2 Abstracts of all lectures

Lexicography and identity, indigenous languages

Historical Comparison of the Iconic Dictionaries of the Three Baltic Nations

Andrejs Veisbergs

Keywords: *Latvian*, *Estonian*, *Lithuanian*, *explanatory dictionary*.

Latvian, Estonian and Lithuanian lexicography are characterised by similar early development, despite different historical development and different language-contact situations. There is a clear dominance of bilingual and multilingual dictionaries, which were initially compiled to serve the needs of the clergy in the main contact-language pairs and triples. After achieving independence early in the 20th century, all three states embarked on large, iconic projects of nation building and prestige, of very different scope and timescale from the bilingual dictionaries. These projects had both extralinguistic prestige objectives (proving the wealth of the language resource, demonstrating it to the outside world, putting the languages on the comparative linguistics map) and linguistic objectives (registering, etymologising, explaining, expanding, purifying and stabilising the wordstock). Elements of language engineering can be observed in prescriptivism (Estonian language planning) and xenophobic purism. These large, iconic projects were led by the well known linguists of the time. Comparing the three, we can see that Latvian and Lithuanian projects are more retrospective (focusing on the heritage) while the Estonian dictionary is more forward-looking. The status of these iconic dictionaries is also different today: only the Latvian project has retained it.

A North Sami translator's mailing list seen as a key to minority language lexicography

Trond Trosterud & Berit Nystad Eskonsipo

Keywords: *North Sami*, *minority language*, *language planning*, *vocabulary planning*.

The topic of this investigation is the set of Norwegian words discussed on a North Sami translator's mailing list during one year, altogether 313 words. The words were grouped according to text domain, and to what extent existing dictionaries were able to meet the translators' needs. Most of the words discussed on the list were missing in relevant reference works. Two reasons for this are the paucity of North Sami text and the fact that Norwegian to North Sami lexicography has had North Sami dictionaries and word lists as their basis. The main finding of the article is that the words put under scrutiny by the

mailing list belong to common, everyday language. The translator list thus may function as a roadmap for future North Sami lexicography.

The Dictionary of the Contemporary Slovak Language: A Product of Tradition and Innovation

Alexandra Jarošová & Vladimír Benko

Keywords: *Slovak explanatory dictionary, prescriptivism and descriptivism*.

After having published the first two volumes of a multivolume monolingual dictionary of Modern Slovak, herein we try to summarise the basic concepts as they have been implemented in the actual dictionary text and introduce some extralinguistic and linguistic contexts relevant to our language and political situation. Both the traditions and innovations that have influenced the actual lexicographic decisions are presented. The extralinguistic contexts are represented above all by the existence of a special linguistic institution authorised to issue codification publications, as well as by the existence of the 'Act on the State Language', the amendment to which was passed in 2010. In Slovak lexicography, the linguistic contexts are governed by two contradicting traditions of prescriptivism and descriptivism. The presented discussion of the macro- and microstructure of the dictionary introduces some novel lexicographical solutions.

Cypriot Greek Lexicography: A Reverse Dictionary of Cypriot Greek

Charalambos Themistocleous, Marianna Katsoyannou, Spyros Armosti & Kyriaki Christodoulou

Keywords: *reverse dictionary, Cypriot Greek, orthographic variation, orthography standardisation, dialectal lexicography*.

This article explores the theoretical issues of producing a dialectal reverse dictionary of Cypriot Greek, the collection of data, the principles for selecting the lemmas among various candidates of word types, their orthographic representation, and the choices that were made for writing a variety without a standardized orthography.

TEDIPOR: Thesaurus of Dialectal Portuguese

Sandra Pereira & Raïssa Gillier

Keywords: *dialectology, lexicography, vocabulary, geolinguistics, database*.

The Thesaurus of Dialectal Portuguese (TEDIPOR) is a dialectal tool under construction. It is a fact that the changing and decreasing of the rural world and of its ways of life lead to the banning and to the extinction of a huge quantity of dialectal lexical variation. Therefore, TEDIPOR is a testimony of a disappearing lifestyle that will be preserved in a rich lexical database. It also aims to make available to the scientific community and to the society in general an important amount of dialectal, ethnographic and cultural information that is often difficult to access and handle. The sources to be integrated in the database include the glossaries of academic monographs (between 1940s and 1970s), atlases, dialectal inquiries and other papers containing dialectal information. In order to demonstrate the usefulness of this tool, all the designations concerning the concepts *bebedeira* (drunkenness), *bêbedo* (drunk) and *embebedar* (to get drunk) were gathered and are analysed under a lexical approach. The results demonstrate that many of the denominations found in TEDIPOR are not attested by Portuguese dictionaries, revealing that these materials are an important source for lexicographic research. Furthermore, the geographical distribution of the concepts *bebedeira* (drunkenness), *bêbedo* (drunk) and *taberna* (tavern) is also presented from a cartographic perspective. The maps show that it is possible to identify dialectal areas for some of the designations. Both the lexical and geographical analyses illustrate the potential of TEDIPOR, especially for Dialectology and Geolinguistics.

The Lexicon of Buda. A Glimpse into the Beginnings of Mainstream Romanian Lexicography

Bogdan Harhata, Maria Aldea, Lilla Marta Vremir & Daniel-Corneliu Leucuta

Keywords: *Romanian lexicography*, *Transylvania*, *academic*, *tradition*.

This paper is the result of a project aimed to e-ready a dictionary dating back to 1825, namely the *Lexicon of Buda* (1825) that is often referred to as the starting point of Romanian modern lexicography. The expressed aim of this paper is to illustrate that The Lexicon of Buda anticipates a log tradition in the academic Romanian lexicography. In order to provide a better understanding of why this lexicon holds its place among lexicographers and linguists, there is a brief description of the status of Romanian lexicography previous to 1800, followed by a short historical development. The second part illustrates the technical novelties inherited by Romanian Academy's lexicographic works, and shows that what this lexicon and the academic dictionaries have in common are the central position in the Romanian cultural establishment and the fact that they are normative and aim to unify the linguistic norm of Romanian.

Kinship terminology in English–Zulu/Northern Sotho dictionaries – a challenge for the Bantu lexicographer

Danie J. Prinsloo & Sonja Bosch

Keywords: *lemmatisation, kinship terminology, Zulu, Northern Sotho, decision tree algorithm.*

The lemmatisation and treatment of kinship terminology in general dictionaries, and in learners' dictionaries in particular, is an established lexicographic tradition. However, due to the nature and complexity of kinship terminology in certain languages, comprehensive guidance is needed for the correct use of kinship terms especially for text and speech production purposes. In such cases the lexicographer plays an important role as the mediator between a complex kinship terminology system and the target user of the dictionary. The aim of this paper is to suggest strategies for the treatment of kinship terms in paper and electronic dictionaries with English as the source and Zulu/Northern Sotho as the target language. Zulu as well as Northern Sotho belong to the Bantu language family of Africa, and can be regarded as variations of the Iroquois type of kinship terminology system (Murdock 1949), a unilineal descent system which distinguishes between Father's and Mother's Kin.

In this paper, we firstly critically compare the kinship terminology structures of English and Zulu/Northern Sotho, and secondly evaluate the treatment (or lack thereof) in Zulu and Northern Sotho dictionaries. Given that in traditional paper dictionaries, it was not possible for lexicographers to do justice textually to the description of complex kinship terms, we suggest an innovative design for an interactive electronic dictionary with English as the source language and Zulu/Northern Sotho as the target that guides the user step-by-step through a sequence of selection processes utilising a decision tree algorithm, to the correct term. Such a design could result in a dynamic as well as a static system. Links to various types of corpora will not only ensure authentic examples, but also collocations and frequency of occurrence.

