

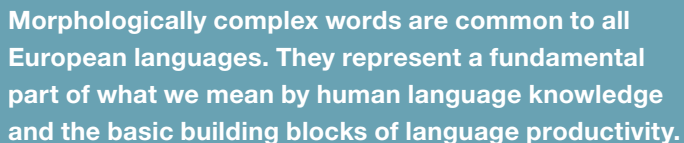
neural networks storage processing agents acquisition computing fMRI neurons brain sciences cortex dyslexia priming memory imaging time learning

Research Networking Programme

The European Network on Word Structure (NetWordsS)

Standing Committee for the Humanities (SCH)

linguistics morpheme grammar language paradigm word structure cognition space lexicon mind meaning aphasia ERP



Scientists all over Europe are currently pursuing important lines of work on word structure, mostly supported by nationally-funded projects or bi-lateral 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 to 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.

Moreover, the Research Networking Programme aims to have a highly interdisciplinary profile, to promote training and development of young scientists through short visits and exchange grants, and to encourage the integration of new partners.

The running period of the ESF NetWordS Research Networking Programme is four years, from May 2011 to April 2015.

Scientific Context

Words are the basic building blocks of language productivity, establishing the most immediate connections between language and our conceptualisation of the outside world. Besides, they represent complex interface units, which are not only parts of larger constructions such as phrases or sentences, but are themselves, in all European languages, made up out of simpler meaningful sub-lexical constituents, such as roots and affixes.

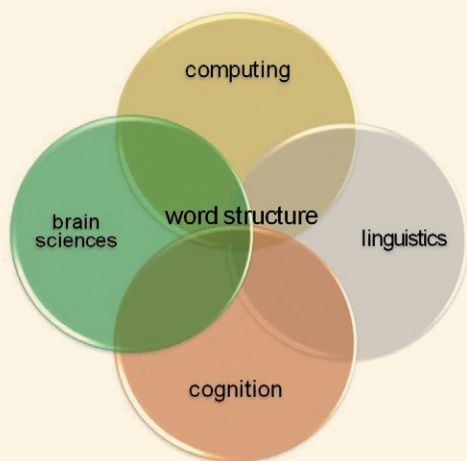
Such a dual status of morphologically complex words, at the interface between lexicon and grammar, raises a number of fundamental questions, many of which are still unanswered. How are words processed in working memory? Are they stored in long-term memory as a whole or rather composed 'on-line' in working memory from stored sub-lexical constituents? Do both knowledge-based factors, such as formal regularity and semantic transparency, and usage-driven factors, such as word length and frequency, play any role in this? Does word-level knowledge require parallel development of form and meaning representations, or do the latter develop independently at a different pace? How do word meanings

function and combine in communicative contexts and evolve through learning? Does lexical knowledge affect on-line processing? Do the dramatic differences in word structures across the languages of Europe impact on processing models worked out on the basis of a single language? What neurobiological patterns of connectivity sustain word processing and storage in the brain?

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 thus lies at the heart of our comprehension of the basic mechanisms serving both language and cognition and is key to addressing the three fundamental challenges for the study of the physiology of grammar that are described in more detail below.

Lexicon and 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. Algorithmically, step ii) is taken if and only if step i) fails to find any matching access entry in the lexicon. 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. Although such a direct correspondence is probably



the most straightforward model of the grammar-processing relation, it may only be the artefact of outdated views of memory as rote storage. In fact, 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. We believe that increasing availability of multilingual 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 and theories of word structure. Recent evidence of automatic sub-lexical

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 connectionist one-route 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 also with what we know about the connection between language acquisition and processing and the human ability to retain sequences of symbols 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, in line with neurobiological approaches according to which Long Term Memory refers to structural networks and Short Term Memory benefits from activation of the same networks.

Word Knowledge and Word Use

People are known to understand, memorise and parse words in a context-sensitive, construction-based 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 paradigm-like structures (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.

Figure 1.

Receptive letter fields on a Temporal Self-Organising Map of 1600 nodes trained on the full text of "The Little Prince" by Antoine de Saint-Exupéry (1900-1944). Receptive fields identify the nodes of the map that are most highly activated upon completion of each input word form. Fontsize and levels of grey of receptive fields are proportional to the frequency of word forms in the training text. Topological organisation reflects word proximity at the level of morphological structure.

© Courtesy of Dylan Lab, CNR Institute for Computational Linguistics, Pisa, Italy

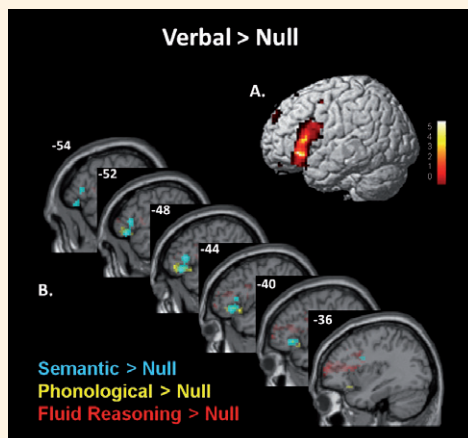


Figure 2.

