

## First NetWordS Workshop on

# Understanding the Architecture of the Mental Lexicon: Integration of Existing Approaches

Pisa (Italy), 24-26 November 2011

Convened by: NetWordS Steering Committee

**Organizing Committee** 

Vito Pirrelli Claudia Marzi Marcello Ferro Claudia Caudai

**SCIENTIFIC REPORT** 

Claudia Marzi (ILC-CNR)





## 1) Summary

The 1<sup>st</sup> NetWordS Workshop, held on the 24<sup>th</sup>, 25<sup>th</sup> and 26<sup>th</sup> of November 2011 in the Research Area of the Italian National Research Council, brought together 37 participants (scholars, Post-Docs, PhD students) from various European countries.

Eighteen speakers, experts of various scientific domain and with different theoretical inclinations, discussed cross-disciplinary approaches to the *Understanding of the Architecture of Mental Lexicon*, reflecting the interdisciplinarity and synergy fostered by NetWordS, the European Research Networking Programme on Word Structure

The workshop was organised with the ambitious goal of paving the way towards a European interdisciplinary research agenda on the Mental Lexicon for the coming 10 years, with particular emphasis on the three main challenges that NetWordS is intended to address:

- Lexicon and Rules in the grammar
- Word knowledge and word use
- Words and meanings

Leading scholars, mostly connected through NetWordS, were invited to address three basic questions:

- What are, in the speaker's area of expertise, the most pressing open issues concerning the architecture of the Mental Lexicon?
- What and how can progress in other research areas contribute to addressing these issues?
- What can advancement in our understanding of these issues contribute to progress in other areas?

Each contribution was planned to have a 35' duration and to be followed by a 10' discussion involving the entire audience.





## 2) Scientific content of the event

Scientists involved in NetWordS are playing a leading role in the European scenario in:

- Theoretical issues in morphology and its interfaces
- Typological, variationist and historical aspects of word structure
- Cognitive issues in lexical architecture
- Short-term and long-term memory issues
- Neuro-physiological correlates of lexical organization and processing
- Psycho-linguistic evidence on lexical organization and processing
- Machine-learning approaches to morphology induction
- Psycho-computational models of the mental lexicon
- Distributional Semantics

Words are the basic building blocks of language productivity, establishing the most immediate connections between language and our conceptualisation of the outside world.

Almost all levels of language knowledge and processing (from phonology to syntax and semantics) are known to be affected by knowledge of word structure at varying degrees.

A better understanding of the human strategies involved in learning and processing word structure lies at the heart of our comprehension of the basic mechanisms serving both language and cognition and is key to addressing the fundamental challenges for the study of the physiology of grammar.

According to dual-route approaches to word structure, recognition of a morphologically complex input word involves a preliminary full-form access to the lexicon, and an optional morpheme-based access of sub-word constituents of the input word, resulting from application of combinatorial rules taking care of on-line word segmentation. This view, challenged by several scholars, rests on the hypothesis of a direct correspondence between principles of grammar organisation (lexicon vs rules), processing correlates (storage vs computation) and localisation of the cortical areas functionally involved in word processing.

Other theoretical models have put forward a more nuanced indirect correspondence hypothesis. For instance, in the Word-and-Paradigm tradition, fully inflected forms are associatively related through possibly recursive paradigmatic structures, defining entailment relations between forms. Any serious appraisal of such an indirect correspondence requires extensive empirical testing on a wide array of morphologically rich languages of the sort spoken in Europe, and is likely to exceed the limits of both human intuition and box-and-arrow models of cognition. Increasing availability of multi-lingual data sets and computer models of language learning and processing will have much to say in this respect in the near future.





Another fundamental open issue is how theoretical models relate to neurobiologically-grounded models of word structure. Evidence of automatic sublexical segmentation of monomorphemic words triggered by pseudo inflectional endings lends support to a less deterministic and modular view of the interaction between stored word knowledge and on-line processing, based on simultaneously activating patterns of cortical connectivity reflecting distributional regularities in the input at the phonological, morphosyntactic and morphosemantic levels. At the same time, this evidence argues for a more complex and differentiated neurobiological substrate for human language than current models are ready to acknowledge, suggesting that brain areas devoted to language processing maximise the opportunity of using both general and specific information simultaneously, rather than maximize processing efficiency and economy of storage.

