

PHRASALEX-II

Phraseological Approaches to Learner's Lexicography

Online Workshop on Thursday 22 July and Friday 23 July 2021

WORKSHOP DESCRIPTION

PHRASALEX aims to bring together experts in the fields of (computational) lexicography, lexicology, phraseology, cognitive linguistics and foreign language teaching, in order to reflect on theoretical and methodological aspects related to

- the **interdependence of lexis and grammar**,
- the **phraseological nature of language**,
- and the implications of these perspectives for the **design of (learner's) dictionaries**.

The first PHRASALEX workshop (sensepattern dictionaries.net), which was held at Modena University in September 2019, was inspired by PhraseBase, a project in learner's lexicography carried out by a small team of researchers at Hildesheim University and Heidelberg University. The idea at the core of the project, namely the existence of lexical units of a language endowed with a phraseological nature, originates from linguistic approaches covering, on the one hand, the interplay between lexical patterns and grammatical frameworks, such as lexicogrammar (cf. Halliday 1992), Gross's *classes d'objets* (1994), or Herbst's notion of *Konstruktikon* (2016) and, on the other hand, a cognitive view of the lexicon (cf., among others, Oakley 2010, Casasanto 2017, Taylor 2017, and Ostermann 2015).

The lexicographic approach at the core of PhraseBase is informed in particular by Sinclair's notion of *normal patterns of usage* as the true lexical units of a language (Sinclair 1986; cf. also Sinclair 1996). This concept laid the foundations of the Collins COBUILD English Dictionary in 1987, as well as by Hanks's Theory of Norms and Exploitations (2013). Our approach distinguishes itself from other lexicographic approaches in that all relevant meaning patterns of a word are systematically and

explicitly specified (DiMuccio-Failla & Giacomini 2017a/2017b). The notion of "linguistic restrictions" intended as lexical, cognitive, and conceptual constraints allows us to discern the different sense-specific lexical units of a word and to order them on the basis of their closeness to a prototypical meaning.

The project has taken up diverse corpus-based methods for the identification and (in)formal description of phraseological sense patterns in lexicography, gradually adapting them to the specific objectives (cf. Giacomini & DiMuccio-Failla 2019, Giacomini, DiMuccio-Failla & Lanzi 2020). The main language of investigation in the PhraseBase project is Italian, but parallel multilingual experiments with English, German, and French are being carried out to provide relevant parameters for method and data validation (cf. Orlandi, Giacomini & DiMuccio-Failla 2019).

The international and interdisciplinary context of PHRASALEX II will foster interaction and collaboration among participants while promoting exchange of ideas on topics close to the project, in particular on possible ways of (formally and informally) representing the interface between lexis and grammar as well as the phraseological dimension of language in learner's lexicography.

The workshop will also be a platform for disseminating the latest results of the PhraseBase project and some new ideas regarding:

- A structured model for the phraseological description of the lexicon,
- Extraction of simple and complex collocations from corpora,
- New NLP experiments on lexicographic data extraction from corpora (e.g. sentence embeddings),
- Observations on new multilingual data,
- PoS-specific issues, e.g. aspectual behaviour of verbs and its cognitive formalisation,
- The formal representation of data (e.g. definitions, polysemy, hypernymy),
- Aspects of online data presentation,
- The modelling of a wordnet-like phraseological resource based on the core phraseological lexical units (a candidate for WordNet 2.0).

PHRASALEX II is organised in four sessions, each dedicated to specific research questions:

– **Session I: Introduction to the project**

The first session of the workshop is dedicated to project-related contents, including a description of theoretical backgrounds, multilingual data analysis, applications and challenges (Giacomini / Di Muccio-Failla), observations on the importance of suitable context length for the

detection of aspectual features of verbs (Piepkorn) as well as new experiments with sentence embeddings in the analysis of argument structure (Kliche & Giacomini).

– **Session II: Corpus Analysis & Polysemy**

Session II offers special insights into the analysis of corpora, concentrating in particular on the advantages of Corpus Pattern Analysis from a language learning perspective (Moze) and on corpus-related issues in research on verb valency, e.g. the status of different argument types and polysemy (Fasciolo). Polysemy is the main topic of two further talks, which deal with the analysis of verbal polysemy through pattern-based resources (Jezek & Marini) and with regular polysemy and disambiguation of meaning in corpora (Renau).