Left lateral PFC involvement in verbal fluid reasoning.

(A) Brain rendering showing activations for Verbal Fluid Reasoning > Null contrast for all participants. (B) Left lateral PFC regions active in the semantic reasoning task (cyan), in the phonological reasoning task (yellow), and common fluid reasoning regions across verbal and visual domains (red). x coordinates are provided for sagittal views.

© Courtesy of Basque Center on Cognition, Brain and Language, Donostia, San Sebastián, Basque Country, Spain

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

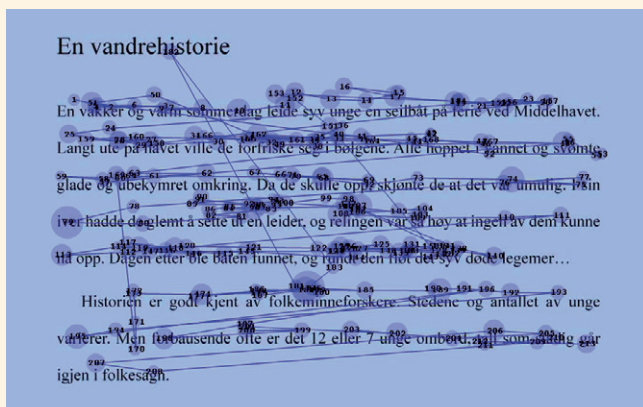
Figure 3.

Text processing in dyslexia.

An example of reading patterns from a study comparing dyslexic and non-dyslexic participants.

Dots represent fixations and the dots' sizes are related to fixation length. Numbering of dots and lines between dots indicate the order of the fixation sequence.

© Courtesy of Language Acquisition and Language Processing Lab, Norwegian University of Science and Technology, Trondheim, Norway



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 and 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

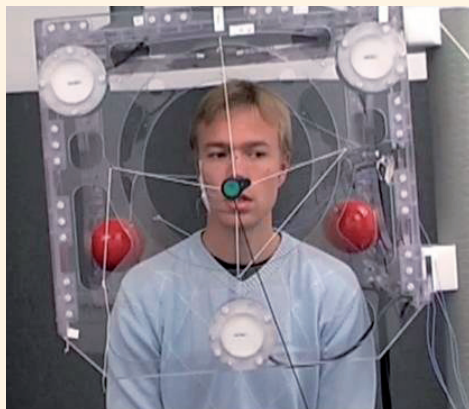


Figure 4.

Articulograph AG500 allowing sensors placed on the tongue, lips and reference points to be tracked in 3D using an electromagnetic field with high spatial and temporal resolution.

© Courtesy of Humanities Lab, Lund University, Sweden

account for our ability to extract meaning from language. Intermediate hypotheses need to 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 such as ‘bread knife’ requires 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, as illustrated by so-called ‘one-off compounds’ such as Downing’s ‘apple juice seat’ example.

Aims and Objectives

Europe has firm and deep roots in as diverse disciplinary fields as theoretical models of language architecture, brain sciences, cognitive modelling, 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 diastatic 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 data based on the large array of European languages
- transfer best practice in use of new computational and statistical techniques for lexicon modelling
- share experimental evidence, software and equipment
- facilitate, through community building, the development of optimum cross-disciplinary and cross-linguistic research strategies
- prompt and extend collaboration between partners and link European activities with the wider community world-wide.

Programme 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 catalysing 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, are aimed 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.

Scientific Meetings

NetWordS organises yearly **workshops** on inter-disciplinary 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.

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 (Lexicon and Grammar, Word Knowledge and Word Use, Words and Meanings) 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.

Global Dimension

NetWordS promotes a global world-wide collaborative dimension through established links with other cognate research initiatives such as the *Mental Lexicon Research Group* in Canada.

