PsyComMT 2020

Implications of psycho-computational modelling for Morphological Theory

19th International Morphology Meeting
Vienna, 7 February 2020

Booklet of abstracts

Foreword

The workshop offers an international forum for discussing interdisciplinary prospects of integration between advances in computer modelling of word knowledge and novel theoretical approaches to morphology. Contributions are intended to focus on germane theoretical issues addressing (but not limited to) the following questions:

What are the optimal representation units of human morphological competence and how are they acquired? What role do they play in the way speakers process and store words? Do speakers combine these units in a linear way, as in chaining Markov models, or rather structure them hierarchically, as suggested by the literature on sentence processing? Do they store them in their long-term lexical repository economically, or rather multiply them redundantly, as a function of their context and use? In addition, are these units represented as independent items, or are they mutually related as nodes in a network of paradigmatic relations? What is the contribution of lexical semantics to this picture, and what type of influence is exercised on lexical units by their referential context? What is the status of the processes combining these units into larger units? Are they implemented by a single mechanism? Or should we rather hypothesize that more than one mechanism is in place? What evidence do we have of the anatomical and functional localization of different combinatorial mechanisms in the brain? And in what ways do their neural implementations differ? Given the mounting evidence that children learn words in chunks and that ready-made stretches of assorted words are committed to the long-term memory by speakers, what does this evidence tell us about the separation between Morphology and Syntax for language learning? Can computer modelling sharpen our current understanding of issues of morphological complexity and their impact on lexical acquisition? What is its potential for modelling language learning, contact and change in multilingual contexts?

Vito Pirrelli
Claudia Marzi
Marcello Ferro

Pisa, February 2020

Program Committee

Farrell	Ackerman	University of California, San Diego US
Sabine	Arndt-Lappe	Universität Trier, Germany
Harald	Baayen	Eberhard Karls Universität Tuebingen, Germany
Thomas	Berg	Universität Hamburg, Germany
James	Blevins	George Mason University, Virginia US
Olivier	Bonami	Université Paris Diderot, France
Marc	Brysbaert	Universiteit Gent, Belgium
Davide	Crepaldi	Sissa, Trieste Italy
Walter	Daelemans	University of Antwerp, Belgium
Mirjam	Ernestus	Radboud University, Nijmegen the Netherlands
Marcello	**Ferro (co-chair)**	Institute for Computational Linguistics, CNR Pisa Italy
Dusica	Filipovic Durdevic	University of Novi Sad, Serbia
Christina	Gagné	University of Alberta, Edmonton Canada
Francesco	Gardani	Universität Zurich, Switzerland
Hélène	Giraudo	CNRS, Toulouse France
Kyle	Gorman	City University of New York, US
Emmanuel	Keuleers	Tilburg University, the Netherlands
Robert	Malouf	San Diego State University, US
Christina	Manouilidou	University of Ljubljana, Slovenia
Marco	Marelli	Università Milano Bicocca, Italy
Claudia	**Marzi (co-chair)**	Institute for Computational Linguistics, CNR Pisa Italy
Petar	Milin	University of Birmingham, UK
Carita	Paradis	Lund University, Sweden
Martina	Penke	Universität zu Köln, Germany
Vito	**Pirrelli (co-chair)**	Institute for Computational Linguistics, CNR Pisa Italy
Ingo	Plag	Universität Duesseldorf, Germany
Michael	Ramscar	Eberhard Karls Universität Tuebingen, Germany
Dorit	Ravid	University of Tel Aviv, Israel
David	Saldaña	Universidad de Sevilla, Spain
Abdelhadi	Soudi	Ecole Nationale Supérieure des Mines de Rabat, Morocco
Thomas	Spalding	University of Alberta, Edmonton Canada
Daniela	Traficante	Università Cattolica, Milano Italy
Madeleine	Voga	Université Paul Valéry, Montpellier France
Mila	Vulchanova	NTNU, Trondheim Norway
Gert	Westermann	Lancaster University, UK
Rok	Žaucer	University of Nova Gorica, Slovenia

Representational agnosticism

James P. Blevins[1], Petar Milin[2]

[1])George Mason University USA, [2]) University of Birmingham UK

Among the enduring legacies of Bloomfield's (1926) 'Postulates' is an emphasis on the study of linguistic units. Morphological models have been largely defined by — or, at least, strongly associated with — fundamental claims about unit size. Approaches may differ on whether units are 'abstracted', 'constructed', 'realized', 'retrieved', etc. However, there has been a broad consensus regarding the pivotal dual role of units, as stable 'elements of analysis' and as persistent 'elements of storage'. The primacy of units has in turn contributed to a more general inventory-based perspective, in which the goal of morphological analysis involves the identification of minimal units and the enumeration of well-formed arrangements of units. Relations between units are of mainly subsidiary importance, and in many models serve principally to guide the disassembly and reassembly of forms.

Yet nearly a century of research within unit-based approaches has not produced anything like a consensus regarding the number, size or composition of linguistic units. If anything, the problems of identifying units, especially units with some claim to cross-linguistic validity, seem more recalcitrant than ever. The sceptical positions expressed in Dixon and Aikhenvald (2002) and Haspelmath (2011) suggest that the more that is known about linguistic variation, the less tenable the idealizations that underlie unit-based approaches become. The indeterminacy of units propagates through unit-based approaches, undermining, as Haspelmath (2011) argues, fundamental divisions between syntax and morphology, and even between lexis and grammar.

One solution to this conundrum can be found in the European branch of the 'structuralist' tradition. The conception of structure that Hjelmslev (1948) attributes to Saussure inverts the Bloomfiedian perspective. On this view, relations are of primary importance and units — paradigmatic as well as syntagmatic — are abstracted from observable dimensions of form variation. The resulting conception defines the organization of a linguistic system in terms of networks of relations, not in terms of inventories of segments or sequences:

> *[t]he real units of language are not sounds, or written characters, or meanings: the real units of language are the relata whic*h *these sounds, characters, and meanings represent. The main thing is not the sounds, characters, and meanings as such, but their mutual relations within the chain of speech and within the paradigms of grammar. These relations make up the system of a language, and it is this interior system which is characteristic of one language as opposed to other languages, whereas the representation by sounds, characters, and*

> *meanings is irrelevant to the system and maybe changed without affecting the system.* (Hjelmslev 1948: 27)

A Hjelmslevian perspective accords well with approaches to morphology that have grown out of the information-theoretic and discriminative learning traditions. Morphological analyses in these contemporary 'word and paradigm' approaches are abstracted from the two observable dimensions of form variation: variation in shape and variation in distribution. These dimensions are in turn modeled by means of a pair of basic relations: form discriminability and form predictability/predictiveness.

The present paper sets out some of the recent results that highlight the integrative character of this tradition. A result on the information-theoretic side shows how languages balance global uncertainty by allowing entropy spikes along the syntagmatic or paradigmatic axis but not simultaneously along both. A result on the discriminative learning side offers a characterization of patterns of allomorphy in Polish by constructing parallel Naive Discriminative Learner models of the phonotactic structure and syntagmatic distribution of case forms. In both cases, a comparison of entropy estimations and learning outcomes can also help to clarify the relationship between sources of uncertainty at the system level and at the level of a learning model.

A conception that treats units as abstractions offers some more general benefits. The move from units to relations contributes to resolving the false dichotomy between 'word structure' and 'paradigmatic structure', by integrating both as complementary rather than mutually exclusive types of abstracted structure. The assignment of forms to inflectional paradigms and inflection classes is clearly guided by patterns of word-level variation. Conversely, the 'intramorphological' variation exhibited by inflection class systems encodes information about the shape of other forms of an item and cannot be interpreted purely in isolation. At the same time, both types of structure may have no persistent status, apart from their role in lexical acquisition and processing. As Robins (1959:128) observes, "grammatical statements are abstractions, but they are more profitably abstracted from words as wholes than from individual morphemes". Parallel remarks apply to paradigmatic structures. In information-theoretic measures of informativity, simple forms are of considerably less predictive value than forms paired with 'paradigm cells' that specify sets of grammatical properties. However, paradigms, cells and features are all abstractions. A 'cell' in an inflectional paradigm is essentially a proxy for a distribution class; it is in a syntagmatic context that speakers are able to determine the function of a form. Relabeling cells as 'morphosyntactic representations' amounts to another way of describing distributional classes in terms of features, such as case, that cross-classify form and distribution classes. Inflectional paradigms and inflection classes likewise represent higher-level abstractions.

A relational approach is also compatible with perspectives on the relation between grammar and lexis that emerge from network-based architectures. In a model where units and structures are uniformly abstracted, there is no motivation for representing lexical knowledge in terms of a discrete 'lexicon'. This opens the intriguing possibility that not only the grammar but also the lexis resides in the distribution.

References

Bloomfield, L. 1926. A set of postulates for the science of language. *Language* 2, 153–164.

Dixon, R. M. W. & Aikhenvald, A. Y. Eds. 2002. *Word: A Cross-Linguistic Typology*. Cambridge: CUP.

