

THE EUROPEAN NETWORK ON WORD STRUCTURE (NetWordS)

Standing Committee for the Humanities (SCH)

Scientific context

Introduction

Morphologically complex words are common to all European languages. They represent a fundamental part of what we mean by human language knowledge and the basic building blocks of language productivity. Nonetheless, words remain a challenging realm of scientific inquiry, at the interface between lexicon and grammar, requiring integration of a number of orthogonal disciplines and approaches, ranging from psycho- and neurolinguistics, to theoretical, variationist and historical linguistics, to memory processes and computational models of (sub)symbolic processing.

Scientists all over Europe are currently pursuing important lines of work on word structure, mostly supported by nationally-funded projects or bilateral cooperation programmes. There nonetheless seems to be a growing need for a larger-scale integrated European effort, focusing on common medium-term objectives, to promote interdisciplinary cross-fertilisation and synergy, and optimise research investments in terms of more convergent and complementary efforts. The European research scenario is particularly conducive to these goals, due to the robustly empirical character of its methodological stance and the unique range of relevant scientific domains where European scientists appear to have, at present, a huge potential for major breakthroughs.

By bringing together experts of various research fields (from brain sciences and computing to cognition and linguistics) and of different theoretical inclinations, **NetWordS** intends to advance our current awareness of the theoretical, typological, psycholinguistic, computational and neurophysiological evidence on the structure and processing of words, with a view to promoting novel methods of research and assessment for grammar architecture and language physiology. This will be achieved through knowledge networking and dissemination and scientific meetings organised over a four year period. Moreover, the Network aims to have a highly interdisciplinary profile, promote training and development of young scientists through short visits and exchange grants, and encourage the integration of new partners.

The running period of NetWordS is four years: from **May 2011** to **April 2015**.

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 three fundamental challenges for the study of the physiology of grammar, as detailed below.

Lexicon & Grammar

According to dual-route approaches to word structure, recognition of a morphologically complex input word involves two interlocked steps: i) preliminary full-form access to the lexicon, ii) 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. Such a view, recently 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. Recent 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 (possibly redundant) distributional regularities in the input at the phonological,

morpho-syntactic and morpho-semantic levels. At the same time, this evidence argues for a more complex and differentiated neuro-biological 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 maximise 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.

Word Knowledge & Word Use

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". In turn, anticipatory mechanisms of language comprehension may be closely related to mechanisms for Short Term memory content rehearsal such as Baddeley's phonological loop.

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 blurring the dichotomy between language knowledge and usage.

Words & Meanings

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 neuro-cognitive 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 independently accessed from the lexicon. However, it has recently 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, schematised information, memorised 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. Mediation of all these factors is key to understanding the ontogenesis of word meaning and its creative usage in daily conversation.

Aims and Objectives

Europe has firm and deep roots in as diverse disciplinary fields as theoretical models of language architecture, brain sciences, cognitive modeling, language development, short-term and long-term memory processes, psycho-computational models of lexical processing and storage, predictive models of language behaviour, machine learning, diachronic, diamesic and diastratic evidence of language varieties. The present healthy condition of European research makes time ripe for a larger-scale cross-disciplinary European effort into word structure aiming at

- * exploring the implications of domain-specific approaches for other fields
- * testing claims by broadening the empirical basis for their support
- * examining whether extensions of theoretical claims can be developed
- * promoting interdisciplinary cross-fertilisation and synergy
- * focusing on common medium-term objectives
- * optimising research investments in terms of convergent and complementary efforts

There is growing awareness that interdisciplinary cooperation in this area will have much wider

chances of success than traditional specialist work in highly focused knowledge domains and that it will consolidate European excellence in the field. By bringing together European (and non European) experts in complementary knowledge areas, **NetWordS** rises to this challenge by setting common research priorities, developing joint training programmes and establishing virtual cross-disciplinary laboratories and research infrastructures. Collaboration will unfold through the following steps:

- * discuss and develop consensual word representations in context
- * establish common experimental protocols and suggest novel ones
- * take stock of and integrate multilingual evidence based on the large array of European languages spoken and investigated in the Network
- * transfer best practice in use of new computational and statistical techniques for lexicon modeling
- * share experimental data, software and equipment
- * facilitate, through community building, the development of optimum cross-disciplinary and cross-linguistic research strategies
- * prompt and extend collaboration between partners
- * link European activities with the wider community world-wide.

