated for high frequency words. This pattern of results is depicted in Figure 1.

With regards to the random-effects structure, by-participant and by-item adjustments to the intercept significantly contributed to the model's goodness-of-fit. Word frequency and trial order needed additional by-participant adjustments for the slopes. Similar by-participant adjustments for the slope were held by the word length, which also revealed significant correlation between adjustments for the intercept and the slope ($r = -.72$), indicating that slower and more careful participants were slowed less as item length increased.

3 Discussion

Radanović, Feldman, and Milin (2014) suggested that cognate facilitation in L2 and inhibition in L1 might be specific to the particular pairing of first and second language and/or to the level of proficiency in the L2. Results of the present study show that the particular L1/L2 combination is not critical in the sense that the same pattern generalized

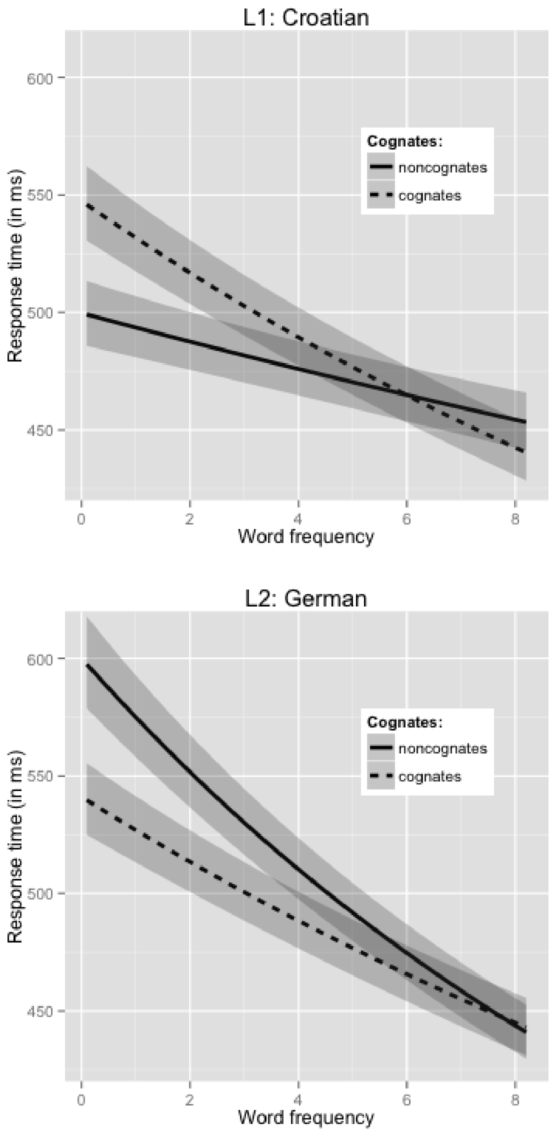


Figure 1: Three-way interaction language by cognates by frequency to reaction time latencies in visual lexical decision task.

to another sample of participants (studying L2 as their major) and another L1/L2 combination. The fact that target frequency played an important role seems more compatible with an account based on proficiency.

However, to find a general explanation and testable hypotheses we turn to learning theory. Arnon and Ramscar (2012), who investigated how adult learners acquire an artificial L2, convincingly demonstrated that "the way in which learning is structured has a considerable impact on what gets learned" (p. 302). In general, knowledge acquisition is codetermined by discrepancies between expectations based on our previous experience and the constellation of cues available in the learning environment. In particular, knowledge in L1 as well as learning history will determine the degree and style of interference that we encounter when learning an L2. This kind of **blocking effect** is well documented in learning theory (Kamin, 1969).

A blocking effect describes failures of learning that arise when a target cue is presented with another cue whose informativity with respect to an outcome has already been established. Arnon and Ramscar (2012) demonstrated in great detail how blocking may influence L2 acquisition when cues from the two languages are competing for the same outcome (a symbolic lexical representation).

Cue blocking does not apply directly to cognates, however, because typically, cues are identical and, thus, cannot compete and/or block each other. Nonetheless, Arnon and Ramscar's general observation regarding the way in which learning is structured helps to make sense of the present findings. All that is needed is to extend it to the distinctive properties of cognates whereby learning entails mapping the very same cues (cognate word forms) onto the same outcome.