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia Linguistica* 45(1), 31–80.

Hjelmslev, L. (1948). Structural analysis of language. In Essais linguistiques. *Nordisk Sprog- og Kulturforlag*, 27–35.

Robins, R. H. (1959). In defence of WP. *Transactions of the Philological Society* 58, 116-144.

CAOSS and Transcendence: On role-dependent constituent meanings

Fritz Günther & Marco Marelli,
University of Milano-Bicocca

Many theories on the role of semantics in morphological representation and processing focus on the interplay between the lexicalized meaning of the complex word on the one hand, and the individual constituent meanings on the other hand. However, Libben (2014) argues that the constituent meaning representations involved in compound representation and processing are not the free-word meanings of the constituents, but rather role-dependent, "morphologically transcended" as-constituent meanings. These can be semantically shifted from their corresponding free-word meanings: For example, the word bill, usually referring to laws or checks, has s strong shift towards its rare beak meaning when used as the head -bill in compounds (hornbill, shoebill, razorbill). These as-constituent representations are assumed to be formed through repeated experience with the constituent in its respective role. However, these as-constituent meanings are extremely difficult to operationalize using the standard psycholinguistic toolkit, and existing studies have addressed the issue only indirectly (Libben, Goral, & Baayen, 2018; Smolka & Libben, 2017).

Recent developments in computational modelling however, and more specifically in compositional distributional semantics, allow us to operationalize these as-constituent meanings directly: In the data-driven, fully implemented CAOSS model by Marelli, Gagné and Spalding (2017), compositional compound meanings c are estimated as $c = M*u + H*v$, with u and v being the modifier and head meanings (all meanings are represented as high-dimensional distributional vectors) that are updated via role-dependent weight matrices M and H (estimated via training on all existing compounds in a language) before being combined into the compositional compound meaning. Thus, this model in its very architecture produces a direct, quantitative representation of as-constituent meanings: $M*u$ (the as-modifier meaning) and $H*v$ (the as-head meaning).

We employ these model-derived representations to investigate several empirical hypotheses by Libben (2014) concerning morphological transcendence. First, we present a qualitative analysis of as-constituent meanings that the model predicts to experience strong semantic shift (*worm-* , as in *wormwood*, *wormhole*; or *-bill*, as in *shoebill*, *hornbill*) or almost no semantic shift (*stair-*, as in *staircase*, *stairwell*; or *-boat*, as in *lifeboat*, *motorboat*). Second, we test and confirm the hypothesis that modifiers, on average, experience stronger semantic shift than

heads, but that they remain linked through their free-word meaning. Third, we test and confirm the hypothesis that the transparency of compound heads is on average higher than of modifiers, and that more compound types are semantically transparent with respect to their head than with respect to their modifier. We observe that this relation is better captured via the as-constituent meanings than via the free-word meanings of the constituents. Finally, we examine whether semantic shift is predicted by experience with the constituent, as operationalized via family size: Libben (2014) assumes morphological transcendence to be the result of repeated experience with constituents in their respective roles, which can result in semantic shift; thus, semantic shift could be expected to be stronger or more likely for constituents with large family sizes. However, we observe the opposite pattern, which characterizes semantic shift as a rather idiosyncratic phenomenon (i.e., with a large variance between constituents) that is nevertheless captured by our model architecture. This opens a system-based rather than constituent-specific view on morphological transcendence.

References

Libben, G. (2014). The nature of compounds: A psychocentric perspective. *Cognitive Neuropsychology*, 31 , 8–25.

Libben, G., Goral, M., & Baayen, R. H. (2018). What does constituent priming mean in the investigation of compound processing? *The Mental Lexicon*, 13 , 269–284.

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as Abstract Operation in Semantic Space: A data-driven, large-scale model for relational effects in the processing of novel compounds. *Cognition*, 166 , 207–224.

Smolka, E., & Libben, G. (2017). 'Can you wash off the hogwash?'–semantic transparency of first and second constituents in the processing of German compounds. Language, *Cognition and Neuroscience*, 32 , 514–531.

Perceptual Salience of Affixes

Karla Orihuela, Hélène Giraudo
University of Toulouse, CNRS, France

Recently, Beyersmann, Ziegler & Grainger (2015) tested affix chunking in a letter search experiment. They found an advantage for suffixed nonwords (e.g., filmure) over pseudo-suffixed (e.g., filmire) but not for prefixed nonwords (i.e ., propoint is not > than cropoint). This asymmetry was interpreted as a reflection of different underlying processes for the recognition of suffixed and prefixed items. A chunking pre-lexical mechanism would operate on suffixed while prefixed would be represented holistically at the word level. As pointed out by Giraudo and Grainger (2003), functional and positional differences could reflect different processes. In the present study we performed two experiments where we tested morphemic salience (as defined according to Giraudo & Dal Maso, 2016) for real stems and positional effect of affixes by comparing prefixed and suffixed words (the use of real words as stimuli is crucial because they are entries with a real lexical representation). For both experiments, 30 native French speakers performed a letter search task in which participants were asked to find, as quickly as possible, a target letter.

In our first experiment, we selected words instead of nonwords because their use allows us to create a pseudo-affixed condition. Hence formal effects can be differentiated from morphological effects. Letter detection performances on real affixed words (e.g., injuste 'unfair'; tueur 'killer') were tested against pseudo-affixed (e.g., insecte 'insect'; fleur 'flower') and unrelated word controls. While results replicated the asymmetry found by Beyersmann et al. (2015), we found for suffixed words a genuine morphological effect differing from both pseudo-suffixed and the control condition.

On the other hand, for prefixed, only formal effects emerged without showing a significant difference between real prefixed and the unrelated control words. In order to dissociate functional from positional effects of prefixes and suffixes, we carried out a second experiment in which we only manipulated simple words. The target letter was present in a cluster either at the beginning or at the end, for example: 'E' in RE in chèvre 'goat' vs. requin 'shark'.

Results showed an advantage for the beginning over the ending letters, suggesting that prefix and suffix asymmetry is due to linguistic or functional factors rather than to the left to right reading direction (i.e., positional effect). Taken together, the results show morphemic salience in the suffix condition (possibly due to a functional effect). As for the prefixed words, word salience seems to guide the letter detection. This data has to be interpreted relative to the

masked priming data (Giraudo & Grainger, 2003) which also showed an asymmetry but in the opposite direction (morphological facilitation effect only for prefixed words). We interpreted our results within an interactive activation model integrating morphology as a central organizational principle (i.e, word representations are connected in morphological families and series through their paradigmatic relationships) as explained by the supra-lexical model which postulates that construction morphology is the main principle of organization of the mental lexicon (Giraudo & Dal Maso, 2016).

References

Beyersmann, E., Grainger, J., Casalis, S., & Ziegler, J. C. (2015). Effects of readingproficiency on embedded stem priming in primary school children. *Journal of Experimental Child Psychology*, 139 , 115–126

Giraudo, H., & Dal Maso, S. (2016). The salience of complex words and their structure: Which comes first? *Frontiers in Psychology*, 7:1778. doi: 10.3389/fpsyg.2016.01778

Giraudo, H., & Grainger, J. (2003). On the role of derivational affixes in recognizing complex words: Evidence from masked priming. In Baayen, R. and Schreuder, R. (Eds.), *Morphological Structure in Language Processing*, (pp. 211-234). Berlin: Mouton de Gruyter.

Morphology? Which units? Which mechanisms?

Ingo Plag
Heinrich-Heine-Universität Düsseldorf, Germany

There is a long-standing debate in morphology about what units are assumed to be the building blocks and which kinds of mechanism must be assumed to manipulate these building blocks. Focusing on derivational morphology, this paper will review some of the evidence for certain assumptions and develop a number of prerequisites that any psycho-computational model of derivational morphology must meet. We will look at discriminative models (NDL, LDL, e.g. Rescorla & Wagner 1972, Baayen et al. 2011, Baayen et al. 2018) and analogical models (e.g. Skousen 1989, Arndt-Lappe 2014) in more detail to illustrate some of the problems that arise from the said prerequisites.

References

Arndt-Lappe, S. 2014. Analogy in suffix rivalry: the case of English -ity and -ness. *English Language and Linguistics* 18(03). 497–548.

Baayen, R. H., P. Milin, D. Filipović Durdević, P. Hendrix & M. Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118. 438–482.

Baayen, R. H., Chuang, Y. Y., & Blevins, J. P. 2018. Inflectional morphology with linear mappings. *The Mental Lexicon*, 13(2), 230-268.

Rescorla, R. A. & A. R. Wagner. 1972. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Abraham H. Black & William F. Prokasy (eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts.