Corpus-driven lexicography

The first Slovene automatically compiled dictionary of abbreviations

Mojca Kompara

Keywords: *computational lexicography*, *dictionary*, *abbreviation*.

Abbreviations are difficult to deal with (Gabrovšek, 1994) and represent a growing phenomena present in all languages. The scope of this article is to present the first Slovene automatically compiled dictionary of abbreviations. In the paper we present how we automatically extract abbreviation-expansion pairs out of newspaper texts and obtain genuine pairs, how we cope with the automatic editing phase and add language qualifiers to expansions and transform non-nominative expansions into nominative. The first Slovene automatically compiled dictionary of abbreviations is available online, free of charge, on the web site of Termania. It is the first dictionary produced automatically from newspaper articles with the help of algorithms. Algorithms represent a link between the text and the semi automatic production of a dictionary of abbreviations. That is why the production and further development of algorithms is essential and useful for lexicographers.

Ein Korpus als Garant zuverlässiger lexikographischer Informationen? Eine vergleichende Stichprobenuntersuchung

Ulrich Schnörch & Petra Storjohann

Schlüsselwörter: *Zuverlässigkeit*, *Authentizität*, *Korpusmethoden*, *Analysemethoden*, *Arbeitsgrundlagen*.

Current working practice of established German dictionaries incorporates large corpora as the basis of most analyses, descriptions and presentations. It is, however, individual lexicological and/or different corpus-methodological approaches that play a crucial role in the process of extracting and documenting lexicographic information in individual reference works. This paper addresses the question of how reliable information is in some electronic German dictionaries. Objects of our investigation are different types of corpus dictionaries, e.g. a digitized dictionary, a reference work that compiles its data fully automatically, a lexicographic system combining different electronic resources, and a corpus-assisted dictionary that examines and interprets its corpus data lexicographically. Critical examinations of such reference works inevitably come up with questions of authenticity and reliability of the given dictionary information. The advantages and disadvantages of various lexicographic or corpus-linguistic methods which are individually implemented will be outlined and critically analyzed with the

help of examples. According to an extensive study (cf. Müller-Spitzer 2011) reliability of given information is one of the key criteria assigned to any reference work by users. We will elicit how different corpus methods expose different descriptions of natural discourse and how they answer questions of authenticity, typicality and reliability with regard to phenomena such as meaning spectrum, collocations, antonymy and hyperonymy. Overall, this paper is a critical account of the current German lexicographic developments. It will include discussions on meta-lexicographic demands and focus on whether there are suitable complementary corpus approaches providing authentic dictionary information to a satisfactory extent.

Electronic Corpora and Dictionary Definitions: the Word "Patriotism" in COCA and Online Merriam-Webster Dictionary

Maria Konovalova & Igor Tolochin

Keywords: *Merriam-Webster, patriotism, COCA, word meaning*.

The article analyses the word '**patriotism**' in the Contemporary Corpus of American English (COCA) and the results are compared with the definition of the same word in the Online Merriam-Webster Dictionary. The comparison points out that one of the most comprehensive dictionaries of American English does not provide a consistent and clear structure of the senses of the word '**patriotism**' as it is used today. Some suggestions for the improvement of the definition are offered.

Word lists in Reference Level Descriptions of CEFR (Common European Framework of Reference for Languages)

Carla Marello

Keywords: *CEFR, reference description level, learner's dictionaries, corpus-driven lexicography*.

In this paper we consider how profiles, or sets of Reference Level Descriptions (hereon RLDs), of the CEFR (Common European Framework of Reference for Languages) for English, German, French, Spanish and Italian present their word lists. We focus on B2 because it is the RLD level reached and published by all the profiles and also because vocabulary for C1 and C2 levels cannot be delimited. RLDs sets provide detailed information about the language that learners can be expected to demonstrate at each level and their word lists are corpus-based. We comment on their actual or prospective links with learner's dictionaries and conclude that learner's dictionaries need not enter in the profiles, which are meant for professionals, including curriculum planners, material writers and teachers. Learner's dictionaries enter German and English profiles because

RDLs planners want to instruct teachers how to go beyond their lists and train students to conduct better look-ups. It should be rather the other way: learner's dictionaries should take advantage of the fact that in the profiles CEFR levels are assigned to each individual meaning of these words, either openly as in the German and English profiles or more implicitly as in the Italian, French and Spanish.

Domain Specific Corpora from the Web

Avinesh PVS, Diana McCarthy, Dominic Glennon & Jan Pomikálek

Keywords: *domain corpus, DANTE, WebBootCat*.

Language usage is dependent on domain and, as a consequence, domain specific corpora are extremely useful for language learning and lexicography. It is possible to label heterogeneous data for domain either manually or automatically using human knowledge or machine learning. State-of-the-art text classification uses supervised techniques whereby a system learns from previously annotated data. This works well when such data is available in sufficient quantities for supervised machine learning, though often that is not the case depending on the domain and language required. Moreover, this approach assumes that the heterogeneous data in the available corpus covers the required domains. In this paper we present the results of an approach using WebBootCat to retrieve data from the web in eight specific domains. A key component of this work was the use of the DANTE database for generating seed words for initial web data retrieval. To tailor the corpus to the nuances of the domain categorisation that we required, we used some of our own corpus data already annotated with subject codes (domain codes) to help refine the seed words used at the start of the iterative web retrieval process. Human effort was needed to refine a whitelist of words for each domain to reduce the chance of irrelevant data due to ambiguous terms in the seeds and extracted keywords used for subsequent retrieval. The domain corpora retrieved are loaded in the Sketch Engine. The word sketches and sketch difference functionality help reveal appropriate domain specific behaviour of words in the respective corpora.

Automatic example sentence extraction for a contemporary German dictionary

Jörg Didakowski, Lothar Lemnitzer & Alexander Geyken

Keywords: *example extraction, digital dictionary, practical lexicography, natural language processing*.

The integration of illustrative examples into monolingual dictionaries provides an intuitive means for grasping the meaning of a word. Tight space constraints of print media no longer apply with online dictionaries. Thus, the inclusion of examples is obviously a useful complement or substitute for the traditional ways of meaning

exemplification. In this article, an approach is presented to automatically extract example sentences from a large German corpus collection. The extraction is done on the basis of the notions of sentence readability and complexity and word usage. The extracted examples are a good pre-selection for further integration into a digitized version of a contemporary German dictionary by lexicographers. A quantitative and qualitative evaluation of the extraction results is presented in the article. The work is related to the dictionary project *Digitales Wörterbuch der deutschen Sprache* (The Digital Dictionary of the German Language, DWDS in short) which integrates multiple dictionary and corpus resources and language statistics on the German language in a digital lexical information system which can be accessed on-line.

Using CPA to represent Spanish pronominal verbs in a learner's dictionary

Irene Renau & Paz Battaner

Keywords: *Corpus Pattern Analysis (CPA), learners' dictionaries, Spanish pronominal verbs*.

In this paper, we deal with different aspects of Spanish pronominal verbs, which can be classified in various types that are often confused and mixed up in real usage. The question is considered a major linguistic problem because this particle has equivalences in the rest of the Romance languages, in spite of their differences. In particular, we will discuss the way in which *se* can be analysed from a corpus-driven approach for lexicographical use, and for this purpose we chose CPA (Hanks, 2004a) as a model of analysis. Our aim is to raise the question of whether CPA is an appropriate model to deal with pronominal verbs and if it is useful to represent them in a dictionary of Spanish as a foreign language. Finally, we also formulate a lexicographical proposal to represent this kind of constructions in the verb entries of a learner's dictionary.