Further insights derive from the **highlighting effect** (Kruschke, 2009) on the target cues. First, the theory predicts that contextual (ambient) cues are informative about the learning cues, but not about outcomes (Kruschke and Hullinger, 2010). Therefore, temporal and/or contingency aspects of the situation are useful for discriminating between specific contexts of learning. Second, learning cues can be unambiguous or ambiguous for a particular outcome, and the highlighting effect predicts that early ambiguous and late unambiguous cues are more informative (Kruschke, 2009). Thus, the availability of either L1 or L2 (but not both) provides a context for a given cognate cue (actively present in the sensory input). Given highlighting mechanism, with cognate forms are unambiguous cues we expect facilitation for a latter learned outcome. Conversely, ambiguous cues should facilitate an earlier learned outcome as in an L1 context and, hence, noncognates ought to be faster in L1 but slower in L2.

In summary, in the case of ambiguous cues highlighting is in essence a blocking effect: firstly learned relationships will be favored. This outcome is fully consistent with the account by Arnon and Ramscar (2012). In the case of unambiguous cues, such as cognate words, competition between cues does not emerge and the latter learned relationships will show some preference. Previous research on highlighting indicates that this pattern might be even more pronounced when the cues are verbally (i.e., linguistically) encoded (Kruschke et al., 2005; Kruschke, 2009). This is what present results confirm as well.

Acknowledgments

The research for this paper was financially supported by the Short visit grant 4784 received by NetWordS-09-RNP-089, as well as by the Ministry of Education, Science and Technological Development of the Republic of Serbia grants ON179006 and ON179033. Furthermore, we wish to thank the Ministry of Science, Education and Sports of the Republic of Croatia for supporting this research within the framework of the project 130-1300869-0826 "Croatian and German in Contact – sociocultural aspects and paradigms of communication".

References

[Arnon and Ramscar2012] Inbal Arnon and Michael Ramscar. 2012. Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, 122(3):292–305.

[Baayen and Milin2010] R.H. Baayen and P. Milin. 2010. Analyzing reaction times. *Journal of Psychological Research*, 3(2):12–28.

[Bates et al.2014] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

[Caramazza and Brones1979] Alfonso Caramazza and Isabel Brones. 1979. Lexical access in bilinguals. *Bulletin of the Psychonomic Society*, 13(4):212–214.

[De Groot and Nas1991] Annette MB De Groot and Gerard LJ Nas. 1991. Lexical representation of cognates and noncognates in compound bilinguals. *Journal of memory and language*, 30(1):90–123.

[Dijkstra et al.1999] Ton Dijkstra, Jonathan Grainger, and Walter JB Van Heuven. 1999. Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and language*, 41(4):496–518.

[Dijkstra et al.2010] Ton Dijkstra, Koji Miwa, Bianca Brummelhuis, Maya Sappelli, and Harald Baayen. 2010. How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and language*, 62(3):284–301.

[Duyck2005] Wouter Duyck. 2005. Translation and associative priming with cross-lingual pseudohomophones: evidence for nonselective phonological activation in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6):1340.

[Kamin1969] Leon J Kamin. 1969. Predictability, surprise, attention, and conditioning. In B Campbell and R Church, editors, *Punishment and aversive behaviour*, pages 279–296. Appleton-Century-Crofts.

[Kelley and Kohnert2012] Alaina Kelley and Kathryn Kohnert. 2012. Is there a cognate advantage for typically developing Spanish-speaking English-language learners? *Language, speech, and hearing services in schools*, 43(2):191–204.

[Kroll et al.2002] Judith F Kroll, Erica Michael, Natasha Tokowicz, and Robert Dufour. 2002. The development of lexical fluency in a second language. *Second language research*, 18(2):137–171.

[Kruschke and Hullinger2010] John K Kruschke and Richard A Hullinger. 2010. Evolution of attention in learning. In Nestor Schmajuk, editor, *Computational models of conditioning*, pages 10–52. Cambridge University Press, Cambridge, UK.

[Kruschke et al.2005] John K Kruschke, Emily S Kappenman, and William P Hetrick. 2005. Eye gaze and individual differences consistent with learned attention in associative blocking and highlighting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5):830–845.

[Kruschke2009] John K Kruschke. 2009. Highlighting: A canonical experiment. *Psychology of Learning and Motivation*, 51:153–185.