Skousen, R. 1989. *Analogical Modeling of Language*. Dordrecht: Kluwer.University

Collocational frequency and context effects in idiom processing in advanced L2 speakers

Evelyn Milburn, Mila Vulchanova, & Valentin Vulchanov
Norwegian University of Science and Technology

Multi-word chunks have attracted attention recently following suggestions that they are acquired chunk-wise by children in the first language, while adults learning a second language might use strategies other than chunking (McCauley & Christiansen, 2017). Idioms are multiword phrases in which the literal meanings of the component words do not (always) directly contribute to overall phrasal meaning, resulting in a figurative interpretation. Figurative meaning access is speeded both by idiom-internal characteristics, like higher collocational frequency (Cacciari & Corradini. 2015), and idiom-external characteristics, like supportive contexts (Vulchanova et al., 2019a). The tensions between these internal and external levels additionally make critical predictions for models of idiom processing.

We contrasted idiom-internal and -external influences on idiom interpretation by examining the relationship between the collocational frequency of idioms' component words and the context in which an idiom is embedded. In a visual world eye-tracking study, advanced non-native English speakers heard incomplete English phrases embedded within contexts that biased either literal or idiomatic continuations and saw images representing literal or figurative completions, or distractor images. In preliminary results, participants looked primarily at appropriate completions depending on the context, literal or figurative. However, they also showed late gazes to highly frequent, but contextually inappropriate continuations when context biased an idiomatic continuation. This suggests that advanced non-native speakers have trouble inhibiting gazes to highly frequent continuations when they are offered, particularly when processing idioms which may not be fully lexicalized. This finding is also consistent with the idea that L2 speakers are not using chunking strategies to the same degree as L1 learners, and are more likely to attend to properties of individual words. In further analyses, we are using Growth Curve Analysis (Mirman et al., 2008) to examine how gaze patterns change depending on the collocational frequencies of individual idioms as well as participants' English proficiency. We discuss the implications of these results for models of idiom processing, as well as for theories of second-language processing, and provide a parallel with other research where we find traces of literal processing in atypical development (Vulchanova et al., 2019b).

References

Cacciari, C. & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idioms processing. A reading-time study*. J. Cogn. Psychol*. 27: 797–811

McCauley, S. & Christiansen, M. (2017). Computational Investigations of Multiword Chunks in Language Learning (2017). *Topics in Cognitive Science* 9: 637–652

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences*. Journal of Memory and Language*, 59(4), 475–494. doi:10.1016/j.jml.2007.11.006

Vulchanova, M., Milburn, E., Vulchanov, V. & Baggio, G. (2019). Boon or Burden? The Role of Compositional Meaning in Figurative Language Processing and Acquisition. *Journal of Logic Language and Information* 28: 359.

Vulchanova, M., Chahboun, S., Galindo-Prieto, B. & Vulchanov, V. (2019). Gaze and Motor Traces of Language Processing: Evidence from Autism Spectrum Disorders in Comparison to Typical Controls. *Cognitive Neuropsychology*, online first.

Explaining dynamic morphological patterns in acquisition using Network Analysis

Elitzur Dattner, Ronit Levie, Orit Ashkenazi, Dorit Ravid
Tel Aviv University, Israel

The development of dynamic systems with growing complexity is at the heart of language acquisition. The dynamic nature of language development entails growing complexity of networks between forms and functions, as well as within functions and within forms. Network Analysis has been used to explain such relations especially in the realm of semantic networks, analyzing their structure and development (e.g., Beckage, Smith, & Hills, 2011; Bell et al., 2017; Steyvers, & Tenenbaum, 2004). The present paper proposes a novel methodology to account for emerging patterns of use in corpora by analyzing morphological relations (i.e., form-form relations) as networks. Specifically, we account for the relations between the Semitic constructs of roots and verb patterns (*binyan-im* 'building-s'), the morphological building blocks of Hebrew verbs (Ashkenazi, Ravid, & Gillis, 2016; Berman, 1993), asking whether these relations can be captured using network analysis. We analyze several new Hebrew corpora of input to young children and
children's own output in dyadic and peer interactions: Child speech between the ages of 1;8 to 2;2, peer talk of six age groups (2;0-2;6, 2;6-3;0, 3;0-4;0, 4;0-5;0, 5;0-6;0, 7;0-8;0), child directed speech to infants (6 months, 9 months, 12 months), and to toddlers (1;8-2;2), and storybooks targeting young children. Using network analyses of the relations between roots and patterns in each corpus, we reveal emerging patterns of links, manifested as root-based and pattern-based derivational families (Levie et al., 2019). We show that the development of the Hebrew verb category can be captured by the measures of (i) network mean degree, as representing the complexity of these patterns, and (ii) network density, as representing growth potential. Mean degree is the mean number of links for each node in the network. Thus, complexity simultaneously quantifies the number of roots shared by each pattern, and the number of patterns to which a single root is linked. Density is the proportion of observed links relative to the maximum number of possible links, quantifying the growth potential of the network: the potential to form new links between roots and patterns, either in future development, or in later parts of current discourse. This potential is low if the network is exhausted (i.e., high density value), since a relatively low number of non-existed links can be made. On the other hand, the growth potential is high when the system has more room to grow (i.e., low density value). We found that

networks become more complex with age, and density values decline with age, in both child and child directed speech, suggesting that these measures add another facet of quantifying language development.

References

Ashkenazi, O., Ravid, D., & Gillis, S. (2016). Breaking into the Hebrew verb system: A learning problem. *First Language*, 36(5), 505-524.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS One*, 6, e19348.

Bell, M., Perera, S., Piraveenan, M., Bliemer, M., Latty, T., & Reid, C. (2017). Network growth models: A behavioral basis for attachment proportional to fitness. *Scientific Reports* (Nature Publisher Group), 7, 42431.

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

Levie, R., Dattner, E., Zwilling, R., Rosenstein, H., Eitan Stanzas, S., & Ravid, D. (2019). Complexity and density of Hebrew verbs in preschool peer talk: The effect of socio-economic background. *The Mental Lexicon*, 14(2), 237–273.

Steyvers, M., & Tenenbaum J.B. (2004). The large scale structure of semantic - networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1), 41-78.

Stratification effects without morphological strata, syllable counting effects without actual counts, and what's in a trigram? A simulation study of English stress with Naive Discriminative Learning

Sabine Arndt-Lappe[1], Robin Schrecklinger[1], Fabian Tomaschek[2]

[1]) Trier University Germany, [2]) University of Tübingen, Germany

This paper uses a series of simulation studies with Naive Discriminative Learning (' NDL', Baayen et al. 2011) to investigate how two widely-shared theoretical assumptions about the nature of the English lexicon, which have been based prominently on regularities in stress assignment, are emulated in a model like NDL, in which these assumptions have no place. The first is that there are different morphological strata in the English lexicon, which are crucially distinguished by stress assignment. Stress in derivatives with 'stress-preserving affixes' falls on the same syllable as in its base (compare e.g. háppi-ness < háppy); stress in derivatives with 'stress-shifting affixes' may fall on a different syllable (com pare e.g. productív-ity < prodúctive). The other assumption is that grammatical stress algorithms make crucial reference to syllable counting from the word edge, i.e. to abstract levels of phonological representation.

In the present study, we use NDL as a classifier to predict stress position in 38.731 English words from the CELEX Lexical Database (Baayen & Piepenbrock 1995). NDL is a two-layer network that is trained to discriminate discrete outputs on the basis on discrete inputs, using the Rescorla-Wagner learning rule (Rescorla & Wagner 1972) which is a formalization of Discriminative Learning (Ramscar et al. 2010, 2013), a learning theory according to which speakers learn to discriminate the world by means of (visual or acoustic) cues.

We trained three NDL networks to discriminate stress positions on the basis of bigram and trigram letter cues for all words from the CELEX lexicon for English. The three networks differ in how information about the stress position was operationalised as outcomes in the classification task: 1) stress counted from the right, which is in line with traditional phonological perspectives according to which English stress assignment is located in a stress window at the right edge of the word (e.g. Hayes 1982 et seq., Pater 2000. Number of outcomes = 7); 2) stress counted from the left, which counts the position of the stress from the left edge of the word (N outcomes = 8); 3) stressed vowel, which allocates stress to a specific orthographic vowel in the word (N outcomes = 66).

Accuracy of prediction ranges between 59.4% and 85.5%, depending on the cue and outcome structure used; models are robust under cross-validation. The results thus provide general evidence that English orthography encodes stress position, both in simplex and in complex words. Furthermore, accuracy is higher when trigram cues are used than when bigram cues are used (ranging between 74.7% and 84.5% depending on the outcome formalisation used).

Accuracy is lowest for stress counted from the right and highest for stressed vowel. The best model (84.5% accuracy) uses trigrams as cues and vowel letters as outcomes. Contrary to pertinent claims in phonological theory, hence, stress assignment does not seem to require access to syllable count and syllable structure directly. We then studied how and on the basis of what information NDL is able to emulate stratification effects. In a regression analysis, we compared the cue-outcome weight structure of words with stress-shifting suffixes and of those with stress-preserving suffixes in our trigram-based model. We find that the two morphological categories differ significantly in how weights are distributed among trigrams that make up pertinent derived words. Trigrams representing stress-preserving suffixes have significantly lower weights than trigrams representing stress-shifting suffixes. Thus, contrary to pertinent claims in the traditional literature, the distinction between stress-preserving and stress-shifting suffixes is an emergent property of the system rather than an a priori specification in the input. Theoretical implications will be discussed.

