Third NetWordS Workshop

Variation and Adaptation in Lexical Processing and Acquisition

Dubrovnik (Croatia) 19-20 September 2013

Convened by:
NetWordS Steering Committee

SCIENTIFIC REPORT

Claudia Marzi (ILC-CNR)
Ida Raffaelli (FFZG)

Organising Committee

Ida Raffaelli
Claudia Marzi
Vito Pirrelli

Local Organisers

Daniela Katunar
Barbara Kerovec
Matea Srebačić
Krešimir Šojat

Faculty of Humanities and Social Sciences, Zagreb Univeristy, Croatia

Istituto di Linguistica Comnputazionale – CNR, Italy

1) Summary

The 3rd NetWordS Workshop, held on the 19th and 20th of September 2013 in Dubrovnik, brought together 30 participants (scholars, Post-Docs, PhD students) from various European countries.

Nineteen speakers, experts of various scientific domain and with different theoretical inclinations, discussed cross-disciplinary perspectives on *Variation and Adaptation in Lexical Processing and Acquisition*, reflecting the interdisciplinarity and synergy fostered by NetWordS, the European Research Networking Programme on Word Structure.

With these objectives in mind, the workshop gathered PhD students and junior research fellows who carried out interdisciplinary research under the NetWordS granting scheme for 2013 and more senior scholars who are currently involved in European or national initiatives geared towards scientific goals of interest to the NetWordS programme.

Each contribution was planned to have a 45' duration including discussion involving the entire audience for the invited *grantees* and for the invited senior scholars.

2) Scientific content of the event

Recent emphasis on language knowledge as an emergent dynamic system has drawn considerable attention to the role of time in the way speakers acquire and use their own language. There are at least three levels on which time matters. At the processing level, the interaction between processing and memory constraints, and in particular between short-term and long-term memory issues, is understood to shape the way we recode and organise time-bound sequences of linguistic signals. On an ontogenetic scale, the age of acquisition of language input data, and the duration of exposure (in the case of multilingual contexts) are known to interact with issues of cognitive maturation and brain plasticity, yielding different outcomes as a function of different time intervals. In this connection, also the distribution of input data in a particular linguistic environment (both in terms of word type and token frequency) is bound to have an impact on rate and speed of acquisition and on overall knowledge organisation. Finally, all previously mentioned time-effects conspire to make the language system change through usage and acquisition in passing from one generation to the ensuing one.

The present workshop was intended to address some of these issues through the joint contribution of a number of invited leading scholars, and some younger research fellows who are currently carrying out joint interdisciplinary research under the 2013 NetWordS granting scheme.

VEDRAN GALETIĆ in his talk "*Towards capturing latent conceptual spaces"* presented the results of his short visit to professor Gärdenfors at Lund University. Main objective of his research is to enrich theoretical underpinnings to the semantic theory of conceptual spaces. He first outlined the conceptual spaces by defining the set of properties via which they are spanned - e.g. by geometrically structured dimensions such as temperature, brightness, height, weight - and their properties. Secondly, he presented a model for typicality quantification. The most challenging aspect of his work is to combine different approaches to semantic modelling, based on both experimental data (e.g. mental representations, that are perceptually grounded) and distributional data (e.g. use in language, which is applicable for all concepts).

KEVIN TANG presented the results of his short visit to the University of Ghent in his talk on "*The OpenLexicons Project - Development and Uses of Subtlex-Corpora for Investigating Sound Symbolism and Brazilian Portuguese"*. He illustrated his joint work with E. Keeulers and Pavel Mandera on the *Openlexicon*, that aims to collaboratively develop tools required for experimental linguistic research in those languages for which tools such as validated word frequencies, tests for lexical knowledge, pseudoword generators are not available. He documented how a Subtlex

corpus can be enriched beyond tabulating the token frequency of words, e.g. neighbourhood density, lemmatisation, POS tagging, grapheme to phoneme conversion, and N-gram versions of the corpus.