Using Google Books Unigrams to Improve the Update of Large Monolingual Reference Dictionaries

Alexander Geyken & Lothar Lemnitzer

Keywords: *practical lexicography, computational linguistics, corpus statistics, lemma list.*

This paper describes ongoing work to extend a traditional dictionary using a large opportunistic corpus in combination with a unigram list from the Google Books project. This approach was applied to German with the following resources: the *Wörterbuch der Deutschen Gegenwartssprache* (WDG, 1961-1977), the German unigram-list of Google Books and the DWDS-E corpus. Both corpus resources were normalized. The subsequent analysis shows that the normalized unigram list has clear

complementary information to offer with respect to DWDS-E and that a comparatively small amount of manual work is sufficient to detect a fairly large number of new and relevant dictionary entry candidates.

A co-occurrence taxonomy from a general language corpus

Rogelio Nazar & Irene Renau

Keywords: *asymmetric word association*, *computational lexicography*, *co-occurrence statistics*, *distributional semantics*, *taxonomy extraction*.

This paper presents a quantitative approach to the generation of a taxonomy of general language. The methodology is based on statistics of word co-occurrence and it exploits the fact that word association is asymmetrical in nature, in much the same way as hyperonymy relations are. Words tend to be syntagmatically associated with their hyperonyms, though this is not true the other way round. Taking advantage of this phenomenon, and with the help of directed graphs of word co-occurrence, we were able to collect hyperonym-hyponym pairs using a reference corpus of general language as the only source of information, i.e., without using lexico-syntactic patterns nor any kind of pre-existing semantic resources such as dictionaries, ontologies or thesauri. The results obtained by using this method are not precise enough to be used for immediate practical purposes, but they confirm the hypothesis that as a general rule hyperonymy is linked to asymmetric co-occurrence relations. The paper discusses an experiment in Spanish, but we believe the same conclusions apply to other languages as well.

Lexicographic potential of corpus-derived equivalents: The case of English phrasal verbs and their Polish equivalents

Magdalena Perdek

Keywords: *phrasal verbs*, *equivalence*, *parallel corpora*.

The aim of this paper is to investigate Polish equivalents of English phrasal verbs as found in an English-Polish (E-P) parallel corpus *PHRAVERB*. Given the semantic idiosyncrasy exhibited by phrasal verbs, it is assumed that the equivalents generated by *PHRAVERB* will often differ from those found in E-P dictionaries. The qualitative corpus analysis aims to show that arriving at the desirable Polish counterpart involves a detailed semantic breakdown of the English structure, a careful analysis of the context in which it is used, as well as linguistic and translation skills, necessary to detect the nuances and subtleties of meaning in both languages. *PHRAVERB* is used to analyze the lexicographic potential (LP) of corpus equivalents. Four levels of LP have been established – high, average, low and zero – to evaluate which corpus-derived equivalents are eligible for inclusion in E-P dictionaries. To this end, 2,514 occurrences of PVs in the parallel corpus, with their equivalents, have been identified and analyzed.

METRICC: Harnessing Comparable Corpora for Multilingual Lexicon Development

Araceli Alonso, Helena Blancafort, Clément de Groc, Chrystel Million & Geoffrey Williams

Keywords: *comparable corpora, focused web crawler, collocational networks, multilingual dictionaries, Cultural Heritage lexicon*.

Research on comparable corpora has grown in recent years bringing about the possibility of developing multilingual lexicons through the exploitation of comparable corpora to create corpus-driven multilingual dictionaries. To date, this issue has not been widely addressed. This paper focuses on the use of the mechanism of collocational networks proposed by Williams (1998) for exploiting comparable corpora. The paper first provides a description of the METRICC project, which is aimed at the automatically creation of comparable corpora and describes one of the crawlers developed for comparable corpora building, and then discusses the power of collocational networks for multilingual corpus-driven dictionary development.

Lexicography and language technology

Corpus-based lexicography: an initial step for designing a bilingual glossary of lexical units in English and in Spanish

Juan-Pedro Rica-Peromingo

Keywords: *lexicography, corpus-based, phraseology, bilingual glossary, lexical units.*

Lexicography is basically concerned with the meaning and use of words. In previous decades, lexicographers have investigated the meanings of words and synonyms, but recent lexicographic research has been extended using corpus-based techniques to study the way that words are used and, in particular, how lexical associations are used. Lexicography is, therefore, directly connected to phraseology because both disciplines study sets of fixed expressions (idioms, phrasal verbs, etc.) and other types of multi-word lexical units. This paper presents an overview of two major corpora (CEUNF and COEPROF) compiled for phraseological and lexicographical purposes: the use of lexical bundles in the writing of Spanish university students. Both the CEUNF and the COEPROF have been used to analyze the production of phraseological units (lexical bundles and grammatical collocations) present in argumentative texts written in English by Spanish EFL university students. This study, based on corpus linguistics (McEnery, Xiao & Tono, 2006), phraseology (Cowie, 1998; Howarth, 1996, 1998; McCarthy & O'Dell, 2005, Nesselhauf, 2003, 2005; Granger & Meunier, 2008) and lexicography (Atkins & Rundell, 2008; Bergenholtz et al., 2009; Hartmann, 2001, 2003; Nielsen, 2009: Ooi, 1998), uses two taxonomies taken from Biber et al. (1999) for the lexical bundles (linking and stance lexical bundles) and Benson et al. (1986, 1993) for the grammatical collocations (verbs of communication and mental processes). With these two taxonomies a bilingual list of phraseological units in Spanish and English will be devised in order to contrastively analyze the production of such units by both non-native students and professionals writing in English and with the ultimate goal of designing a lexicographical glossary of bilingual lexical units used in argumentative English writing. For the preliminary quantitative analysis of the data and word searching Wordsmith Tools (Wordlist and Collocates tools) has been used. The analysis of these initial data and the use of the appropriate statistical tools (norming of words, T-test for the statistical significance, etc.) may be seen as a starting point for producing a glossary of lexical items in argumentative writing and improved teaching material for Spanish university learners of English.

The lexicographic working environment in theory and practice

Andrea Abel & Annette Klosa

Keywords: *dictionary writing system, lexicographic working environment, dictionary software.*

The changes caused by the growing automatisation of processes in the lexicographer´s workstation and in lexicographic work, together with the ensuing needs of lexicographers and their demands for adequately targeted software, have not been discussed sufficiently in meta-lexicographic research. The aim of this paper is therefore to fill this gap, with a focus on academic non-commercial lexicography. After an introduction into the general functionalities of specific dictionary writing software, with the help of a real-life example we will discuss the lexicographic working environment, the new specific demands to lexicographic software as well as different tools. The final aim is to propose some recommendations for how to structure the lexicographic working environment to meet specific project requirements.

Online English Dictionaries: Friend or Foe?

Gao Yongwei

Keywords: *online dictionary, e-lexicography, English-Chinese lexicography*.

The emergence of online English dictionaries in the past two decades has not only changed the lookup habit of many people and but also influenced the way dictionaries are compiled and presented. The traditional role played by paper dictionaries has been challenged, as witness the sharp decrease of the sales of the so-called "dead-tree" dictionaries and the steady diminishing in their readership. In consequence, many paper dictionaries have been gathering dust on bookshelves in bookstores, libraries or private studies. The ever-increasing popularity of online dictionaries has even made some alarmists suggest the possible demise of paper dictionaries. However, the future of dictionary-making and that of bilingual lexicography in particular is not as dismal as what people usually think. The lexicographical information presented in online dictionaries may prove to be a bonanza for bilingual lexicographers. This paper attempts to research into the major online English dictionaries that are available today, and their advantages and disadvantages will also be discussed. The scene of online English-Chinese dictionaries will also be investigated, and opportunities presented to English-Chinese dictionary-makers in the digital era will be explored.