References

Baayen, R. H., Milin, P., Durdevic, D. F., Hendrix, P., and Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on Naïve Discriminative Learning. *Psychological Review*, 118(3):438–481.

Baayen, R. H., Piepenbrock, R., and Guilkers, L. (1995). *The CELEX Lexical Database* (CDROM). Philadelphia: Linguistic Data Consortium.

Hayes, B. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13(2). 227–276.

Pater, J. 2000. Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* 17(237-274).

Rescorla, R. and Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Black, A. H. and Prokasy, W. (eds), *Classical conditioning II: Current research and theory*, 64–69. Appleton Century Crofts, New York.

Ramscar, M., Dye, M., and McCauley, S. (2013). Error and expectation in language

learning: The curious absence of ‘mouses’ in adult speech. *Language*, 89(4):760–793.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., and Thorpe, K. (2010). The effects of featurelabel-order and their implications for symbolic learning. *Cognitive Science*, 34(6):909–957.

Processing morphology: the view from language disorders

Christina Manouilidou
University of Ljubljana, Slovenia

The general aim of research in acquired language disorders is to understand how brain damage affects language either by focusing on specific issues or by looking at the big picture. The questions that are often asked are: a) To what extent is linguistic knowledge compromised after brain damage? b) What aspects of language are affected? c) How do the various types of disorders affect linguistic knowledge depending on their pathology? Answers to the above are primarily used to better describe the condition under investigation. However, if one looks at the other side of the coin, studies on language disorders often inform linguistic theory by providing independent evidence about a linguistic phenomenon and they offer an alternative way of looking at language processing and representation. The present talk aims to highlight this aspect by focusing on morphological processing in Primary Progressive Aphasia (PPA), a neurodegenerative condition which primarily affects word knowledge.

Psycholinguistic research on the processing of morphologically complex words supported by neuroimaging data has revealed that the human processor accesses morphological information immediately when confronted with a complex lexical item by decomposing it into its constituents (Rastle & Davis, 2008 for a review of behavioural studies; Fruchter et al., 2013 for neuroimaging data). In order to avoid the erroneous (in some cases) interpretation of initial decomposition (e.g. corn+er), validation processes take place during the last stage of lexical access where speakers engage their semantic knowledge in order to resolve incompatible information from stem and suffix. This has been described either as licensing (Burani et al., 1999) or as recombination (e.g. Fruchter & Marantz, 2015), and it refers to the stage where speakers try to make sense of the decomposed parts of the word.

Subsequent psycholinguistic research has also shown that stem + affix combinations are not only evaluated at semantic grounds but also syntactic information is taken into account. For instance, in a behavioral lexical decision experiment, Manouilidou & Stockall (2014), found that pseudowords which violated syntactic specifications of the stem + affix combination (e.g. rehappy) were rejected faster and more accurately than pseudowords in which the same affix was attached to a verb stem of the wrong semantic type (e.g. redance). More recent experiments investigated the neuropsychological correlates of this by

using Magnetoencephalography (Neophytou et al., 2018; Stockall, 2019) and revealed a consistent pattern of syntactic wellformedness versus semantic wellformedness, suggesting a multi-staged evaluation which highlights the contribution of syntactic and semantic information in building complex lexical items.

Within this framework, I will discuss how patterns of morphological loss in PPA come into play and inform the above processing model. I will focus on recent work about the processing of pseudowords by English-speaking populations with PPA (Manouilidou et al, under review) and on compound processing by Greek-speaking populations with PPA (Kordouli et al, 2018; under review) and I will show that indeed different kinds of information are evaluated when putting together a complex word attempting to further disentangle the contribution of the underlying cognitive processes operating when confronted with morphological processing. PPA is a neurodegenerative condition with three distinct variants, agrammatic, semantic and logopenic. Even though the landmark of the condition is impaired word knowledge, each variant has its own particularities as to how it manifests itself. That is, while agrammatic PPA is characterized by impairments of grammar (syntax and morphology), the semantic variant is mostly known for difficulties in processing lexical–semantic information (i.e., word meaning) and the logopenic variant for intermittent word-finding hesitations, impaired phonological memory and problems with repetition.

Specifically, combined results from accuracy and reaction times reveal interesting dissociations between groups in the way they process pseudowords violating (a) syntactic, e.g. rehappy and (b) semantic constraints, e.g. redance. Data suggests that there is a three-way approach, with PPA-L being the most successful, PPA-S the least successful and with agrammatic group (PPA-G) lying in the middle. Important correlations with the grammatical abilities of participants and accuracy rates allow us to identify the source for their performance and explain the different strategies used by the participants. That is, the agrammatic group appears to process both types of pseudowords simply on semantic grounds, where they are both unacceptable, but being unable to tell them apart. This performance correlates perfectly with their performance on grammatical tasks. PPA-S, on the other hand, appear to simply judge the two types of pseudo-words at a coarse structural level, being unable to process finer semantic information.

Likewise, Kordouli et al (2018; under review) examined production and comprehension of compound words, e.g. truckdriver in the three distinct PPA variants. Results showed that patients' performance was significantly worse than controls taking into account both accuracy of responses and reaction times. Qualitative analysis of patients' erroneous responses indicated a discrepancy between PPA-S and PPA-G. Specifically, PPA-S demonstrated retained knowledge of the compound structure, while PPA-G not; similarly, distinct production errors with a lot of neologisms for PPA-S but not for PPA-G suggest that problems in

compound retrieval are strongly related to the underlying deficits in each PPA variant, i.e., structural in PPA-G and semantic in PPA-S.

Combined results from the above studies provide independent evidence for the existence of various types of information that needs to be evaluated when confronted with a complex lexical item and inform the current models of lexical processing providing a view from language disorders.

References

Burani, C., Dovetto, F.M., Spuntarelli, A. & A.M. Thornton. (1999). Morpholexical access and naming: the semantic interpretability of new root–suffix combinations. *Brain and Language* 68(1/2): 333-339.

Fruchter, J., & A. Marantz. (2015). Decomposition, lookup, and recombination: MEG evidence for the Full Decomposition model of complex visual word recognition. *Brain and language* 143: 81-96.

Fruchter J., Stockall, L., & A. Marantz. (2013). MEG masked priming evidence for form-based decomposition of irregular verbs. *Frontiers in Human Neuroscience* 22(7): 798.

Kordouli, K., Manouilidou, C., Stavrakaki, S., Mamouli, D. & P. Ioannidis. (2018). Compound production in agrammatism: evidence from stroke-induced and Primary Progressive Aphasia. *Journal of Neurolinguistics* 47: 71-90.

Manouilidou, C. & L. Stockall. (2014). Teasing apart syntactic category vs. argument structure information in deverbal word formation: A comparative psycholinguistic study. *Italian Journal of Linguistics*, 26 71-98.

Manouilidou, C., Nerantzini, M., Chiappetta, B., Thompson, CK & MM Mesulam (under review) Processing Complex pseudo-words in Primary Progressive Aphasia and Agrammatic Aphasia.

Neophytou, K., Manouilidou, C., Stockall, L., & A. Marantz. (2018). Syntactic and semantic restrictions on morphological recomposition: MEG evidence from Greek. *Brain & Language*, 183: 11-20.

Rastle K., & M.H. Davis. (2008). Morphological decomposition based on the analysis of orthography. *Language and Cognitive Processes* 23: 942-971.

Stockall, L., Manouilidou, C., Gwilliams, L., Neophytou, K., & Marantz, A. (2019). Prefix Stripping Re-Re-Revisited: MEG investigations of Morphological Decomposition and Recomposition. *Frontiers in Psychology*, 10, 1964.

Functional and semantic properties modulate information in inflectional features

Francesca Franzon[1], Chiara Zanini[2]

[1]) SISSA, Trieste Italy, [2]) University of Zurich

Introduction. Grammatical information conveyed by the inflectional features plays a pivotal role at the sentence level in building up head-modifier agreement dependencies. In sentence processing, agreement would favor word prediction by restricting the possible alternatives, thus reducing cognitive effort (Dye et al. 2017; Wicha et al. 2004). In this sense, higher communicative efficiency would come from a maximally discriminative system, that could subdivide the lexicon so that ambiguity is reduced. In such system, all inflectional features would have the same probability of occurrence, which would correspond to the maximal entropy of the system.

However, inflectional features can also convey salient semantic properties of references such as numerousness in the category of Number or sex in the category of Gender (Corbett, 1991; 2000). We will discuss how the distributions of inflectional features can comply on the one hand with general coding and processing principles, and on the other hand with the encoding of semantic and referential information.