In his invited talk on "*The Sources of Non-Concatenative Morphology in Diachronic Change*" STEPHEN ANDERSON developed the argument that the existence of non-concatenative morphology demonstrates the inappropriateness of traditional models of morphology based on the classical morpheme. This evidence should be especially relevant to scholars involved in the development of computational and psychological models of knowledge of words, for whom it is important to understand that the analysis of non-concatenative formations is not just a marginal phenomenon relevant to choices among abstract formal models, but a fundamental point about the nature of the human language faculty in the domain of knowledge of words. The main point of this talk was to persuade that there is nothing deeply unnatural about non-concatenative morphological markers. They arise in perfectly expected ways in the evolution of individual languages. While Anderson acknowledges the non-concatenative markers of morphological categories that result from historical developments are less common than affixes, they are apparently the regular outcome of normal phonological change. Speakers show no reluctance to reanalyse opaque phonological changes as morphologically conditioned. Differences in shape between morphologically distinct forms, whether they consist in the addition of material (an affix) to one form or not, are perfectly valid realizations of morphological content. Hence, theories of morphology that purport to capture the nature of speakers' knowledge of word structure cannot be confined to purely affixal, concatenative models, including those based on the classical morpheme.

STELA MANOVA presented in her talk "*Suffix ordering in Italian and Russian: a cognitive approach*" a comparative, corpus-based study of the Italian and Russian suffix combinations in derivation, carried out with Luigi Talamo and Chiara Celata of the Scuola Normale Superiore in Pisa, where Manova spent her short visit. Main objective of her presentation has been to reveal the mechanism behind suffix order in a cognitive approach, for both suffix combination in Italian and Russian. She suggests that affix ordering is best analysed in terms of binary combinations of affixes, and that to understand the nature of suffix combinations it is not always necessary to relate them to lexical bases, but rather if the lexical-category specification of a suffix and suffix-particular semantics are considered, most of the combinations appear either fixed or predictable.

ALEXANDRA BAGASHEVA reported results obtained during her short visit at the University of Wien, where she worked on "*Semantically-conditioned two-suffix constructions in English and Bulgarian*" together with Dr. Manova and prof. Dressler. The focus of her study is on affix ordering both in English and Bulgarian, and in particular on affix combinability in the two languages that are typological different. By analysing lexical items from various sources, she established a significant

distinction in affix combinability, for both languages, which are associated with the semantics of Personhood and Object-hood, with no hint of argument structure. In this way, the analysis she proposed is informative at the level of basic conceptual categories such as Person and Object.

DANIELA KATUNAR in her "*Constructional effects on prepositional antonymy*" presented results obtained during her short visiting at the University of Copenhagen, where she analysed antonymy relations and tested methods for extracting functional units. She discussed results in the frame of theoretical models for their representation in the lexicon, by proposing a conceptually based model of antonymy relations of prepositions to be used in typological research, and a model for the relation between antonymy and polysemy of prepositional meanings. Danish data were used to elaborate and adjust the standard as devised for Croatian. The model she proposed uses antonymy relations between prepositions to define and separate different polysemous meanings, and therefore incorporates the two semantic relations (antonymy and polysemy) as complementary relations within the lexical structures of prepositions.

In GEERT BOOIJ's talk "*Language use and the architecture of grammar: a Construction Morphology perspective*" I-language (i.e. the system of knowledge which resides in the speakers' brains and which underlies their linguistic performance) and E-language (i.e. the language as encountered in the world) are understood as standing in a mutual relationship. Morphology must be usage-based in order to understand and explain the nature of lexical knowledge, in particular the knowledge and creation of complex words. The Construction Morphology model of a hierarchical lexicon with various degrees of schematicity can do justice to actual language use in the domain of word creation. The approach leads to important implications concerning the relationship between word-level units and above-word-level units (e.g. multi-word units and phrases). Since there are also productive phrasal lexical constructions, and word formation may be based on paradigmatic relationships with phrases, there is no sharp divide between lexicon and grammar. Second order schemas account for a multi-dimensional network of relationships between linguistic constructs, with Construction Morphology allowing for graceful integration of findings in the various subdomains of linguistics.