The Syntax-Semantics Interface of Czech Verbs in the Valency Lexicon

Václava Kettnerová, Markéta Lopatková & Eduard Bejček

Keywords: *valency, lexicon, alternations*.

In this paper, alternation based model of the valency lexicon of Czech verbs, *VALLEX*, is described. Two types of alternations (changes in valency frames of verbs) are distinguished on the basis of used linguistic means: (i) grammaticalized alternations and

(ii) lexicalized alternations. Both grammaticalized and lexicalized alternations are either conversive, or non-conversive. While grammaticalized alternations relate different surface syntactic structures of a single lexical unit of a verb, lexicalized alternations relate separate lexical units. For the purpose of the representation of alternations, we divide the lexicon into data and rule components. In the data part, each lexical unit is characterized by a single valency frame and by applicable alternations. In the rule part, two types of rules are contained: (i) syntactic rules describing grammaticalized alternations and (ii) general rules determining changes in the linking of situational participants with valency complementations typical of lexicalized alternations.

ISA-overload in NorNet

Julie Matilde Torjusen

Keywords: *wordnets*, *ISA-overload*, *hyponymy*, *paranymy*, *orthogonal hyponymy*.

NorNet is a wordnet constructed on the basis of a traditional monolingual dictionary, still undergoing development. The wordnet does for instance still contain many cases of ISA-overload. In this paper I will show examples of ISA-overload in the semantic fields of persons and animals, and see if the relation of paranymy (Huang, Hsiao et al. 2008) or orthogonal hyponymy (Pedersen et al. 2009) are possible ways of solving the ISA-overload from these examples.

Semi-Automatic Analysis of Dictionary Glosses

Rune Lain Knudsen

Keywords: *wordnet*, *semantic networks*, *computational linguistics*, *bioinformatics*.

Automatic methods for information retrieval, knowledge engineering/representation and text classification are important tools for processing large amounts of natural language. Lexicographic databases are being used as part of the toolset for some of these methodologies. At the Department of Linguistic and Scandinavian Studies (UiO), a Norwegian wordnet (NorNet) is being developed by applying a thorough analysis of the semantic parts of the definitions contained in Bokmålsordboka (BOB) in order to generate a network of semantic relations. In addition to the development of a wordnet using this dictionary-based method, the analysis stage of the process is valuable in itself as it can be used to give new insights into the consistency of the source material and gloss structure in general. An overview of the analysis stage is presented in this paper. The analysis is limited to verb definitions for the time being, and should be regarded as a work in progress.

On automated semantic and syntactic annotation of texts for lexicographic purposes

Vladimir Selegey

Keywords: *automated linguistic annotation, syntax and semantic analysis, corpus-based lexicography, Internet as a corpus*.

The main idea of this paper is that automatic annotation is the only means to secure an efficient access to the whole set of linguistic productions rather than merely a small subset of such productions annotated manually.

Why is it necessary for a lexicographer to turn to open unannotated corpora? There are two valid concurrent reasons for that: the ever-growing rate of linguistic changes, on the one hand, and, on the other, the regional, social and professional 'segmentation' of the language, requiring a differential approach to the language phenomena under analysis.

For the past 10 years or so, the line of research based on the 'Internet as a Corpus' approach has seen booming growth. As far as technologies are concerned, the means of access available to the researcher are much more modest in this case. The methods currently used for indexing the World Wide Web by search engines are based on principles that are far from being linguistic. In spite of the fact that there are projects like Semantic Web, the Internet remains so far a raw text corpus with rather unreliable data about the frequency of occurrence.

We are presenting ongoing project ABBYY Syntactic and Semantic Parser that offers technologies for the automated linguistic annotation of text corpora. These technologies make a seamless addition to the technologies for the production of representative sub-corpora relating to the major Internet segments. ABBYY Syntactic and Semantic Parser (SSP) is built on linguistic technologies developed within the scope of the ABBYY Compreno project. It is planned to be part of LingvoPro portal (http://lingvopro.abbyyonline.com/en). Compreno is a multi-language (at the moment English, Russian, German, Spain, French, Chinese) ongoing NLP project based on the combination of sophisticated linguistic modeling and modern methods of language structure analysis (recognition). It is a scalable linguistic technology to use at a basic level for a range of NLP applications. As far as lexicography is concerned, the most important feature of this system is that automatic linguistic annotation is derived from a thorough syntactic and semantic analysis of a sentence.

Multilingual lexicography

Lexical enrichment of bilingual dictionaries with a focus on conversion as a word-formation process

Enn Veldi

Keywords: *bilingual dictionaries, lexical enrichment, conversion, English, Estonian.*

The paper focuses on the treatment of noun-to-verb conversion in English-Estonian and Estonian dictionaries. Because conversion is highly productive in English, it may pose some difficulty for compilers of bilingual dictionaries. It is argued that there is considerable room for lexical enrichment of bilingual dictionaries with regard to both inclusion of conversion verbs and the choice of translation equivalents. From the perspective of Estonian one has to take into account the possibility that an English converse verb could be rendered by means of conversion, suffixation, or a multi-word equivalent. The established equivalents can be used for the enhancement of symmetry between English – language X and language X – English dictionaries.

An Online Dictionary Browser for Automatically Generated Bilingual Dictionaries

Enikő Héja & Dávid Takács

Keywords: *parallel corpus, proto-dictionary, dictionary query system, semantic relations.*

The objective of this paper is to demonstrate that corpus-driven bilingual dictionaries generated fully by automatic means are suitable for human use.

Previous experiments have proven that bilingual resources can be created by applying word alignment on parallel corpora and such resources are useful for bilingual dictionary compilation purposes. Moreover, the corpus-driven nature of the method yields several advantages over more traditional approaches. Most importantly, the exploitation of parallel corpora decreases the reliance on human intuition during dictionary building. However, the proposed technique has to face some difficulties, as well. First, the scarce availability of parallel texts for medium density languages imposes limitations on the size of the resulting dictionary. Secondly, the resulting bilingual resource is not completely clean: that is, wrong translation candidates are also included in the dictionary. In fact, there is a tight correlation between the proportion of wrong candidates and the size of the resulting resource.

Our objective is to design and implement a dictionary a query system that is apt to exploit the additional benefits of the dictionary building method and overcome the disadvantages of it.

Lexicography and semantic theory

Basics of a comprehensive semantic categorization of Dutch verbs

Frans Heyvaert

Keywords: *semantic categorisation*, *verb meaning*, *definition structure*.

Based on a definition analysis project carried out at the INL, this paper puts forward a proposal for a semantic categorisation of Dutch verbs in which existing category systems like the ones in Framenet and in Levin (1993) are integrated with some findings that emerged from the project. A multilayered category structure is proposed in which data about Aktionsart, conceptual field and systematic semantic analysis are all made explicit. This type of categorization is to be used in a Dutch dictionary project, the *Algemeen Nederlands Woordenboek* (ANW). It is meant as a tool to make common verb definitions in the dictionary more uniform and systematic, but also as a service to scholarly dictionary users whom it will enable to extract easily and systematically bodies of semantically related data from the dictionary.

Emotion verbs in Greek. From Lexicon-Grammar to multi-purpose syntactic and semantic lexica

Voula Giouli & Aggeliki Fotopoulou

Keywords: *emotion verbs*, *Lexicon-Grammar tables*, *syntactic structure*, *distributional properties*, *semantic classification*.