Joining NetWordS

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 use the email contact:

www.esf.org/networks

www.networks-esf.eu

info-networks@ilc.cnr.it

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 in Österreich (FWF)**
Austrian Science Fund, Austria
- **Fonds voor Wetenschappelijk Onderzoek-Vlaanderen (FWO)**
Research Foundation Flanders, Belgium
- **Hrvatska zaklada za znanost (HRZZ)**
Croatian Science Foundation, Croatia
- **Det Frie Forskningsråd - Kultur og Kommunikation (FKK)**
The Danish Council for Independent Research – Humanities, Denmark
- **Suomen Akatemia/Finlands Akademi**
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 Alapprogramok (OTKA)**
Hungarian Scientific Research Fund, Hungary
- **An Chomhairle um Thaighde sna Dána agus sna hEolaíochtaí Sóisialta (IRCHSS)**
Irish Research Council for the Humanities and Social Sciences, Ireland
- **Consiglio Nazionale delle Ricerche (CNR)**
National Research Council, Italy
- **Norges Forskningsråd**
Research Council of Norway, Norway
- **Slovenská akadémia vied (SAV)**
Slovak Academy of Sciences, Slovak Republic
- **Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS)**
Slovenian Research Agency, Slovenia
- **Departamento de Educacion, Universidades e Investigacion**
País Vasco, Spain
- **Forskningsrådet för arbetsliv och socialvetenskap (FAS)**
Swedish Council for Working Life and Social Research, Sweden

NetWordS Steering Committee

- **Dr Vito Pirrelli** (Chair)
Institute for Computational Linguistics,
Italian National Research Council,
Pisa • Italy
Email: vito.pirrelli@ilc.cnr.it
- **Dr Paolo Acquaviva**
School of Languages and Literatures,
University College Dublin, Belfield,
Dublin 4 • Ireland
Email: paolo.acquaviva@ucd.ie
- **Professor Manuel Carreiras**
Basque Centre on Cognition, Brain
and Language, St. Sebastian • Spain
Email: m.carreiras@bcbl.eu
- **Professor Walter Daelemans**
Department of Linguistics, Faculty of Arts,
University of Antwerp, Antwerp • Belgium
Email: walter.daelemans@ua.ac.be
- **Professor Wolfgang Ulrich Dressler**
Institute of Corpus Linguistics and
Text Technology, Austrian Academy of
Sciences, Vienna • Austria
Email: wolfgang.dressler@univie.ac.at
- **Dr Tomaž Erjavec**
Department of Knowledge Technologies,
Jožef Stefan Institute, Ljubljana • Slovenia
Email: tomaz.erjavec@ijs.si
- **Dr Radovan Garabík**
Ľ. Štúr Institute of Linguistics, Slovak
Academy of Sciences, Bratislava •
Slovak Republic
Email: garabik@kassiopeia.juls.savba.sk
- **Dr Hélène Giraudo**
Laboratoire CLLE-ERSS (CNRS-UMR
5263), Université de Toulouse le Mirail,
Toulouse • France
Email: helene.giraudo@univ-tlse2.fr
- **Dr Krista Lagus**
Adaptive Informatics Research Centre,
Aalto University School of Science and
Technology, Espoo • Finland
Email: krista.lagus@tkk.fi
- **Professor Carita Paradis**
Centre of Languages and Literature, Lund
University, Lund • Sweden
Email: Carita.Paradis@englund.lu.se
- **Professor Ingo Plag**
English Linguistics, University of Siegen,
Siegen • Germany
Email: plag@anglistik.uni-siegen.de
- **Professor Gábor Prószéky**
Faculty of Information Technology
of Budapest, Pázmány Péter Catholic
University, Budapest • Hungary
Email: proszeky@itk.ppke.hu
- **Professor Ida Raffaelli**
Faculty of Humanities and Social
Sciences, University of Zagreb,
Zagreb • Croatia
Email: ida.raffaelli@ffzg.hr
- **Professor Hanne Ruus**
Department of Scandinavian Studies
and Linguistics, University
of Copenhagen • Denmark
Email: haru@hum.ku.dk
- **Professor Mila Dimitrova-Vulchanova**
Department of Modern Languages,
Faculty of Arts, Norwegian University
of Science and Technology,
Trondheim • Norway
Email: mila.vulchanova@ntnu.no

Programme Coordinator

- **Dr 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

Dr Arianna Ciula/Dr Barry Dixon

Science

Ms Marie-Laure Schneider

Administration

Humanities and Social Sciences Unit

European Science Foundation

1 quai Lezay-Marnésia

BP 90015

67080 Strasbourg cedex

France

Tel: +33 (0)3 88 76 71 39

Fax: +33 (0)3 88 37 05 32

Email: sch-rnp@esf.org

For the latest information on this Research
Networking Programme consult the
NetWordS websites

www.networks-esf.eu

www.esf.org/networks

or write to **info-networks@ilc.cnr.it**

Cover design by Dylan Lab

The European Science Foundation (ESF) was established in 1974 to provide a common platform for its Member Organisations to advance European research collaboration and explore new directions for research. It is an independent organisation, owned by 78 Member Organisations, which are research funding organisations and research performing organisations, academies and learned societies from 30 countries. ESF promotes collaboration in research itself, in funding of research and in science policy activities at the European level.

European Science Foundation

www.esf.org

August 2011 – Print run: 1 000