In the invited talk on "*Variation in Morphology*" ANGELA RALLI addressed the issue of language variation as the outcome of the dynamic relationship between change, which is often difficult to predict and mostly triggered by language-external factors such as ease of articulation, prestige pronunciation, language contact etc., on the one hand, and language-internal constraints, such as structure of the paradigmatic grid hosting an intervening change, on the other hand. Such a general claim is supported by a wealth of evidence from Modern Greek and Modern Greek dialects, covering inflection, derivation and compounding. To give but one example, variation in the selection of inflectional markers (heteroclisis) in two Modern Greek dialects is associated with patterns of stem

allomorphy. The most vulnerable cells to intrusion of suffixes from another declension class are exactly those that exhibit variation at the level of stem selection. This corroborates Stump's claim, according to which when two inflection classes are involved in heteroclisis, the 'intrusive' class is generally expected to occupy the more marked set of cells, and confirms Maiden's assertion that heteroclisis is morphologically conditioned and related to stem allomorphy.

In the invited talk "*Adaptation Effects in Lexical Processing*" CHRISTINA L. GAGNÉ and THOMAS L. SPALDING questioned the dominant assumption in experimental psycholinguistics that underlying representations and processing in the mental lexicon are not hugely altered during the experiment and that by averaging results across trials one can successfully filter out random effects and obtain more stable estimates of experimental outcomes. In fact, they argue, only by taking into account the inherently dynamic nature of the mental lexicon we can understand more about the influence of context on lexical processing and the incremental pressure exerted by past trials on future behaviour. This approach has both methodological and theoretical implications. Considering the adaptability of language and cognitive systems alters the scientific strategy because the experimental protocol need be able to evaluate changes occurring during the experiment. Besides considering what the system structure is, it is useful to consider what the system can do and how it does.

The nowadays traditional notion of language resources has considerably extended its scope and coverage over the last ten years, to include, besides digitally stored and accessible language data (dictionaries and text corpora), possibly encoded in a standard and uniform way, also tools for processing language data (e.g. standalone applications or environments) and web-based on-line accessible services. In his invited talk "*New language resources, tools and services for morphological processing*", MARKO TADIĆ illustrated two current networking initiatives in the domain of language resources, dedicated to build the technological foundations of a multilingual European information society: META-NET and Cesar. META-NET is aimed at paving the way to an ambitious joint international effort seeing Language Technology as one of the means for realising a united European digital market and information space. Cesar operates as part of the META-NET network of excellence to promote a coordinated effort in the development of Central and South-Eastern European Language resources and tools, and to raise awareness of the role of language resources amongst policy makers, media and the general public.

According to traditional wisdom in Linguistics, morphemes are represented in the mental lexicon at the phonological level, there being no phonetic difference between, to give an English example, different -*s* morphemes (plural, genitive, 3rd person singular etc.) or in the speech production of morphemic and non-morphemic /-s/. In his talk "*Against homophony: the acoustic of English {s} morphemes*", INGO PLAG questions this assumption to provide, through statistical (re)analysis of both old (Walsh and Parker's 1983) and new (Buckeye corpus) behavioural and experimental English data,

fresh evidence supporting the hypothesis that speakers rely on more detailed (phonetic) representations of morphologically complex lexical forms. This analysis has significant theoretical implications. In theoretical accounts of English phonology, both plural –s and 3rd person singular –s present nearly identical phonological realisations in phonologically-defined contexts. Current models of speech production do not postulate another level of form representation besides context-sensitive selection/adjustment of phonologically distinct allomorphs. Plag concluded that phonetic detail must have a place in the description of the formal aspects of the morphemes involved, thus questioning the empirical underpinnings of both theoretical models of the phonology-morphology interface such as Lexical Phonology and more recent models of speech production, where post-lexical phonology has no access to morphological structure.

HANNE RUUS in her presentation on "*Diachronic Variation in Danish Morphology"* analyses change and variation in Danish inflection morphology during two hundred years – from 1500 to 1700. Digitally accessible texts put together ballads and songs (approximately 1000), and prose texts on different subject fields, offer evidence of variation in gender, case, number and person. The Lexical Multi-Level Text representation framework shows findings and parallel forms. The proposed textbased investigation highlights proper changes, stable old forms, new gradually taking-over forms, concurrent old and new forms, and different morpho-syntactic systems.