We hereby present work aimed at giving an account of Greek verbs denoting emotion that is placed within a larger context, aimed towards defining and describing the semantic field of emotions by means of identifying, selecting, classifying and organizing a *core* lexicon of emotions in a conceptual Data Base. The ultimate goal is the exhaustive description of Modern Greek and the development of a wide-coverage lexical resource that will be appropriate for a range of Natural Language Processing Applications.

Cognitive lexicography of emotion terms

Carolin Ostermann

Keywords: *lexicography and semantic theory, English monolingual learner lexicography, cognitive linguistics, new lexicographic approach, semantics of emotion terms, user-study*.

At a glance, lexicography and cognitive linguistics are two branches of linguistics that do not seem to have a lot in common. While the lexicography of English on the one hand has followed established principles for decades or even centuries, cognitive linguistics on the other hand only emerged a few decades ago. But since the systematic description of the language is the basis for lexicography, linguistics also has a significant influence on the latter (cf. Béjoint 2010). I furthermore argue that it would be especially beneficial to use cognitive linguistics as a new basis for lexicography, - leading to something called 'cognitive lexicography' - since this new branch of linguistics tries to explain how humans perceive and conceptualise the world and has provided the basis for an entire new conception of semantics. A description of language in dictionaries based on cognitive linguistics would therefore be more realistic (cf. Geeraerts 2007) and more tangible. This is demonstrated here for emotion terms, which are generally hard to define. Emotion terms have received a fair amount of treatment in literature (cf. Kövecses 2000), but dictionary definitions of emotion terms are usually vague and circular. For this class of abstract nouns, a new lexicographic defining format has been developed which is not only based on traditional principles of lexicography, but also on cognitive linguistic semantic information concerning emotion terms, for example the prototypical emotion scenario and metaphors and metonymies (cf. Kövecses 2000). Definitions of the nine basic emotions terms anger, disgust, hate, fear, sadness, desire, love, happiness and joy written in this new format were scrutinised in a user study whereby test subjects had to name the correct term for a given definition. It has been demonstrated that definitions following this new cognitive linguistic defining scheme yield significantly better results compared to traditional dictionary definitions.

Étude contrastive de la lexicographie synonymique distinctive en France et en Europe aux XVIII^{e} et XIX^{e} siècles

Alice Ferrara-Léturgie

Keywords: ***l****exicography*, *synonymy*, *distinction*, *diachrony*.

This study aims at comparing both French and European dictionaries of synonyms of the XVIIIth and XIXth centuries. Girard wrote the very first dictionary of distinctive synonymy in French in 1718. His dictionary was the very first of this kind in Europe. It is only after Girard's dictionary, which introduced the methodology of the distinction of synonyms, that other dictionary writers in Italy, Spain, Great Britain, Germany or Russia have made dictionaries akin to Girard's. Thus, Girard's part into the growth of

dictionaries of synonyms across Europe is a major issue. By using Spanish and Italian dictionaries of synonyms, we will show that Girard was actually the model for all dictionaries of synonyms writers. However, the aim of this study is not to demonstrate that Girard is a model, but rather that all European synonymists started to consider and theorize synonymy in the same way as one single person, in other words by using distinction between synonym words. After translating French synonymists, European synonymists began to write their own dictionaries of distinctive synonymy.

Translation equivalents in translation corpora and bilingual dictionaries: the case of approximators in English and French

Diane Goossens

Keywords: *translation equivalents, bilingual dictionaries, translation corpora, quantity approximation*.

This paper reports on an investigation of the translations of 'approximator + number' combinations (e.g. *about 200*) in English and French using a translation bidirectional parallel corpus of news reporting and examining the entries for the approximators selected in three bilingual dictionaries. This study analyses the major tendencies that are found in the corpus when translating six commonly used approximators occurring around numbers into French and into English: the approximators *plus de*, *près de*, *environ*, *dépasser*, *quelque* and approximators formed using a number and the suffix *–aine* are analysed for the French to English translation direction, and the approximators *more than*, *about*, *up to*, *around*, *over* and *some* are investigated for the English to French translation direction. The entries for these approximators are also scrutinized in three bilingual dictionaries: the *Harrap's Unabridged, Le Grand Robert et Collins électronique* and the *Grand Dictionnaire Hachette Oxford*. The paper focuses more specifically on the types of examples given for each approximator, examining whether the quantity approximation meaning of these items is well represented. Several bidirectional translation equivalence issues are also discussed as some translation equivalents mentioned in one direction are not listed in the other direction in the same dictionary. Based on corpus evidence, the study suggests several ways of improving the treatment of the items under study in bilingual dictionaries. These include the introduction of labels that would inform the user about the context in which the approximator is preferred, for instance in informal contexts or with certain types of quantities. The variety of items identified in the corpus may also help lexicographers list translation equivalents from a wider variety of grammatical categories.

Optimizing semantic granularity for NLP - report on a lexicographic experiment

Silvie Cinková, Martin Holub & Vincent Kríž

Keywords: *corpus pattern analysis*, *semantic tagging*, *semantic granularity*, *English*, *verbs*.

Experiments with semantic annotation based on the Corpus pattern Analysis and the lexical resource PDEV (Hanks and Pustejovsky, 2005), revealed a need of an evaluation measure that would identify the optimum relation between the semantic granularity of the semantic categories in the description of a verb and the reliability of the annotation expressed by the interannotator agreement (IAA). We have introduced the Reliable Information Gain (RG), which computes this relation for each tag selected by the annotators and relates it to the entry as a whole, suggesting merges of unreliable tags whenever it would increase the information gain of the entire tagset (the number of semantic categories in an entry). The merges suggested in our 19-verb sample correspond with common sense. One of the possible applications of this measure is quality management of the entries in a lexical resource.

On the Nature of Signposts

Janet DeCesaris

Keywords: *signposts*, *pedagogical lexicography*, *English*, *Spanish*.

Dictionary entries for highly polysemous words have long proved difficult for lexicographers and dictionary users alike. From the lexicographer's point of view, senses and possibly subsenses need to be identified, and tough decisions must be made about the order of senses within the entry. From the user's standpoint, long entries require a certain amount of time and patience, because users must often wade through large amounts of information before finding the answer to their initial query. In response to this, lexicographers working on English monolingual learner's dictionaries have introduced "access facilitating devices" Lew's (2010), also known as pointers, guide words or signposts, to help users disambiguate and thus find information more quickly. This paper addresses the nature of signposts: what sort of information do they convey, and what semantic relationship do they have with the headword? In our paper, we will analyze several entries for nouns and adjectives in four learner's dictionaries of English (CALD, LDOCE, MEDAL and OALD) and discuss the differences across dictionaries. Our analysis shows a preference for synonyms, as opposed to superordinates or contextual information, in the English dictionaries analyzed. We then show how signposts are being used in the DAELE, an ongoing project of a learner's dictionary of Spanish.

Terminology, LSP and lexicography

Visual Analytics and the Language of Web Query Logs - A Terminology Perspective

Daniela Oelke, Ann-Marie Eklund, Svetoslav Marinov & Dimitrios Kokkinakis

Keywords: *co-occurrence analysis, web search log, visual analytics, medical terminology*.