Such a dynamic view of the brain language processor makes contact with the human ability to retain symbolic sequences in Short Term Memory. Elements that are frequently sequenced in the subject's input are stored in Long Term Memory as single chunks, and accessed and executed in Short Term Memory as though they had no internal structure. Such an interaction between Short Term and Long Term Memory structures points to a profound continuity between word repetition/learning and other levels of grammatical processing in language.

People are known to understand, memorise and parse words in a context-sensitive and opportunistic way. Not only can speakers take advantage of token-based information such as frequency of individual, holistically stored words, but they are also able to organise them into paradigmatic structures (or word families) whose overall size and frequency is an important determinant of ease of lexical access and interpretation. Quantitative and analogy-based approaches to word interpretation lend support to this view, capitalising on stable correlation patterns linking distributional entrenchment of lexical units with productivity, internal structure and ease of interpretation.

These aspects agree with well-established psycholinguistic evidence that language comprehension is highly incremental, with readers and listeners continuously updating the meaning of utterances as they parse them. Much recent research suggests that language comprehension can be highly predictive, as long as the linguistic and non-linguistic context supports these predictions. Prediction can also be used to compensate for problems with noisy or ambiguous input and may explain the human advantage in parsing morphologically irregular forms (where morphosyntactic and morpholexical features are marked through extended exponence) over morphologically regular forms (where a morphological exponent systematically follows a full stem).

A parsimonious explanation of anticipatory mechanisms of language comprehension is that prediction uses some components for language production. There is indirect empirical evidence pointing in this direction: listeners activate the appropriate articulatory cortical areas for tongue and lips while listening to speech and brain areas that are associated with production during aspects of comprehension from phonology to narrative structure. This is in keeping with evidence of activation





of mirror neurons in monkeys by perceptual predictions and perceived actions, but may also be understood as involving context-sensitive language "emulators".

All of this points to a converging trend between computational and cognitive lines of scientific inquiry, supporting the view that grammar and lexical competence are acquired through minimal steps, shaped up by performance-driven factors such as memory limitations, frequency-based sensitivity, and modality-specific constraints, ultimately blur-ring the dichotomy between language knowledge and usage.

By exchanging words in ecological settings, we share, assess, modify, extend and structure our "semantic memory". Yet, the nature and content of such memory, the principles of its associative organisation and internal structure, the developmental role of the dynamic interaction between linguistic form, meaning and sensing are among the most controversial issues in the current linguistic and neurocognitive debate.

Suggestions in the literature range from relatively abstract representations, including hierarchical semantic networks and lexical conceptual structures, to more concrete perceptual- or motor-based representations. Each of these approaches faces difficulties. Abstract representations elude the issue of symbol interpretation by severing meaning from our system of experiences of the external world. On the other hand, linguistic units can combine and behave distributionally in ways that are not strictly predictable from their semantic properties. Inferences, sense extensions, metaphors and processes of concept composition and coercion show that grounded sensory motor knowledge does not suffice to account for our ability to extract meaning from language. Intermediate hypotheses need be entertained and empirically assessed, casting meaning as abstract, schematic representations, based on linguistically articulated, structured knowledge and word co-occurrences in large text samples, which are nonetheless embodied in human perceptual and motor systems. Researchers working in a neurocomputational framework have recently addressed issues of semantic knowledge arising from patterns of combinatorial information using more brain-like neural network simulations.

Interpretation of Noun-Noun compounds seems to require integration of the meaning representations associated with the two constituent nouns and in-dependently accessed from the lexicon. However, it has been shown that access to conceptual representations is considerably more dynamic and context-sensitive, so that the whole construction appears to prompt a process of selective activation of contextually-relevant semantic properties. From a computational standpoint, constraint-satisfaction approaches made the interesting suggestion that the interpretation of a complex construction makes use of pre-compiled, schematized information, memorized in the mental lexicon and applied probabilistically.





These aspects bring in the issue of interactive negotiation of referential and intentional word meanings in the process of learning word usages in daily communicative exchanges. Lexical pragmatics investigates the processes by which linguistically-specified (i.e. literal) word meanings are modified in use on the basis of factors related to pragmatic competence, such as knowledge of the specific communicative context, knowledge about the co-conversant(s), knowledge about the specific ongoing task and general knowledge of the world.