PETAR MILIN outlined in his talk "*Morphology Acquisition with Naive Discriminative Learning"* theoretical approaches and processing models of language and morphology acquisition, by focussing on visual comprehension based on naïve discriminative learning. His analysis inspected experimental results and simulated predictions, which highlight language as a complex adaptive system based on both rich dynamics and optimality constraints. In this direction, information theory could be a fruitful tool in helping to understand constraints and how they emerge, and naïve discriminative learning model can represent a useful support for detailed linguistic and psychological analysis, far from requiring hard statistics on a hidden layer.

CLAUDIA MARZI in her talk "*On memory and computation: a reappraisal of German noun plural inflection"* presented one of the main objectives on which the bilateral Italian-Belgian action is focussing, namely linguistic and extra-linguistic factors involved in mono- and bi-lingual word recognition. She suggests a multi-factorial view of morphology adaptive processing, in presenting a usage-based perspective, by investigating a few properties of the German noun plural system and focusing on the dynamic relation between regularity, productivity and competition of inflection patterns through computer simulations (Temporal Self-Organising Maps - TSOMs) of type/token-frequency effects. In detail, she highlighted a few formal properties of the *–s* plural class, and concluded that although relatively infrequent, *-s* plurals seem to pattern in fairly regular sub-classes which suffer from no competition by members of other inflectional classes.

EMMANUEL KEULEERS and MARCELLO FERRO in the frame of the bilateral Italian-Belgian action in their "*A psycho-computational view on wordlikeness*" focus on the issue of word similarity and its representation. In line with recent psycholinguistic studies, measures of node activation in a TSOM connectionist model are used to test hypotheses about perceived levels of word-likeness. In compliance with some desiderata for a model of lexical similarity, namely context-driven, spatially-independent, alignment-free, language-independent, TSOMs offer a cognitively plausible model of lexical representation as a dynamic long-term store, based on words as sequence of symbols where a similarity measure is based on topological proximity and activity synchronization.

MILA VULCHANOVA and WOLFGANG DRESSLER, in the frame of an international collaboration for *"The LangEqual project"* (Early language development: Language input's impact on (un)equal language opportunities), outline the interdisciplinary large-scale European project proposal on the causal chain to educational and social inequality. Main objective is to focus on language-related and environmental mediator factors leading to academic success in the early school years, by investigating the input and output of preschool children in verbal interactions with their caregivers at home and in kindergarten and pre-school according to 2 dimensions: low vs. mid/high socio-economic status (SES), monolingual children acquiring the respective national language as their first language vs. bilingual children acquiring an immigrant language as their first and the respective national language as their second language. The expected results consist of (i) the creation of novel and enriched datasets for Germanic languages on the gap in the causal chain of language acquisition and its dependence on environmental (language-external) factors; (ii) ecologically valid and valuable sociological data on the daily routines and practices of families of different SES and different backgrounds (immigrant vs. non-immigrant) in five European welfare states; (iii) new interdisciplinary knowledge feeding into educational policies (updates in teacher-training programs for kindergartens and higher education); (iv) new knowledge for caregivers (parents, kindergarten teachers), and the public(institutions, educators, policy makers); (v) substantial contributions to public and specialists' knowledge about language acquisition.

3) Assessment of the results and impact of the event on the future direction of the field

Time and context appear to play a prominent role in the way speakers perceive, decode, acquire and use linguistic knowledge. Not only do speakers prove to be able to memorise and capitalise on phonetically detailed information concerning word forms in both perception and production, but their use of this information may change over time, as a function of strategic adaptation, repetition and frequency. Incremental learning and functional specialisation of linguistic structures appear to be key notions in this context. Redundancy in the mental lexicon is another key-factor. The co-existence of different levels of linguistic representations, rangin from fairly specific item-based and context-sensitive information, to more context-free general schemata appears to underpin both the capability of a speaker to selectively exploit this information depending on the task at hand, and the composite nature of the language system as a whole, where different subsystems, both old and new ones, can be in operation at the same point in time. Understanding the neurocognitive and developmental mechanisms accounting for such a high degree of opportunistic adaptivity is one of the most challenging intedisciplinary goals in language research for the years to come.