This paper explores means to integrate natural language processing methods for terminology and entity identification in medical web session logs with visual analytics techniques. The aim of the study is to examine whether the vocabulary used in queries posted to a Swedish regional health web site can be assessed in a way that will enable a terminologist or medical data analysts to instantly identify new term candidates and their relations based on significant co-occurrence patterns. We provide an example application in order to illustrate how the co-occurrence relationships between medical and general entities occurring in such logs can be visualized, accessed and explored. To enable a visual exploration of the generated co-occurrence graphs, we employ a general purpose social network analysis tool, visone (http://visone.info), that permits to visualize and analyze various types of graph structures. Our examples show that visual analytics based on co-occurrence analysis provides insights into the use of layman language in relation to established (professional) terminologies, which may help terminologists decide which terms to include in future terminologies. Increased understanding of the used querying language is also of interest in the context of public health web sites. The query results should reflect the intentions of the information seekers, who may express themselves in layman language that differs from the one used on the available web sites provided by medical professionals.

Terminology of Higher Education: Towards International Harmonization

Vera Budykina

Keywords: *dictionary compilation, terminology of higher education, international harmonization of terminology, dictionary of higher education, dimensions of terminology development*.

The paper describes the evolution of dictionary of education since the first special dictionary of this kind was compiled in 1945. With the spread of globalization and the Bologna process the problem of harmonization of terminology is up-to-date. In

particular, the terminology used by teachers, students, and educators in European higher educational institutions is of great interest to those who are not native speakers of English.

This paper describes the new project of compiling an English-Russian Dictionary of Higher Education. The paper also highlights the results obtained from experiments, which have proven that educational terminology in English is difficult to understand due to the differences in educational systems. To fill this void, and compile the bilingual dictionary of higher education it was necessary to identify three dimensions of terminology development: the cognitive, linguistic and communicative. The last part of the paper describes the methodology based on the three dimensions and the tools of the project, which is aimed at harmonizing the terminology used within higher education on an international level, and the compilation of bi- and multilingual dictionaries of higher education, which are few at the moment.

The project will benefit further developments of the European higher education sphere, creating a mutually beneficial cooperation between countries and stimulating collaborative university partnerships. It will also favor both international understanding and multiculturalism and hopefully contribute not only to enrichment of lexicography but science, research, and technology.

Inclusion of verbal syntagmatic patterns in specialized dictionaries: the case of EcoLexicon

Beatriz Sánchez Cárdenas & Miriam Buendía Castro

Keywords: *terminology*, *specialised knowledge representation*, *verbal lexicon*, *syntagmatic patterns*.

One of the main drawbacks of specialized lexicographical resources is the lack of combinatorial patterns in word descriptions. Various authors have highlighted the need to include verbs in specialized lexicographic resources (William 2010; L'Homme & Leroyer 2009; López-Ferrero & Torner Castells 2008; Alonso Campos & Torner Castells 2008). In this sense, apart from some initiatives (Williams 2008; Williams & Millon 2010 inter alia), verbs have not yet deserved enough attention in terminographic resources. In this research we aim to evaluate how verbs should be ideally described in dictionaries for specific purposes. To this end, we first analyze how the existing specialized resources deal with phraseology and word combination. Based on their main advantages and shortcomings, we present here a new proposal for verb description in EcoLexicon, a specialized knowledge base of environmental sciences. Accordingly, a fine-grained description of the macrostructure and microstructure of a verb entry is provided, based on the main tenets of Frame-Based Terminology (Fillmore 1985, 2006; Faber 2009, 2011, 2012), Role and Reference Grammar (Van Valin 2005) and the Lexical Grammar Model (Faber & Mairal 1999, Ruiz de Mendoza & Mairal 2008). The terminological entry proposed accounts for the combinatorial patterns of terms and verbs and, therefore, is thought to be very useful for translators who are due to produce texts in the target language in the same way natives would do.

Developing Academic Word Lists for Swedish, Norwegian and Danish – a joint research project

Sofie Johansson Kokkinakis, Emma Sköldberg, Birgit Henriksen, Kari Kinn & Janne Bondi Johannessen

Keywords: *academic word list, Nordic languages, higher education, language learning and teaching*.

This paper reports on a joint multi-disciplinary Nordic project aimed at developing three new academic lexical resources based on corpora consisting of texts from Swedish, Norwegian and Danish academic settings. An academic word list exists for English, but no such lists exist for the Nordic languages. Such a list would be an important resource for both L1 and L2 students in their first years of study, a period when many students struggle to cope with the demands of academia. Moreover, the word lists would be of use to students and teachers at the higher levels of secondary education. An inventory of academic words and phrases would also be a useful tool for researchers of academic language use and for test developers. The paper outlines the initial stages of work on an academic word list for Swedish. Three potential research approaches have been explored: the translation of the English list, extracting academic words from existing corpora, and the compilation of parallel academic corpora where an academic word list is extracted from these. The paper will discuss the advantages and drawbacks of the different approaches and the benefits of carrying out a joint project involving several languages. The question of entry selection and the information categories of the dictionary entries and the interplay between the entries in the dictionaries and the corpora will also be briefly addressed.

Exemples de lexicographie juridique à orientation pédagogique en France: le *Vocabulaire du juriste débutant* et le *Guide du langage juridique*

Chiara Preite

Keywords: *legal lexicography, pedagogical specialized lexicography, knowledge-orientated function, communication-orientated function*.

Pedagogical specialised lexicography has only recently developed as an independent field of study (Fuertes-Olivera / Arribas-Baño 2008, Fuertes-Olivera 2010). Most research is carried out within the broader framework of Bergenholtz et al.'s (1995, 2003, 2006) Modern Theory of Dictionary Functions, which distinguishes between knowledge-orientated dictionary functions – focusing on the user's need for cultural and encyclopaedic information – and communication-orientated functions – which address communication, translation and production needs. With Fuertes-Olivera and Arribas-Baño (2008: 139), we claim that pedagogical specialised dictionaries should address

both functions. While traditional specialised dictionary meets knowledge-orientated needs, we shall shift the focus to communication orientated functions. To this purpose, we shall identify and describe the communication orientated items present in Bissardon's Guide du langage juridique and Lerat's Vocabulaire du juriste débutant, mainly along the lines of Groffier and Reed's (1990) work on the microstructure of legal dictionaries.

Multidimensional Categorization in Terminological Definitions

Pilar León Araúz & Antonio San Martín

Keywords: *multidimensionality*, *terminological definition*, *EcoLexicon*, *recontextualization*, *contextual variation*.

EcoLexicon (http://ecolexicon.ugr.es) is a terminological knowledge base on the environment that currently holds 3,351 concepts and a total of 17,475 terms in English, Spanish, German, Russian, French, and Modern Greek. Concepts are linked by means of hierarchical and non-hierarchical relations in dynamic networks and in definitions. The environmental domain is interdisciplinary and its concepts can be categorized from different perspectives, thus conceptual representation needs to be multidimensional. Although, unlike other knowledge resources, conceptual representations in EcoLexicon reflect multidimensional categorization, this has also produced an information overload, particularly at upper concept levels. This means that many concepts show overloaded networks partly caused by multiple inheritance, as many of them have several hyperonyms. However, all conceptual dimensions do not occur at the same time but rather are context-dependent. Since the context of a concept is the set of concepts relevant to its intended meaning, we solved the information overload problem by recontextualizing networks in terms of discipline-based domains. The recontextualization of concepts constrains their relations with other concepts, depending on the activation scenario. By no means, does this imply that these are different senses of a polysemic term, but concepts also vary by context regardless of sense variation. Given that terminological definitions are also an integral part of the representation of multidimensionality, we applied the same contextual constraints to definitional propositions. The result is what we call flexible terminological definitions. This paper describes the representation of context-dependent multidimensionality in EcoLexicon and, more specifically, how this phenomenon is managed in terminological definitions.

Term candidate extraction for terminography and CAT: an overview of TTC

Ulrich Heid & Anita Gojun

Keywords: *terminology extraction*, *computer-assisted translation*, *term alignment*.