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

All workshop contributions were devoted to shedding light on all these factors as key to understanding the ontogenesis of word competence, creative usage of words in daily conversation, and the architecture of the mental lexicon.

In his inaugural talk "The impact of types of analogy on language acquisition" Wolfgang Dressler illustrated investigations on the early phases of morphological development by showing typological variations in the use of analogy-based induction. Contrastive plural formation in German, together with verb onflection, diminuitive formation and compound formation, provide ample evidence of overgeneralisation cases according to productive or unproductive rules.

Radovan Garabik in his talk "A Database of Slovak Verbs" illustrated methodological issues in the computer-assisted development of a database of the Slovak verb system, emphasizing anomalies and sub-regularities in Slovak morphology.

Paolo Acquaviva in his "Lexical decomposition meets Conceptual Atomism" illustrated the relation between the structure of words and the structure of lexical concepts, bringing to bear work in the psychology and pholosophy of concepts on theories of word meaning, and arguing for a specific view of lexical concepts on the basis of word semantics.

A continuous measure of irregularity based on the computational analysis of an arbitrary inflectional system was used by *Emmanuel Keuleers* to shed light on regularity and irregularity in morphological processing in his "On lazy learning, Irregularity and Lexical Decision megastudies". In particular, he investigated correlation patterns between classes of word neighbours in memory and acceptability of novel and non-existing forms by native speakers.

*Ida Raffaelli* illustrated a cognitive approach to "Morphosemantic patterns in lexical architecture" by emphasizing the connection between morpho-syntax and semantics as two levels that equally participate in lexical organisation.

Carita Paradis reported on an interdisciplinary approach to "Antonymy in language, thought and memory" by bringing together language technology, Linguistics, Psycolinguistics, and Neuropsycology.

"The role of morphology in acoustic reduction" was illustrated by *Mirjam Ernestus* in connection with the structure of the mental lexicon.





"Word stress systems and the role of stress in lexical processing" were illustrated by *Richard Wiese* by examining typological aspects of prosodic and lexical representations, providing evidence that processing strategies at the prosodic level may vary as a function of the stress patterns being processed and their level of lexical conditioning.

Principles of paradigms organisation were illustrated by *Fabio Montermini* in his "<u>Stem spaces and regularity in verbal inflection</u>" as a network of connected forms.

*Greville Corbett* presented the high variability of paradigms, within and across languages in "Shapes of lexemes: a typological perspective".

Models of processing printed words are presented by *Ram Frost*in his "What determines principles of word recognition and lexical structure: Evidence from cross-linguistic research". Every language presents a different solution for representic phonological, morphological and semantic information, and principles of lexical organization are language specific, reflecting the language phonological space and morphological structure.

Mila Vulchanova in "Morphology in ASD: Local processing bias and language" provided evidences for a connection between the local processing bias and the acquisition of morphology and grammar.

Christina Manouilidou provided evidence of lexical semantic representation in her "<u>The architecture</u> of the mental lexicon: insights from compound processing".

Dominiek Sandra by illustrating how young children are sensitive to morphological relations in his "Orthographic representations of homophonous word forms", suggested the existence of orthographic representations of whole-word forms in mental lexicon.

Harald Baayen presented a quantitative modeling of acoustic duration in his talk "Why is the signal smooth?" to provide evidence in favour of a highly combinatorial and distributed nature of word-based information in the brain.

Madeleine Voga investigated morphology in the bilingual lexicon in her "Inflectional and derivational processing in a morphologically rich language: an overview of Greek bilingual and monolingual data" presenting psycholinguistic experiments.

*Vito Pirrelli* in his "Word alignment and Morphology Induction" presented explanatory models of word learning and word processing as memory-driven processes.

In his presentation "Two words, one meaning: lexical organization and processing in bilinguals" Manuel Carreiras presented possible anwers to how bilinguals access lexical information from different languages in reading, by taking in examination inflection, derivation and compounding.