This study has been conducted on Italian nouns, which are mandatorily inflected for Number (singular and plural) and Gender (masculine and feminine). Number usually encodes numerousness, but it can underdetermined, surfacing in a 'default' value for functional purposes, which is usually singular as in the case of mass references (e.g. poco latte 'some milk'). Gender does not encode semantic features in most of the Italian lexicon; however, it can mark the sex of the referent in animate nouns, leading to systematic oppositions, as in gatto (masculine singular, male cat), gatta (feminine singular, female cat), gatti (masculine plural, male cats), gatte (feminine plural, female cats). Also Gender in animate nouns can be underdetermined: masculine plurals can include the reference to female entities, and masculine singulars can denote entities without strictly specifiying they are male.

Methods. We measured the distribution of nouns across the inflectional features. We compared the information of the obtained distributions to a maximally informative distribution for 2x2 discrete values (H=2).

A total of 210,325,942 tokens of 22,638 noun types were obtained by merging

Morph-It!, a morphologically tagged lexicon (Zanchetta & Baroni 2005), with ItWaC, a web-based 1.9 billion token corpus (Baroni et al. 2009). We excluded forms relatable to more than one inflectional feature (e.g. cameriere, masculine singular 'waiter' or feminine plural 'waitresses'). The measures were taken in the whole set of nouns and in a subset of 90 nouns denoting animate referents (4,404,872 tokens), showing the full inflection for Gender and Number mentioned above. Moreover, vectors were created collecting the lexical words occurring in a 10-word window around the animate nouns in ItWaC. The entropy of such vectors (Context Entropy) was measured, in order to assess the variability of context in which nouns occur, as an index of semantic specificity.

Results and discussion. Noun types are quite evenly distributed across the inflectional features, with 20% feminine singulars; 20% feminine plurals; 30% masculine singulars; 30% masculine plurals. The almost maximal entropy of type distribution (H=1.974) suggests that the nominal lexicon is potentially arranged in a way that allows to reduce uncertainty in sentence parsing.

Nevertheless, when considering token frequency, singular nouns increase (feminine singulars 34%; feminine plurals 13%; masculine singulars 35.6%; masculine plural 17.4%). This can possibly be the signpost of a semantic effect. In fact, within Number features, the singular can be used as a default value to express no reference about the numerosity while still providing a feature for functional operations. In turn, the increased information in the plurals corresponds to their less ambiguous encoding of a semantic interpretation about numerousness (Arcara et al. 2019). The distribution of tokens leads to a small decrease of information (H=1.881), with respect to the one potentially provided by types, displaying similarly distributed token frequencies across the inflectional features (Fig.1).

The selection criteria for the animate nouns forcibly lead to a maximal entropy in the types distributions (H=2). Noteworthy, when token frequency is computed, the information decreases to 1.502. Feminine singulars, and feminine plurals show lower token frequency when compare to masculine singular and plurals (Fig.2). The semantic counterpart of this is the use of default values for Gender underdetermination. As suggested by Context Entropy measures, feminine nouns occur in more specific contexts, whereas the less informative use of masculine nouns corresponds to the wider variability in the words they co-occur with (Fig.3).

In Italian, the inflectional features sustain a subdivision of the nominal lexicon, that potentially allows efficient discrimination and prediction processes. However, part of the space of probabilities of inflectional distributions seem to be also ruled by semantic specification and underdetermination constraints.

Fig. 1

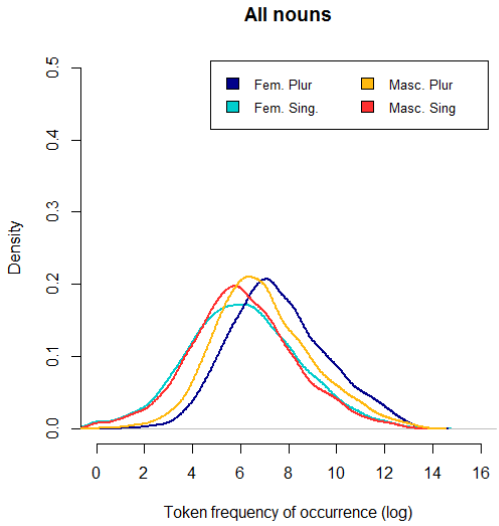


Fig. 2

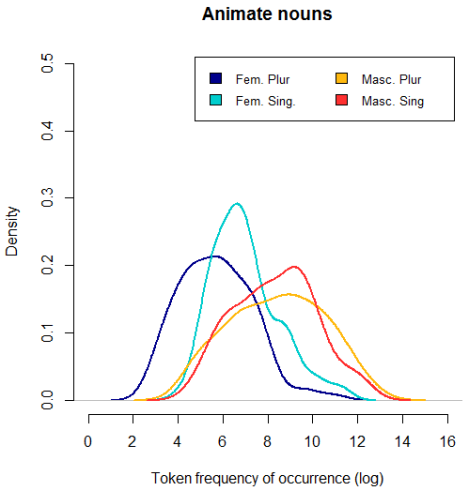
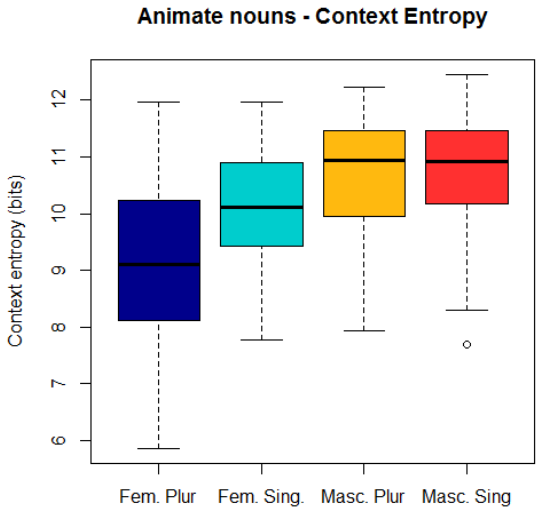


Fig. 3



References

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

Corbett, G. (2000). *Number*. Cambridge: Cambridge University Press.

Dye, M., Milin, P., Futrell, R., & Ramscar, M. (2017). A functional theory of gender paradigms. In *Perspectives on Morphological Organization* (pp. 212-239). Brill.

Wicha, N. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of cognitive neuroscience*, 16(7), 1272-1288.

Zanchetta, E., & Baroni, M. (2005). Morph-it! A free corpus-based morphological resource for the Italian language. In *Proceedings Of Corpus Linguistics*, http://dev. sslmit. unibo. it/linguistics/morph-it. php.

Simulating phonological and semantic impairment of English tense inflection with Linear Discriminative Learning

Maria Heitmeier[1], Harald R. Baayen[2]

[1])University of Edinburgh, UK, [2]) University of Tuebingen, Germany

Studies of the production of English verbs in aphasia have reported dissociations between regular and irregular verbs (Ullman, 1997). It has been claimed agrammatic patients (anterior aphasia, Parkinson's disease) have problems with inflecting regulars while patients with memory loss (Alzheimer, posterior aphasia) have problems with irregulars. However, Joanisse et al (1999) noted that the patients with memory problems actually had semantic deficits and that the agrammatic patients suffered from phonological difficulties. They therefore proposed a connectionist model with banks of units for semantics, auditory input, and speech output, as well as multiple hidden layers. Depending on whether they damaged the production layer or the semantic layer, the model selectively had more problems with regular verbs or with irregular verbs. In order to get their model to produce the desired interaction of type of impairment by regularity, Joanisse et al. had to add Gaussian noise to the semantic represention, without adding noise to the phonological representations. In other words, in this study, the desired interaction was hand-crafted into the model architecture.

Joanisse et al. assumed that semantics is irrelevant for past-tense forms in English. However, Baayen and Moscoso del Prado Martin (2005) showed that English irregular verbs have higher semantic densities than regular verbs. In the present study, we show that when topographically coherent areas of phonological and semantic networks of a discriminative lexicon (Baayen et al., 2019) constructed for English are lesioned, the interaction of regularity by type of impairment follows straightforwardly. The reason that specifically irregular verbs are vulnerable to semantic lesioning can be shown to follow from discrimination learning and the distributional properties of English. Because of their greater semantic density, irregular verbs are more likely to have edges to other irregular verbs in the semantic graph. As a consequence, in the twodimensional graph layout of the semantic topological map, irregular verbs will be more likely to be close together. It follows that under lesioning of the map, irregular verbs are more likely to be affected jointly.

References

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. (2019a). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*.

Baayen, R. H. and Moscoso del Prado Martín, F. (2005). Semantic density and past-tense formation in three Germanic languages. Language, 81:666–698. Joanisse, M. F. and Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, 96(13):7592–7.

Ullman, M. T., Corkin, S., Coppola, M., Hickok, G., Growdon, J. H., Koroshetz, W. J., and Pinker, S. (1997). A neural dissociation within language: Evidence that the mental dictionary is part of declarative memory, and that grammatical rules are processed by the procedural system. *Journal of Cognitive Neuroscience*, 9(2):266–276.

Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning

Yu-Ying Chuang[1], Kaidi Lõo[2], James P. Blevins[3], Harald R. Baayen[1]

[1]) University of Tuebingen, Germany, [2]) University of Tartu, Estonia, [3]) George Mason University, USA

The declension system of Estonian nouns appears to be far from straightforward from the perspective of decompositional theories that seek to associate units of form with corresponding semantic operations. However, within the framework of Word and Paradigm (WP) morphology (Matthews, 1974; Blevins, 2016), the noun system emerges as highly regular and predictable, with most inflectional forms falling out straightforwardly from the analogical relations within and across paradigms (Blevins, 2008). Recently, a computational implementation of WP morphology was introduced, the model of the "discriminative lexicon" and the associated theory of Linear Discriminative Learning (LDL, Baayen et al. 2019).

In this theory, both words' forms and meanings are represented as numeric vectors. At the heart of the model are two networks, neither of which contains hidden layers. One network maps form vectors onto meaning vectors (comprehension), and the other network maps meaning vectors onto form vectors (production). The mappings defined by these networks are mathematically equivalent to multivariate multiple regression of meaning on form, and form on meaning. Importantly, without designing any formal system with stems, and exponents, and inflectional classes, computational implementations of the model achieve high accuracies for a range of languages, including English, Russian, Polish, Korean, Turkish, and Biblical Hebrew.

The declension system of Estonian nouns provides an an interesting testcase for the model, for several reasons. First, paradigms, which can contain as many as 28 inflectional forms (14 cases and 2 numbers), contain subsets of forms that are constructed from different stems. Second, a substantial number of different declension classes govern how exactly forms within a paradigm are constructed. Third, for many nouns, only a subset of paradigm cells is in actual use. Finally, some cells of a given noun can be occupied by multiple alternative forms.

The goal of our study is to clarify whether our theory of the "discriminative lexicon" is able to perform with high accuracy for a nominal system with this high level of complexity. Of specific interest to us is how model perfomance varies when paradigms are underspecified (not all forms are used) or overspecified (some cells have multiple occupants).

A dataset with 232 nouns was constructed. We first considered the situation of complete paradigms without alternative forms. All the 28 inflectional forms of each noun in the dataset were spelled out by a native Estonian speaker. In cases where more than one inflectional forms are possible for a given paradigm slot, the most natural-sounding and frequent form was selected. For the computational model, word forms were represented by triphones (i.e., sequences of three phones), and word meanings were represented by vectors obtained by summing the semantic vectors of the word's meaning and those of its inflectional functions. Similar to previous studies on other languages, both comprehension and production accuracies were high, attaining 99.2% and 99% respectively.

To construct a dataset with incomplete paradigms, we consulted the online Balanced Corpus of Estonian (www.cl.ut.ee/korpused/grammatikakorpus/). Out of the 6496 inflectional forms, 2708 are absent in the corpus and were removed from the dataset. Despite the empty cells in the paradigms, comprehension and production accuracies were still close to error-free (both over 99%). In addition, when presented with the remaining 2708 unseen forms, comprehension accuracy remained high (97%). For production, however, accuracy dropped substantially to 64.5%. Many of these errors result from novel phone transition: given that pertinent triphones occuring in the testset are not available in the training data, the model cannot generate the appropriate inflectional forms. This limitation of the model is to a considerable extent due to the limitations that come with the way our modeling experiment was set up. We only considered (a subset of) nouns, without taking into account the many other words in the language (verbs, adjectives) in which triphones occur that are in our simulation 'out-of-vocabulary'.

Interestingly, the model also produced novel forms that can actually be understood by native speakers of Estonian. With respect to over-represented paradigms, i.e., paradigms with multiple alternatives in specific paradigm cells, we added the 144 alternative forms to the dataset that we were able to retrieve from the Balanced Corpus of Estonian. For comprehension, accuracy was at 99.3%, and for production, at 98.1%.

Considered jointly, the results suggest that computationally, WP morphology provides a promising framework for understanding the inflectional morphology of Estonian. The challenge for the next step in this line of research is to broaden the coverage of the model, by substantially extending the number of nouns taken into consideration, and by including, in addition to nouns, also verbs and adjectives.

Reference

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and

lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity*, 1–39.

Blevins, J. P. (2008). Declension classes in Estonian. *Linguistica Uralica*, 44(4): 241–267.

Blevins, J. P. (2016). *Word and paradigm morphology*. Oxford University Press, Oxford.

Chuang, Y.-Y., Lõo, K., Blevins, J. P., and Baayen, R. H. (2019) Estonian case inflection made simple. A case study in Word and Paradigm morphology with Linear Discriminative Learning. *PsyArXiv*, 1-19.

Matthews, P. H. (1974). *Morphology. An Introduction to the Theory of Word Structure*. Cambridge University Press, Cambridge.

How systems of morphological contrasts contribute to the discriminative process of communication

Michael Ramscar
Eberhard Karls Universität Tübingen, Germany

Traditionally, theories of language have assumed an atomic model in which linguistic signals comprise discrete, minimal form elements associated with discrete, minimal elements of meaning. Since linguistic production has been seen to involve the composition of messages from an inventory of form elements, and linguistic comprehension the subsequent decomposition of these messages, studies of linguistic morphology have focused on attempting to identify and classify these elements, along with the lossless processes of composition and decomposition they support. Time has shown that this program raises more questions than answers, especially when it comes to the nature of form-meaning associations, and the inductive processes that are thought to underlie their acquisition.

While ideas about compositionality have dominated linguistic thinking, research on learning has found it to be a fundamentally discriminative process. Learning models assume competitive mechanisms that serve to tune a system of cues that model the environment by iteratively eliminating or reducing uncertainty, such that at both a behavioral and neuroscientific level learning has been shown to be best characterized in terms of a deductive process. Further, and notably, the picture of understanding that emerges from learning models has much in common with information theory, which assumes a deductive model in which a common code designed to maximize the discriminability of messages while minimizing the cost of signaling is used to iteratively eliminate a receiver's uncertainty about the message communicated by a sender.

However, although the parallels between natural languages and information theoretic models of communication have often been noted, some important differences between the two are typically ignored. Not the least of these is the fact that while information systems rely on the existence of a shared source code, it is clear that human communicative codes are not shared across the users of a language in anything like the same way, because the statistical properties of natural languages make it clear that no one ever learns a natural communicative code in its entirety. Moreover, whereas information theoretic models of communication assume that models of the underlying probabilities of events in a system are shared, the fact that humans learn from experience guarantees that

individuals' samples of any code will vary enormously, making it difficult to see how individual models can ever align in the way presupposed by information theory.

In this talk I will show how the distribution of forms in natural languages appears to square this apparent circle, enabling probability models to align even though individual samples vary enormously, and endowing the communicative process with systematicity even though codes are 'incomplete.'

I will first describe how when analyzed in relation to the contexts in which they are used, form distributions across a diverse set of languages are geometric, such that the power law distributions typically thought to characterize lexical distributions only emerge when empirical distributions are aggregated. Because geometric distributions are memoryless, they are thus also far more resistant the effects of sample variance as compared to other distributions, suggesting a solution the first problem.

With regard to the apparent incompleteness of human codes, it has been shown empirically that many morphological forms are systematically organized into neighborhoods. These neighborhoods suggest one way in which human communicative codes might manage to achieve a level of systematicity that is sufficient for communicative purposes, because they allow for the scope of possible variations to be realized by inference: regular paradigms enable language users to generate forms that have not explicitly been experienced because they are implicitly coded in the distribution of form/semantic associations in the system, thereby endowing codes with systematically even in cases where experiential gaps can inevitably be expected to occur. By contrast, because irregular forms are both frequent and well discriminated, they serve to emphasize important communicative contrasts that are less saliently marked regular patterns, making code more efficient and ensuring transmission across generations.

From this perspective, the coexistence of regular and irregular patterns can be understood in terms of a functional trade-off between two opposing pressures on communication: predictability and discriminability. I will illustrate these points – and the very different picture of communication that emerges from the discriminative perspective – in relation to the oblique genitive construction. This construction can lay claim to be the most productive in the language, and is an aspect of English syntax that few native speakers master with confidence. I will show how successful transmission of it seems to rely on an interaction between its argument structure and the properties of the irregular and regular morphological forms contained within it.

Grounding transparency in vision: image-based distributional models and the perceptual side of compound-word comprehension

Marco Marelli, Marco Petilli, Fritz Günther

University of Milano-Bicocca, Italy

Semantic transparency, namely the degree to which the meaning of a compound word can be inferred from its parts, has been a central topic in the investigation of compound processing (Libben, 1998). Recent results suggest that its impact may be better understood when framed in compositional terms (Günther & Marelli, 2018), in line with the conceptual combination literature (Ji et al., 2011). However, conceptual combination is believed to be grounded on sensorimotor aspects, that is, compositional processes should not build on linguistic information only, but also rely on perceptual and bodily experiences (Lynott & Connell, 2010). The present study investigates this hypothesis by evaluating to what extent image-based transparency estimates can explain behavioral data in a series of compound-processing tasks.