[Kuznetsova et al.2014] Alexandra Kuznetsova, Per Bruun Brockhoff, and Rune Haubo Bojesen Christensen, 2014. *lmerTest: Tests in Linear Mixed Effects Models*. R package version 2.0-20.

[Lemhöfer and Dijkstra2004] Kristin Lemhöfer and Ton Dijkstra. 2004. Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, 32(4):533–550.

[Midgley et al.2011] Katherine J. Midgley, Phillip J. Holcomb, and Jonathan Grainger. 2011. Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, 23(7):1634–1647.

[Mulder et al.2011] Kimberley Mulder, M Dijkstra, and T Schreuder. 2011. Are eye-fixations in cognate processing dependent on entropy? In *The 17th Meeting of the European Society for Cognitive Psychology [ESCOP 2011]*.

[Peeters et al.2013] David Peeters, Ton Dijkstra, and Jonathan Grainger. 2013. The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, 68(4):315–332.

[Pérez et al.2010] Anita Méndez Pérez, Elizabeth D. Peña, and Lisa M. Bedore. 2010. Cognates facilitate word recognition in young Spanish-English bilinguals' test performance. *Early childhood services*, 4(1):55.

[R Core Team2014] R Core Team, 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Radanović et al.2014] Jelena Radanović, Laurie Beth Feldman, and Petar Milin. 2014. Cognates and frequency in Serbian and English. In *The 3rd Rijeka Days of Experimental Psychology – REPSI*, page 37, Rijeka, Croatia.

[Rosselli et al.2012] Mónica Rosselli, Alfredo Ardila, María Beatriz Jurado, and Judy Lee Salvatierra. 2012. Cognate facilitation effect in balanced and non-balanced spanish–english bilinguals using the boston naming test. *International Journal of Bilingualism*, 18(6):649–662.

[Strijkers et al.2009] Kristof Strijkers, Albert Costa, and Guillaume Thierry. 2009. Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, 20:912–928.

[Van Hell and De Groot2008] Janet G Van Hell and Annette MB De Groot. 2008. Sentence context modulates visual word recognition and translation in bilinguals. *Acta psychologica*, 128(3):431–451.

[Van Hell and Dijkstra2002] Janet G. Van Hell and Ton Dijkstra. 2002. Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9(4):780–789.

The role of grammar factors and visual context in Norwegian children's pronoun resolution

Camilla Hellum Foyn
Department of language and literature
NTNU
camilla.foyn@ntnu.no

Mila Vulchanova
Department of language and literature
NTNU
mila.vulchanova@ntnu.no

Rik Eshuis
Department of language and literature
NTNU
hendrik.eshuis@ntnu.no

1 Introduction

Most personal pronouns have one entry in the mental lexicon, but they can have different referents depending on the context they appear in. They are sometimes fairly ambiguous. There is also evidence that pronoun resolution is impaired in many developmental deficits. Children have to learn how to find the intended referent, but we do not know much about how resolution strategies are acquired. How do visual context and syntactic context influence children's pronoun processing? Using eye-tracking, we investigate for the first time the development of Norwegian children's pronoun resolution competencies in their L1.

2 The study

The participants were monolingual 3-, 5-, and 7-year-old children, as well as a control group of monolingual adults. There were between 25 and 28 participants in each group. In the first of three experiments, they listened to it-cleft sentences with either subject focus (2a) or object focus (2b), while they watched illustrations of two animals (corresponding to the subject and the object) on a screen. It-clefts provide a good environment for testing syntactically expressed focus, and appear to be more frequent in Norwegian than e.g., English (Gundel, 2002). The animals were sometimes shown performing the actions from the cleft-sentences, and other times not (see Table 1 for overview of conditions). Thereafter, the participants heard an ambiguous pronoun sentence (3), and eye-tracking data were collected to determine whether they looked at the subject or object referent. In addition, offline data were collected, by asking the participants to name or point at the pronoun referent (4).

Example of the stimulus sentences:

1. Introduction sentence:

Der er hesten og kaninen
There are the.horse and the.rabbit

2a. Subject-cleft:

Det er hesten som kiler kaninen
It is the.horse that tickles the.rabbit

2b. Object-cleft:

Det er kaninen hesten kiler
It is the.rabbit the.horse tickles

3. Ambiguous pronoun sentence:

Han kan telle til ti
He can count to ten

4. Question sentence:

Hvem kan telle til ti?
Who can count to ten?