More specifically, and in line with the main goal of the Project, namely to focus on our comprehension of the basic mechanisms serving both language and cognition, as key to understanding the human strategies involved in learning and processing word structure, most workshop contributions were devoted to shedding light on understanding the architecture of the mental lexicon as a dynamic system. It is important to emphasise at this stage that most contributions promoted this view successfully as a result of interdisciplinary integration of different methodological approaches. The research community at large appears now to be more aware of the importance of sharing data, methods and protocols through the on-line interoperable infrastructures which are common and widely available in the Language engineering sector.

Another notable reason of the success of the event is ascribable to the joint 9th MMM conference. Most remarkably, some of the speakers of the MMM conference attended the whole NetWordS workshop – both as speakers and participants. Moreover, the most mature research work of short visiting applicants and bilateral actions were presented at the 3rd NetWordS workshop.

In line with the succesful event, a third call for grants (both short-visit grants and action supporting grants) will be launched for year 2014, to foster new partnerships meeting the challenges of 2014 European research and development funding programme (Horizon 2020), as well as promote training of young scientists and a new generation of truly interdisciplinary scholars.

4) PROGRAMME

Thursday, 19 September 2013

9:30   ***Welcome by convenors***

9:45   Vedran Galetić - University of Zagreb, Croatia
               Towards capturing latent conceptual spaces

10 :15   Kevin Tang - University College London, United Kingdom & Paweł Mandera - Ghent
               University, Belgium

               *The OpenLexicons Project - Development and Uses of Subtlex-Corpora for Investigating
               Sound Symbolism and Brazilian Portuguese*

10:45   ***coffee break***

11:15   Stephen Anderson – Yale University, USA
               The Sources of Non-Concatenative Morphology in Diachronic Change

12:00   Stela Manova - Austrian Academy of Sciences, Institute for Corpus Linguistics and Text
               Technologies – Vienna, Austria
               Suffix ordering in Italian and Russian: A cognitive approach

12:30   Aleksandra Bagasheva - Department of English and American Studies, Sofia University "St.
               Kliment Ohridski", Bulgaria
               Semantically-conditioned two-suffix constructions in English and Bulgarian

13:00   Daniela Katunar - University of Zagreb, Croatia
               Constructional effects on prepositional antonymy

13:30   ***lunch***

15:00   Geert Booij, Leiden University, The Netherlands
               Language use and the architecture of grammar: a Construction Morphology perspective

15:45   Angela Ralli – University of Patras, Greece
               Variation in Morphology

16:30   ***coffee break***

17:00   Christina L. Gagné, Thomas L. Spalding – University of Alberta, Edmonton, Canada
               Adaptation Effects in Lexical Processing

17:45   Marko Tadić – University of Zagreb, Croatia
               New language resources, tools and services for morphological processing

18:30   ***End of the day***

20:00   ***Dinner***

Friday, 20 September 2013

9:00   Ingo Plag - Heinrich-Heine-Universität Düsseldorf, Germany
               Against homophony: the acoustics of English {s} morphemes"

9:45   Hanne Ruus, University of Copenhagen, Denmark
               Diachronic Variation in Danish Morphology

10:30   Petar Milin – University of Tuebingen, Germany – University of Novi Sad, Serbia
               Morphology Acquisition with Naive Discriminative Learning

11:15	***coffee break***
11:45	Claudia Marzi - ILC-CNR, Pisa, Italy & Walter Daelemans, University of Antwerp, Belgium *On memory and computation: a reappraisal of German noun plural inflection*
12:15	Emmanuel Keuleers- Ghent University, Belgium & Marcello Ferro - ILC-CNR, Pisa, Italy *A psycho-computational view on wordlikeness*
12:45	Mila Vulchanova - Norwegian University of Science and Technology, Trondheim, Norway & Wolfgang Dressler – University of Vienna, Austria *The LangEqual project*
13.30	**End of Workshop**
	* * *
15.00	**3rd NetWordS Steering Committee Meeting**