– **Session III: Lexicography**

The third session opens with a look at issues concerning (learner's) dictionaries, namely computer-mediated communication (Abel) and language awareness in dictionary use (Nied Curcio). Further two talks are dedicated to the features and lexicographic representation of partially filled constructions (Piunno) and to the role of discourse-oriented corpus studies in lexicography (Gabrielatos).

– **Session IV: Argument structures**

The last session is fully dedicated to theoretical and practical aspects of research on argument structures. Starting from distributional semantics and Machine-Learning-based approaches, namely the application of transformers and distributional models to the analysis of argument structures (Chersoni), Session IV moves on to the modelling of constructions in a construction (Herbst & Uhrig). The third talk discusses semi-automated procedures for valency generation, analysis and lexicographic implementation (Domínguez Vásquez), and is followed by a study on argument structures in the syntax-lexicon continuum (Mereu & Piunno).

The main language of the workshop is English, however some talks will be given in a different language.

Papers from the workshop will be published later this year.

References:

Casasanto, D. (2017). Relationships between Language and Cognition. In: B. Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press, 19-37.

DiMuccio-Failla, P. & Giacomini, L. (2017a). Designing an Italian learner's dictionary based on Sinclair's lexical units and Hanks's corpus pattern analysis. In: *Proceedings of the Fifth eLex Conference, Electronic Lexicography in the 21st Century*. Leiden, Netherlands.

DiMuccio-Failla, P. & Giacomini, L. (2017b). Designing a learner's dictionary with phraseological disambiguators. In: R. Mitkov (ed.), *Computational and Corpus-Based Phraseology. Second International Conference, Europhras 2017, LNAI 10596*. Springer, 290-305.

- Giacomini, L. & DiMuccio-Failla, P. (2019). Investigating Semi-Automatic Procedures in Pattern-Based Lexicography. In: Proceedings of the eLex 2019 conference. Electronic lexicography in the 21st century. Sintra, Portugal.
- Giacomini, L., DiMuccio-Failla, P. & Lanzi, E. (2020). The interaction of argument structures and complex collocations: role and challenges for learner's lexicography. In: Proceedings of the XIX EuraLex International Congress, Alexandroupolis, Greece.
- Gross, G. (1994). Classes d'objets et description des verbes. In: *Langages*, 15-30.
- Halliday, M. A. K. (1992). Some lexicogrammatical features of the zero population growth text. *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*. Amsterdam: John Benjamins, 327-358.
- Hanks, P. (2013). *Lexical analysis: Norms and exploitations*. MIT Press.
- Herbst, T. (2016). Wörterbuch war gestern. Programm für ein unifiziertes Konstruktikon. In: S. J. Schierholz, R. H. Gouws, Z. Hollós & W. Wolski (eds.), *Wörterbuchforschung und Lexikographie*. Berlin/Boston: de Gruyter.
- Oakley, T. (2010). Image schemas. In: D. Geeraerts & H. Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*. Oxford/New York: Oxford University Press, 214-235.
- Orlandi, A., Giacomini, L. & DiMuccio-Failla, P. (2019), I disambiguatori fraseologici nella lessicografia di apprendimento: una proposta per l'italiano e il francese. In: *Repères DoRiF n.18 – Phraséodidactique: de la conscience à la compétence*, DoRiF Università, Roma.
- Ostermann, C. (2015). *Cognitive lexicography: A new approach to lexicography making use of cognitive semantics (Lexicographica Series Maior, Vol. 149)*. Berlin, München, Boston: De Gruyter.
- Sinclair, J. (1986). First throw away your evidence. In: Leitner, G. (ed.), *The English Reference Grammar*, Max Niemeyer Verlag, 56-65.
- Sinclair, J. (1996). The search for units of meaning. In: *Textus: English Studies in Italy* 9(1), 75-106.
- Taylor, J. (2017). *Lexical Semantics*. In: B. Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press, 246-261.