In order to obtain a measure of "perceptual compound transparency", we exploited convolutional neural-network systems from computer vision (Vedaldi & Lenc, 2015), trained on image databases, to induce perceptual vector representations for 388 compounds and their corresponding constituents. These vector representations were then used to train a compositional distributional model (the CAOSS model; Marelli et al., 2017) that can produce representations for compounds on the basis of any pair of elements. Through this model we were hence able to compute estimates of perceptual compound transparency, capturing to what extent the visual representation of a compound can be obtained by the visual features of its constituents (e.g., to what extent a swordfish is, visually, a combination of a sword and a fish). Perceptual compound transparency for 736 compounds was then evaluated against response latencies in three behavioral tasks (lexical decision, word naming, and timed sensibility judgment).

Perceptual compound transparency was found to have a consistent effect across tasks: in all tasks, response latencies were found to be shorter for compounds which are perceptually more transparent. This effect is independent from the impact of linguistic predictors, including semantic transparency as estimated through the CAOSS model trained on text-corpus data. Indeed, these

latter metrics have effects that are task-specific and less robust than the ones of their visual counterparts.

These results confirm that word comprehension entails activations of the perceptual features of the lexically-denoted objects. This does not only apply to the word itself, but also to its morphemic units. Perceptual representations of compound constituents seem to undergo a compositional process that is similar to the one characterizing their language-grounded counterparts. More generally, the present results indicate that perceptual information plays a crucial role in characterizing semantic transparency from a cognitive perspective. Models providing an approximation of the speakers' perceptual experience can today inform such characterization in a previously unparalleled manner.

References

Günther, F., & Marelli, M. (2018). Enter sandman: Compound processing and semantic transparency in a compositional perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, 65(4), 406-430.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and language*, 61(1), 30-44.

Lynott, D., & Connell, L. (2010). Embodied conceptual combination. *Frontiers in Psychology*, 1, 212.

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as abstract operation in semantic space: Investigating relational effects through a large-scale, data-driven computational model. *Cognition*, 166, 207-224.

Vedaldi, A., & Lenc, K. (2015, October). Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 689-692). ACM.

Wide learning of the comprehension of morphologically complex words: from audio signal to semantics

Elnaz Shafaei-Bajestan and Harald R. Baayen
Quantitative Linguistics, Eberhard Karls University of Tübingen
{elnaz.shafaei-bajestan,harald.baayen}@uni-tuebingen.de

This study makes use of Linear Discriminative Learning (LDL; Baayen et al., 2019) to model computationally auditory comprehension of English inflected words. Unlike previous computational models such as TRACE (McClelland and Elman, 1986), our model takes real speech as input. Unlike in the study of Magnuson et al. (2018), who used synthesized speech, we are studying the performance of our model for spontaneous spoken American English, as sampled in the UCLA Library Broadcast NewsScape corpus of television news broadcasts.

In our model, comprehension starts with the extraction from the audio signal of discrete auditory features, namely, the Frequency Band Summary Features (FBSFs) introduced by Arnold et al. (2017). These features register patterns of change at the 21 frequency bands that characterize the tonotopic structure of the cochlea's basilar membrane. These patterns of change are extracted for stretches of speech bounded by minima in the Hilbert envelope of the speech signal's amplitude. Following Arnold et al. (2017) and Shafaei-Bajestan and Baayen (2018), we extracted FBSFs for the speech signals of individual word tokens sliced out of the speech stream. The number of different FBSFs for a word typically ranges between 21 and 84. From the features detected for a word, a form vector is constructed with 1s for those features that are present in the word and 0s for those features that are not realized in the word.

A word's form vector is mapped onto a corresponding semantic vector using a simple two-layer network without hidden layers. Mathematically, this is equivalent to the use of multivariate multiple regression. Semantic vectors (a.k.a. word embeddings) are constructed using an implementation from distributional semantics, introduced in Baayen et al. (2019). This implementation derives from a corpus, pre-processed with the treetagger software (Schmid, 1999), semantic vectors for both content words and inflectional functions. The inflectional functions we consider for this study are plural for nouns, past, perfective, and continuous for verbs, and comparative and superlative for adjectives. The semantic vectors for inflected words are constructed by summing the vectors of the base and the inflectional function.

As a first step, we trained and tested the model on only monomorphemic words, using 131,372 audio tokens representing 4,741types. The networkrecognized 34% of theseaudiotokenscorrectly. When we consider the number of targeted semantic vectors among the top 5 and top 10 best-supported vectors, LDL accuracy reaches 55% and 62%. The model's ability to generalize to unseen test data is at 8%, averaged over 10-fold cross-validation runs.

As a next step, we trained and tested the model on both monomorphemic and complex words using 153,644 audio tokens of 9,606 unique surface forms from the NewsScape database.

inflectional function	proportion in data (%)	true recognitions overall (%)	true recognitions per category (%)
Uninflected	80.0	25.5	31.9
Plural	6.8	4.5	66.4
Past	3.9	1.7	44.6
Continuous	2.9	1.6	54.7
3rd person	2.8	1.3	46.4
Perfective	2.4	1.1	43.5
Comparative	0.8	0.3	33.6
Superlative	0.3	0.1	41.3

Table 1: Proportions of 8 different inflectional functions across all data points are presented in the second column. Percentages of correct recognitions over all data points and within each inflectional function are presented in columns 3 and 4, respectively.

This model performs slightly better with an overall accuracy of 36%. Table 1 breaks down accuracy by morphological class. As can be seen, accuracy for inflected words is invariably higher than for uninflected words. Interestingly, model accuracies and the productivity of inflectional functions are correlated. We quantified the degree of productivity of an inflectional function using the Good-Turing estimate $P = n_1/N$ (see, e.g., Baayen and Lieber, 1991, n_1 denotes the number of types with a given inflectional function that occur only once and N the total number of tokens with this inflectional function), which approximates the probability of encountering unseen types. Degree of productivity and model accuracy were found to enter into a strong positive correlation: $r = .77, p < .05$.

Mean cross-validated accuracy of the model over 10 folds is 6%. To place the performance of the model in the context of current automatic speech recognition (ASR) systems, a Mozilla DeepSpeech model (Hannun et al., 2014), pre-trained on more than 5000 hours of speech, was able to recognize 5.8% of the tokens. The system was not able to make use of its language model, as the tokens used in the study are excised out of running speech and presented to the ASR system without context. End-to-End CTC-based ASR systems such as DeepSpeech clearly rely heavily on language models to select between the many different alternative lexical candidates that are compatible with the acoustic input. The low accuracy

of state-of-the art ASR contrasts with human comprehension accuracy, which was found for English to range between 20% and 60% (Pickett and Pollack, 1963). The present LDL model is developed in the hope that by taking seriously the separation in the cochlea of acoustic information at different frequencies, headway can be made in better approximating human auditory comprehension. Our current, biologically inspired, model, when trained on acoustics of single English inflected words excerpted from real spontaneous speech, performs surprisingly well, given its simplicity. An important next step in our research programme is to extend our model to continuous speech.

References

Arnold, D., Tomaschek, F., Sering, K., Lopez, F., and Baayen, R. H. (2017). Words from spontaneous conversational speech can be recognized with human-like accuracy by an error-driven learning algorithm that discriminates between meanings straight from smart acoustic features, bypassing the phoneme as recognition unit. *PloS one*, 12(4):e0174623.

Baayen, H. and Lieber, R. (1991). Productivity and english derivation: A corpus-based study. *Linguistics*, 29(5):801–844.

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., and Blevins, J. P. (2019). The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de) composition but in linear discriminative learning. *Complexity*, 2019.

Hannun, A., Case, C., Casper, J., Catanzaro, B., Diamos, G., Elsen, E., Prenger, R., Satheesh, S., Sengupta, S., Coates, A., et al. (2014). Deep speech: Scaling up end-to-end speech recognition. *arXiv preprint arXiv:1412.5567*.

Magnuson, J., You, H., Nam, H., Allopenna, P., Brown, K., Escabi, M., Theodore, R. M., Luthra, S.,Li,M.Y.-C.,andRueckl,J.(2018).Earshot: Aminimal neural network model of incremental human speech recognition. *PsyArXiv*.

McClelland, J. L. and Elman, J. L. (1986). The trace model of speech perception. *Cognitive psychology*, 18(1):1–86.

Pickett, J. and Pollack, I. (1963). Intelligibility of excerpts from fluent speech: Effects of rate of utterance and duration of excerpt. *Language and speech*, 6(3):151–164.

Schmid, H. (1999). Improvements in part-of-speech tagging with an application to german. In *Natural language processing using very large corpora*, pages 13–25.

Springer.

Shafaei-Bajestan, E. and Baayen, R. H. (2018). Wide learning for auditory comprehension. In *Proceedings of the Interspeech 2018*, pages 966–970.

Morphological errors in typical and atypical reading development. Is it deep dyslexia or morphosyntactic competence?