In this paper, we present a tool chain for terminology extraction and term alignment which is under development in the EU-project TTC. The tool components comprise the

crawling of domain-specific text from the internet, in different languages, linguistic pre-processing of the corpus collected in this way, and the extraction of term candidates. Extracted term candidates of two languages are aligned into pairs of source and target term equivalents. This output can be used both in interactive translation setups (e.g. computer-aided translation) and in machine translation.

The communicative situation as frontier between words and constituents of terminological variants

Paula de Santiago

Keywords: *terminology*, *lexicography*, *communicative situation*, *corpus linguistics*, *denominative variants*.

The article describes the importance of the analysis of language in use. In this respect, it has been appreciated that many of the sweeping differences between lexicography and terminology are seen as conflicting ideas in contrast with the descriptive theories of terminology. In this study, it is believed that the limits between these disciplines become blurred when we take into account pragmatic and discursive criteria. On the basis of a corpus composed of popularized scientific articles, attention will be paid to the identification, with more or less difficulty, of terminological variants in a certain communicative situation.

The purpose of our study is to support the status of terms whenever they are used in a specialized communicative situation, considering that the participants can have different degrees of knowledge. In addition, it will be shown that the terminology of a particular subject field is never completely fixed due to the range of discourses where it can appear; as a consequence, it is proposed to use genre restrictions when including variants of an original term in a dictionary.

Juristische Kollokationen in norwegischen Arbeitsverträgen

Julia Pujsza

Schlüsselwörter: *Juristische Fachsprache*, *Juristische Kollokationen*, *Korpusarbeit*, *Norwegische Arbeitsverträge*.

This article describes the research of legal collocations in the corpus of Norwegian employment contracts. It is characteristic for the legal language, as a language for specific purposes, that terms and collocations have a special meaning, different from the general language usage. In this article the problem of the definition of collocations is described as well as the differences between collocations in a language for specific purposes and a language for general purposes. The aim of the research was to compose the list of the legal collocations and the possible classification of them in the Norwegian employment contracts. Some examples of them are given and explained in the article.

The results could be a starting point for other lexicographic works. The study of the collocations in Norwegian employment contracts could be useful for lawyers, interpreters or even laymen who have contacts with this type of text in their everyday life.

Dictionary use, pedagogical lexicography

A Comparison Between COBUILD, LDOCE5 and CALD3: Efficacy and Effectiveness of the Dictionaries for Language Comprehension and Production

Alice Yin Wa Chan

Keywords: *language comprehension and production, dictionary use, comparisons of monolingual dictionaries*.

This paper reports on the results of a research study which compared the effectiveness of different monolingual dictionaries for language comprehension and production by advanced Cantonese ESL learners in Hong Kong. A group of 31 students majoring in English participated in the study. This included a meaning determination task which required students to use a dictionary to determine the meanings of nine familiar words used in unfamiliar contexts, a sentence completion task which required students to use a dictionary to complete ten English sentences based on some given Chinese contexts, as well as a sentence construction task which required students to use a dictionary to construct ten English sentences using some given English prompts. Different monolingual dictionaries were used in the tasks by different sub-groups of participants, namely *Collins COBUILD Advanced Dictionary 6th edition (COBUILD6)/ Collins COBUILD Learner's Dictionary Concise Edition (COBUILD Concise)*, *Longman Dictionary of Contemporary English 5th edition (LDOCE5)*, and *Cambridge Advanced Learner's Dictionary 3rd edition (CALD3).* The accuracy rates at which the participants performed the tasks were calculated, and their perception of the usefulness of the dictionaries was collected. It was found that monolingual dictionaries are effectiveness not just for language comprehension but also for language production, yet successful dictionary consultation does not depend on the dictionary being used. Learners' dictionary skills and their abilities to extract relevant information from a dictionary are more important than the choice of dictionaries.

'To Teach Little Boys And Girls What It Is Proper For Them To Know': Gendered Education and the Nineteenth-Century Children's Dictionary

Sarah Hoem Iversen

Keywords: *children's dictionaries, education, gender, nineteenth century, Britain*.

This paper explores the role nineteenth-century children's dictionaries in the gendered education of children. Children's dictionaries have been widely regarded as mid-twentieth-century phenomena. Pre-twentieth-century lexicography, meanwhile, has been

traditionally regarded as an exclusively male pursuit. Contrary to these assumptions there were, in fact, many dictionaries specifically written for children in the eighteenth and nineteenth centuries. Several of these were compiled by women who drew on their experience as educators. Children's dictionaries in this period aimed, not simply to impart the meaning of words, but also to provide a social and moral education. This moral didacticism can be seen to form part of an ongoing construction of gender identities for children in this time. As lexicographer Anna Murphy put it in her 1813 A First, Or Mother's Dictionary for Children, to educate was 'To teach little boys and girls what it is proper for them to know'. Through dictionary definitions, illustrative examples, and pictorial illustrations, girls and boys were constructed in different ways, and as exhibiting different virtues (or vices). Although this paper focuses mainly on dictionaries compiled by female lexicographers, and the ways in which these works addressed female readers, dictionaries compiled by men are also considered for comparative purposes. Similarly, though the discussion centres on constructions of the prototypical 'good girl', the 'good boy' is also considered, especially since these prototypes were often seen to define each other by antithesis. The extent to which individual lexicographers' personal and political positions came into play is significant and could lead to ideological patterns deviating from dominant gender ideologies; some female compilers, for instance, actively contested some of the limitations placed on feminine identity.

Are dictionaries of lexical blends efficient Learners' dictionaries?

Arnaud Léturgie

Keywords: *lexical blending, learners' dictionaries*.

The scope of this paper concerns both learners' lexicography and lexical blending. It will focus on the potential utilization of dictionaries of fanciful lexical blends as learning tools, able to ensure an educational role in the learning of the lexicon. In order to deal with this issue, the phenomenon of blending in regard to a learning perspective will be briefly introduced. This will allow an exploration of the ways of using dictionaries of blends in a didactic manner, and at the same time evaluate the limitation of this method.

General Monolingual Persian Dictionaries and Their Users: A Case Study

Saghar Sharifi

Keywords: *lexicography, Persian, general monolingual dictionaries, user*.

User needs and user satisfaction have unfortunately been neglected in the compilation of Persian dictionaries. This article aims to investigate five general monolingual Persian

dictionaries in terms of their meeting user needs and the extent of user satisfaction with them. The investigated dictionaries are *Dehkhoda*, *Mo'een*, *Amid*, *Farhange Farsie Emrooz,* and *Sokhan*. To assess user needs, different groups of users, based on Assi (1995), filled up questionnaires, and some were interviewed; some statistical procedures, such as the chi-square significance test, were used. The objectives of this study were to identify the users' reference needs and the relationship between these needs and social variables. Moreover, the extent of the users' satisfaction with the mentioned dictionaries, the relation of this satisfaction to the social variables, and the necessity of certain qualifications in users were assessed. It was found that the users' educational background was the only determining factor in their amount of dictionary use, in their finding the desired information, and in their satisfaction with the dictionary.

Using FrameNet in Communicative Language Teaching

Karin Friberg Heppin & Håkan Friberg

Keywords: *FrameNet*, *communicative language teaching*, *language learning*, *frame semantics*.

This article describes how a lexical database such as FrameNet in its different language versions can be used for communicative language teaching, an approach which focuses on communicative rather than grammatical competence. Using the semantic frames of FrameNet to illustrate situations on which to base teaching can bring about a natural flow in the organisation of teaching materials, in syllabus construction, and in the planning of individual lessons. FrameNet can also support language students in learning to communicate in different situations. The frames can guide them in choosing lexical units and sentence patterns.