Activities and instruments

Scientific partnerships

NetWordS promotes development of interdisciplinary transnational scientific partnerships through **short-visit grants**, that are assigned yearly on the basis of **open calls** for short-term project proposals. Scholars taking part in interdisciplinary activities funded through **NetWordS** grants convene periodically to discuss and disseminate results. Mature results are also expected to be disseminated and published through newsletters, existing major international conferences, journals and edited books. Short-visit grants are also geared towards planning focused collaborative work, with a view to catalyzing credible large-scale proposals within more application-oriented European projects and initiatives.

Training

NetWordS promotes training of young scientists through two instruments: **exchange grants** and **Summer schools**. Exchange grants are meant to cover long stays (up to 6 months) of promising junior researchers at leading hosting labs. Grants are assigned yearly on the basis of **open calls**.

Two Summer Schools, planned to take place in **2012** and **2014**, aim at bringing up a new generation of PhD students and young researchers with a truly interdisciplinary background, and shaping new ways of thinking that are conducive to cross-disciplinary breakthroughs.

Discussion and Networking

To maximise synergy, NetWordS sets itself the goal of defining a priority list of shared research topics of general interest (e.g. word reading, interpretation of complex nominals, developmental acquisition of inflection etc.) that will be addressed and discussed from the perspective of the three above-listed challenges through NetWordS **internet forums**. Forums are intended to stimulate discussion on domain-specific approaches and explore ways of integrating and extending current approaches also through data sharing.

Scientific meetings

NetWordS organises yearly **workshops** on interdisciplinary issues in word structure. The usual time frame is between late November and early December. A major final conference is planned to take place in 2015.

Funding

ESF Research Networking Programmes are principally funded by the Foundation's member Organisations on an à la carte basis. NetWordS is supported by:

- **Fonds zur Förderung der Wissenschaftlichen Forschung Österreich (FWF)**
Austrian Science Fund, Austria
- **Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO)**
Research Foundation Flanders, Belgium
- **Nacionalna zaklada za znanost (NZZ)**
National Foundation for Science, Croatia
- **Det Frie Forskningsråd – Kultur og Kommunikation (FKK)**
The Danish Council for Independent Research – Humanities, Denmark
- **Suomen Akatemia**
Academy of Finland / Research Council for Culture and Society, Finland
- **Centre National de la Recherche Scientifique (CNRS)**
National Centre for Scientific Research – France
- **Deutsche Forschungsgemeinschaft (DFG)**
German Research Foundation, Germany
- **Országos Tudományos Kutatási Alprogramok (OTKA)**
Hungarian Research Fund, Hungary

- **Irish Research Council for the Humanities and Social Sciences (IRCHSS)**
Ireland
- **Consiglio Nazionale delle Ricerche (CNR)**,
National Research Council, Italy
- **Norge Forskningsråd (NCR)**
Research Council of Norway, Norway
- **Javna agencija za raziskovalno dejavnost Republike Slovenije (SRA)**
Slovenian Research Agency, Slovenia
- **Departamento de Educacion, Universidades e Investigacion**
País Vasco, Spain
- **Forskningsradet för Arbetsliv och Socialvetenskap (FAS)**
Swedish Research Council, Sweden

Programme Chair

Vito Pirrelli

Institute for Computational Linguistics
Italian National Research Council
Pisa, Italy
Email: vito.pirrelli@ilc.cnr.it

Programme Coordinator

Claudia Marzi

Institute for Computational Linguistics
Italian National Research Council
Pisa, Italy
Email: claudia.marzi@ilc.cnr.it
coordination-networks@ilc.cnr.it

ESF Liaison

Arianna Ciula

Science Officer

Marie-Laure Schneider

Administration

European Science Foundation
Strasbourg, France
Email: sch-rnp@esf.org

Information

NetWordS encourages **integration** of new European partners through an open cooption scheme. For the latest information about the programme and if you wish to know more about how to join NetWordS, consult the NetWordS websites or contact the info mail address:

www.networks-esf.eu
www.esf.org/networks
info-networks@ilc.cnr.it

NetWordS RNP ref. number is **09-RNP-089**