Conditions		
1	Subject-cleft	Depicted action
2	Subject-cleft	No depicted action
3	Object-cleft	Depicted action
3	Object-cleft	No depicted action

Table 1: Conditions.

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

3 Earlier findings

According to Järvikivi et al. (2013), German 4-year-olds and adults show a subject preference regardless of which word the it-cleft focuses on. Moreover, children seem to show a weaker subject preference than adults. We expect similar results from our data.

Hartshorne et al. (2014) discovered that 2- to 3-year-olds have a first-mention preference that seldom is detected because they take longer to process. We thus expect young children to show a preference for subject and/or first-mentioned character, albeit at a later time window, whereas adults will show an earlier preference than children.

Bittner and Kuehnast (2011) have found that German 3-year-olds rely more on context-cues than older German children, who more often use syntax-cues. We thus expect that young children will be more influenced by the presence of visual context, whereas older children will be more sensitive to syntactically expressed focus.

4 Results

A mixed design ANOVA showed that 5-year-olds looked more at the subject referent after subject-clefts than object-clefts from 500-1000 ms after pronoun onset ($p > .05$), whereas adults did the same during the first 500 ms ($p = .06$). Adults also showed a general subject preference both offline ($p > .001$) and online ($p > .05$), specifically after subject-clefts as opposed to object-clefts offline ($p > .05$). Moreover, first-look data (first look at subject or object referent after pronoun onset) revealed a stronger subject preference in 7-year-olds after subject-clefts than object-clefts ($p > .05$). We found no significant effect of visual context in the children. However, an interaction effect in adults showed that their stronger subject preference in subject-clefts than object-clefts offline was only present when the action was not depicted ($p > .05$).

5 Conclusions

The results from the time series data suggest that adults process the pronouns faster than children, which supports Hartshorne et al. (2014).

In contrast to the older children, the 3-year-olds performed at chance level in all the different conditions. This may be due to what Hartshorne et al. (2014) found, namely that young children show a first-mention bias that is too slow to detect, or it may simply show that 3-year-olds are too young to comprehend cleft-sentences. In any case, this shows that older children have a stronger preference for the focused referent than younger children do.

Adults showed an overall subject preference regardless of sentence type, except in the condition with object-cleft and no depicted action. This appears to be the only condition that weakens their subject preference, probably because it leaves the subject without syntactic focus and with no visual support. Thus, the effect of syntactic focus and/or a first-mention preference emerges here.

Moreover, depicted action seems to have distracted the adults, since the effect of subject vs. object-clefts offline was only found when the action was not depicted.

In subject-clefts as opposed to object-clefts, 5- and 7-year-olds displayed an online subject preference, although in different manners. Adults also showed this preference, both offline and online. Hence, all these three age groups appear to use syntax cues, but adults seem to be more aware of them, as 5- and 7-year-olds still only reveal their preferences through their gaze behavior. This supports Järvikivi et al.'s (2013) suggestion that children use the same cues as adults, but that they have not fully developed their ability to do so.

References

Dagmar Bittner and Milena Kuehnast. 2011. Comprehension of intersentential pronouns in child German and child Bulgarian. *First Language*, 32(1-2), 176–204.

Jeanette K. Gundel. 2002. Information structure and the use of cleft sentences in English and Norwegian. *Language and Computers*, 39(1), 113–128.

Joshua K. Hartshorne, Rebecca Nappa, & Jesse Snedeker. 2014. Development of the first-mention bias. *Journal of Child Language,* 41(3), 1-24.

Juhani Järvikivi, Pirita Pyykkönen-Klauck, Sarah Schimke, Saveria Colonna, & Barbara Hemforth. 2013. Information structure cues for 4-year-olds and adults: tracking eye movements to visually presented anaphoric referents. *Language and Cognitive Processes*, 0(0), 1–16.

Acknowledgements

The international conference "Word Knowledge and Word Usage: Representations and processes in the mental lexicon" was supported by the European Science Foundation Standing Committee for the Humanities within the framework of the NetWordS ERP programme (May 2011 - April 2015).