To sum up, word storage and word processing have traditionally been modelled according to complementary theoretical paradigms, in line with the classical tenet of so-called dual-route models of word structure (Pinker and Ullman 2002) assuming a modular dichotomy between memory and computation. The extensive linguistic, psycho-linguistic, neuro-linguistic and typological evidence reviewed and discussed in Pisa broached an overall different picture. Finite state automata, hierarchical lexica, stochastic classifiers, dynamic self-organizing memories, together with





"external" evidence of develpmental and process-oriented human language performance prompt a less modular view of the interplay between stored word knowledge and on-line processing. The view entails the co-existence of distributed patterns of combinatorial (rule-like) and lexical knowledge, mutually interacting in non-trivial ways and emphasizes a strong convergence between recent psycholinguistic, neurolinguistic, computational, historical and typological lines of inquiry, suggesting that stored morphological representations imply and are strongly implied by word processing principles. Knowledge of "how" and knowledge of "what" can hardly be decoupled in the current inquiry of Morphology. In this scenario, NetWordS can act as a catalyst in an incipent, promising trend towards mutidisciplinary integration. We believe that the current initiatives launched (on-going grants for exchange visits and small inter-disciplinary projects, the fothcoming Summer School on the Mental Lexicon, the yearly workshop and other related events) will play a significant role in strengthening this trend considerably.

## 4) PROGRAMME

#### Thursday, 24 November 2011

14.30 Welcome by convenors

Wolfgang U. Dressler (Austrian Academy of Sciences)

The impact of types of analogy on language acquisition

Radovan Garabik (Slovak Academy of Sciences)

A Database of Slovak Verbs

16.30-17.00 coffee break

Paolo Acquaviva (University College Dublin, Ireland), Phoevos Panagiotidis (University of Cyprus)

Lexical decomposition meets conceptual atomism

**Emmanuel Keuleers (Ghent University, Belgium)** 

Regularity/irregularity in morphological processing

18.45 End of Thursday Session

20,15 **Social Dinner** 

Friday, 25 November 2011

09.15 Start of Friday Session

Ida Raffaelli (University of Zagreb, Croatia)

Morphosemantic patterns in lexical architecture: a cognitive approach

Carita Paradis (Lund University, Sweden)

Antonymy in language, thought and memory

10.45-11.15 coffee break

Mirjam Ernestus (Radboud University Nijmegen & Max Planck Institute, Holland)

The role of morphology in acoustic reduction.



processing cortes

Richard Wiese (University of Marburg, Germany)

Word stress systems and the role of stress in lexical processing

12.45-14.30 Buffet lunch

Fabio Montermini (CNRS, Université de Toulouse le Mirail, France), Olivier Bonami (Université Paris-Sorbonne, France)

Stem spaces and regularity in verbal inflection

Greville Corbett (University of Surrey, UK)

Shapes of lexemes: a typological perspective

16.00-16.30 coffee break

Ram Frost (The Hebrew University, Israel)

What determines principles of word recognition and lexical structure? Evidence from cross-linguistic research

Mila Vulchanova, Valentin Vulchanov (Norwegian University of Science and Technology, Norway)

Morphology in ASD: Local processing bias and language

Manouilidou Christina and Ralli Angela (University of Patras, Greece)

The architecture of the mental lexicon: insights from compound processing

**Dominiek Sandra (University of Antwerp, Belgium)** 

Orthographic representations of homophonous word forms

19.30 End of Friday Session

20.00 Buffet dinner

#### Saturday, 26 November 2011

09.15 start of final Workshop Session

Harald Baayen (Tuebingen University, Germany)

Naive discriminative learning and construction morphology

Madeleine Voga (Paul Valéry University, France), Hélène Giraudo (CNRS, Université de Toulouse le Mirail, France), Anna Anastassiadis-Symeonidis (University of Thessaloniki, Greece)

Inflectional and derivational processing in a morphologically rich language: an overview of Greek bilingual and monolingual data

10.45-11.15 coffee break

Marcello Ferro, Claudia Marzi, Vito Pirrelli (Institute for Computational Linguistics, CNR Pisa Italy)

Word alignment and Morphology Induction

María Dimitropoulou, Jon Andoni Duñabeitia, Manuel Carreiras (Basque Centre on Cognition, Brain and Language, Spain)

Two words, one meaning: lexical organization and processing in bilinguals.

13.00 End of Workshop

13.15-15.30 NetWordS Steering Committee Meeting (including a buffet lunch)