Emanuele Casani

Università Ca' Foscari Venezia

Morphological and semantic errors in reading can be associated with a deficit in both the lexical and sublexical route of reading, resulting in deep Developmental Dyslexia (DD) (Stuart & Howard, 1995). Function words and morphologically complex words are particularly problematic: the former are generally replaced for visually similar concrete words, for other function words, or they are just omitted; the latter are simplified through omissions and substitutions of morphological affixes. Subjects with deep DD may present syntactic deficits. A deficit in the phonological output buffer (phonological-output-buffer phonological dyslexia) can also be responsible for errors in reading longer morphologically complex words, with omissions, substitutions, and transpositions of some phonemes. In Italian, some studies have highlighted the main effect of morphological information on reading fluency (Burani, 2010) and accuracy (Angelelli et al., 2014). However, the effect of morphosyntactic abilities on reading is not clear.

In this study, we analyze the relations between reading and (morpho)syntactic competence.

To this aim, 42 Italian children in primary school were tested on written decoding, syntactic comprehension, and syntactic production of direct object clitic pronouns. Fourteen children (age 7;5 – 10;09 (M = 9;9, SD = 0;11) had a diagnosis of general DD, and 28 (age 8;4 – 11;3 (M = 9;6; SD = 0;11)) were age-matched typically developing (TD) children. Diagnoses were made by the Italian public health system (ASL) or authorized private clinical centers. Reading accuracy and speed were tested through standardized texts calibrated to the students' age (Cornoldi & Colpo, 2011). Besides, an analysis of reading errors based on an adaptation of the coding scheme by Friedmann & Coltheart (2018) was performed. Syntactic comprehension was tested through a standardized picture-sentence matching task (Bisiacchi et al., 2005). The results of items focused on grammar and lexicon were separately assessed. Syntactic production was tested through a non-standardized elicitation task of direct object clitic pronouns (Arosio et al., 2014). Independent sample tests revealed that DD children are significantly less accurate and slower in reading than TD children. No significant

differences emerged for syntactic comprehension and production.

A generalized robust linear mixed model with the double level of group (DD and TD), syntactic comprehension and syntactic production as fixed effects, and age and 11 types of reading errors including morphological, semantic and phonological-output-buffer errors as random effects, showed that both syntactic comprehension of grammar-focused items and syntactic production of direct-object clitic pronouns significantly predict standard scores of reading accuracy. DD children confirm to be significantly slower and less accurate than TD children after controlling for the above-mentioned effects. Age has a negative predictive effect on the proportion of semantic errors, which might be compatible with the gradual acquisition of the orthographic step of reading (Frith, 1985). Semantic reading errors are predicted by morphological reading errors, which might be evidence of the presence of deep DD in the sample. Moreover, this confirms that morphological information affects lexical decoding. The group level is the only significant predictor of phonological-output-buffer errors, with a negative effect of TD on the error proportion, meaning that DD children show a specific difficulty in decoding longer and morphologically complex words compared to TD children.

Fisher's exact test showed that morphological reading errors are significantly associated with the reading of determiners, clitic pronouns, prepositions, and verbs. This suggests that the decoding of morphology of function words, phonologically weak words, and morphologically complex words turns to be particularly problematic. Significant rates of morphological reading errors consist of substitutions with other visually similar parts of speech, omissions and morphological changes.

The present results suggest that oral (morpho)syntactic skills can play a role in reading accuracy even in a shallow orthography like the Italian one for both TD and DD children. A good (morpho)syntactic command can help recognize, expect and decode function strings while reading, such as free and bound morphemes. It would then be advisable to assess the linguistic profile of children with DD to establish whether some reading errors are related to a weakness in morphosyntax. In this case, a morphosyntactic training aiming at recognizing function strings, which can be easily mistaken for their morphological complexity and/or overlooked for their phonological weakness, might support reading accuracy.

References

Angelelli, P., Marinelli C.V., & Burani, C. (2014). The effect of morphology on spelling and reading accuracy: A study on Italian children. *Frontiers in Psychology* 5, 1373.

Arosio, F., Branchini, C., Barbieri, L., & Guasti, M.T. (2014). Failure to produce

direct object clitic pronouns as a clinical marker of SLI in school-aged Italian speaking children. *Clinical Linguistics and Phonetics*, 28 (9), 639-663.

Bisiacchi, P.S., Cendron, M., Gugliotta, M., Tressoldi, P.E. & Vio, C. (2005). BVN 5-11. *Batteria di valutazione neuropsicologica per l'età evolutiva*. Trento: Erickson.

Burani, C. (2010). Word morphology enhances reading fluency in children with developmental dyslexia. *Lang. Lang*. 9, 177–198.

Cornoldi, C., & Colpo, G. (2011). *Prove di Lettura MT-2 per la Scuola Primaria*, Firenze: Giunti.

Friedmann, N., & Coltheart, M. (2018). Types of developmental dyslexia. In A. Bar-On, & D. Ravid (Eds.): *Handbook of communication disorders: Theoretical, empirical, and applied linguistic perspectives* (pp. 721-751). Berlin, Boston: De Gruyter Mouton.

Frith, U. (1985). Beneath the surface of surface dyslexia. In J.C. Marshall, M. Coltheart, & K. Patterson (Eds.), *Surface dyslexia and surface dysgraphia*. London: Routledge & Kegan Paul.

Stuart, M., & Howard, D. (1995). KJ: A developmental deep dyslexia. *Cognitive Neuropsychology*, 12 (7).

Minimal Parsimonious Chunking Of Written Language

Romain Brasslet, Davide Crepaldi
SISSA, Trieste Italy

Somewhat intuitively, scholars have been holding that words fundamentally mediate the mapping between form and meaning in human languages, but this view has been challenged on several fronts. Theoretical linguists have proposed theories of the human lexicon that dispense of the concept of word (Stump, 2019), while Computational Linguistics has started to build semantic models where meaning is mapped onto sub-word letter chunks (Bojanowski et al., 2016) and Cognitive Neuroscience has found that the brain is sensitive to processing units other than words (e.g., Amenta and Crepaldi, 2012). Obviously, moving to smaller units would require less storage capacity but more processing power, but it is far from clear what the best compromise between the two would be, and if there is such an optimal compromise in the first place. Indeed, Nowak and Krakauer (1999) have suggested that combining smaller, lower-level units devoid of meaning (e.g., letters) into larger, higher–level, meaningful units (e.g., morphemes, words) is necessary in order to overcome an informational bottleneck — this would be exactly what gives the human language its immense communicative power. In fact, all known languages seem to possess this feature, which was proposed as a linguistic universal already in the 60s ("double articulation", Hockett, 1963). From this point of view, however, that the higher-level units should be words is by no means a given—any combination of letters might in principle work.

Proposals have been made that using chunks of progressively larger size is exactly what becoming a proficient reader entails. Children move from letter-by-letter conversion into quick and automatic word identification; and proficient readers can regress at letter conversion after brain damage (e.g., Arguin et al., 2002). Also, Developmental Dyslexia has been interpreted as reading without word representations (e.g., Burani et al., 2008).

The idea that we progressively use larger and larger units to store information is widespread, and reaches far into domains as different as learning to read (the brain computations within the ventral stream, more generally), language evolution, and information theory. Yet, it's unclear whether any particular unit size would be privileged, and what kind of dynamics can describe this tension between storing a great amount of singletons and make little use of combination; and storing a small number of units, which we would then combine extensively.

Here we propose a formal framework to interpret this tension, describe the possible cognitive systems that may emerge from it, and quantify their informational content. Essentially, we try to find the minimal set of chunks with which we can parsimoniously decompose all the words in a language. This framework is based on one unique trade-off parameter between the size of the set of chunks and the number of chunks needed per word on average, which, we argue, can be interpreted psychologically to shed light on reading acquisition, dyslexia, and the progressive mastering of literacy more in general. From our algorithm, we extracted cohesive building blocks of words. Some of them are indeed morphemes, that is, directly interpretable (such as 'ed ', 'ing ' or ' re'); but some others are chunks devoid of any obvious meaning. We also assessed their productivity, i.e. how much they participate in the decomposition of words. We hypothesize that these productive blocks could actually be fundamental in our ability to decompose words. The results of this work generate predictions that can be tested experimentally at several different levels.

References

Amenta, S. and Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. Frontiers in Psychology, 3, 232.

Arguin, M., Fiset, S. and Bub, D. (2002). Sequential and parallel letter processing in letter-by-letter dyslexia. Cognitive Neuropsychology, 19, 535-555.

Bojanowski, P., Grave, E., Joulin, A. and Mikolov, T. (2016). Enriching word vectors with subword information. CoRR, 1607.04606.

Burani, C., Marcolini, S., De Luca, M., and Zoccolotti, P. (2008). Morpheme-based reading aloud: Evidence from dyslexic and skilled Italian readers. Cognition, 108, 243-262.

Hockett, C.F. (1963). The problem of universals in language. In Greenberg, J.H. (Ed.), Universals of Language. Cambridge, MA: MIT Press.

Nowak, M.A. and Krakauer, D.C. (1999). The evolution of language. Proceedings of the National Academy of Sciences, USA, 96, 8028-8033.

Stump, G. (2019). Theoretical issues in inflection. In Audring, J. and Masini, F. (Eds.), The Oxford Handbook of Morphological Theory. Oxford, UK: Oxford University Press.