A golden mean? Compromises between quantity of information and user-friendliness in the bidirectional Norwegian-Lithuanian Dictionary

Aurelija Griškevičienė & Sturla Berg-Olsen

Keywords: *bilingual lexicography*, *user-friendliness*, *bidirectional dictionaries*, *Norwegian*, *Lithuanian*.

This paper explores the concept of user-friendliness in the context of bidirectional bilingual dictionaries, presenting and discussing some of the choices taken by the editors of the Norwegian-Lithuanian Dictionary (NLD). The NLD is a medium-sized paper dictionary compiled by a joint group of lexicographers from the Universities of Vilnius and Oslo. The dictionary is intended both for native speakers of Norwegian and of Lithuanian. Designing a user-friendly bidirectional dictionary necessarily involves

making compromises between the needs of different target groups. User-friendliness in lexicography is a problematic concept, because a feature that enhances the user-friendliness of a dictionary for one group of users often reduces it correspondingly for other groups. This is especially acute in the case of bidirectional dictionaries. The amount of information given and the degree of linguistic precision must be balanced against the danger of information overload. Thus, designing the structure of a dictionary is largely a matter of seeking compromises between quantity of information, precision and user-friendliness. The paper shows concrete examples of how the editors of the NLD have tried to maintain this balance. Many elements in the NLD are based on another bilingual dictionary (Berkov et al. 2003), but the system for information on the target language, Lithuanian, is designed by the editors of the NLD. The paper shows the steps taken to make the dictionary user-friendly from two angles: 1) adapting and improving the lemma list and information on the source language and 2) designing the system for providing information on the target language. In this context it also discusses problems arising from the wish to re-use data from one bilingual dictionary when compiling another dictionary with a different target language.

Online dictionaries – how do users find them and what do they do once they have?

Henrik Lorentzen & Liisa Theilgaard

Keywords: *online dictionaries, search strategies, query log analysis, information retrieval, user behaviour, user survey*.

In general, user behaviour studies on online dictionaries have focused on user behaviour once the user is on the site. But before a potential user even reaches this stage, he or she must succeed in finding the dictionary on the web. In this paper we investigate users' linguistic search strategies before they enter our dictionary site, *ordnet.dk*. What kind of search engine queries are successful and why (not)? Similarly, we have studied the site search queries. Are the search strategies the same? Taking the no-match searches as a starting point, we have asked ourselves if our content and search functionality correspond to the search behaviour of the users, that is if we can give an answer to the users' queries and if data is organized and presented in an appropriate way. Given the results of these analyses, we decided to make several changes to the site in order to optimize user access and attract new users. These changes and their ensuing results are presented. Furthermore, we present and discuss the results of a user survey conducted in October-November 2011.

Usability testing as a tool for e-dictionary design: collocations as a case in point

Ulrich Heid & Jan Timo Zimmermann

Keywords: *electronic dictionaries, usability testing, collocations, access to lexicographic data*.

We report about the application of usability tests to electronic dictionaries; our examples concern the design of dictionary interfaces that allow the user to access lexicographic data about collocations. We thus first summarize options for collocation retrieval, in terms of search criteria and types of data displayed as search results. We then present usability testing methods in general, as well as their application to electronic dictionaries, and we report about two tests, one with existing e-dictionaries, the other with custom-built mock-ups. We interpret this work as a first step towards usability design of electronic dictionaries: we suggest that new concepts for e-dictionary interfaces could be developed by rapid prototyping and tested with users before being integrated into dictionary products.

A study of pupils' understanding of the morphological information in the Norwegian electronic dictionary *Bokmålsordboka* and *Nynorskordboka*

Kjersti Wictorsen Kola

Keywords: *dictionary use, morphological information, electronic dictionaries*.

Do 15-and-16-year-old pupils understand the morphological information in the Norwegian electronic dictionary Bokmålsordboka and Nynorskordboka? That is the question addressed in this study. The informants were given grammatical exercises which they were supposed to answer by making use of the morphological information in the dictionary. The information consisted partly of codes and example words and partly of inflectional suffixes and full inflectional forms. According to the results, the former is easier to understand than the latter, but altogether, the information seems to be difficult to understand. This result suggests a need for changes in the way the morphological information is presented in the dictionary.

Collocations, phraseology and idioms

The presentation of set phrases and collocations in bilingual dictionaries with focus on an Icelandic-French dictionary

Rosa Elin Davidsdottir

Keywords: *bilingual dictionaries, collocations, foreign language learning, Icelandic-French lexicography*.

This paper presents a PhD thesis whose aim is to analyse the methodological concepts pertaining to the composition of bilingual dictionaries with a focus on the language pair Icelandic and French and example entries for an Icelandic-French dictionary. Collocations, for example *se brosser les dents* ('to brush one's teeth') and set phrases such as *Il pleut des cordes* ('It's raining cats and dogs') are important for the language learner but are often neglected in bilingual dictionaries despite various linguists having pointed out the importance of taking them into account in lexicography. Therefore, special attention will be paid to the presentation of set phrases and collocations in a bilingual dictionary destined to help with encoding from a mother tongue to a foreign language (an L1→L2 dictionary). In the thesis, it will be examined how bilingual dictionaries can give more information on set phrases and collocations in the target language and thus be a better tool for the language learner. Propositions will be exemplified with selected entries for an Icelandic-French dictionary with explanations and scientific argumentation for the choices made. We set out to establish a model for an Icelandic-French electronic dictionary that will be as detailed as possible, in terms of examples, and focused on collocations and set phrases.

The thesis is a contribution to research in the field of bilingual lexicography and aims to contribute to the making of bilingual dictionaries in general, regardless of the languages in question. It is hoped that the outcome will also serve as a foundation for a new Icelandic-French dictionary as the need for a new one to meet the expectations of users in the 21st century has become considerable.

Towards more and better phrasal entries in bilingual dictionaries

Sylviane Granger & Marie-Aude Lefer

Keywords: *n-grams, lexical bundles, English, French, phrasal entries, phraseology*.

Although the phraseological coverage of dictionaries has improved considerably in recent years, bilingual dictionaries are still lagging behind. The objective of our paper is to show that including a range of multi-word units (MWUs) extracted via the n-gram method can considerably enhance the quality of English<>French bilingual dictionaries.

We show how multiword units extracted from monolingual corpora can enhance the phraseological coverage of bilingual dictionaries and suggest ways in which the presentation of these units can be improved. We also focus on the role of translation corpora to enhance the accuracy and diversity of MWU translations in bilingual dictionaries.

Finding Multiwords of More Than Two Words

Adam Kilgarriff, Pavel Rychlý, Vojtěch Kovář & Vít Baisa

Keywords: *collocations, multiword expressions, multiwords, corpus lexicography, word sketches*.

The prospects for automatically identifying two-word multiwords in corpora have been explored in depth, and there are now well-established methods in widespread use. (We use 'multiwords' to include collocations, colligations, idioms and set phrases etc.) But many multiwords are of more than two words and research for items of three and more words has been less successful.

We present three complementary strategies, all implemented and available in the Sketch Engine. The first, 'multiword sketches', starts from the word sketch for a word and lets a user click on a collocate to see the third words that go with the node and collocate. In the word sketch for take, one collocate is care. We can click on that to find ensure, avoid: take care to ensure, take care to avoid.

The second, 'commonest match', will find these full expressions, including the to. We look at all the examples of a collocation (represented as a pair/triple of lemmas plus grammatical relation(s)) and find the commonest forms and order of the lemmas, plus any other words typically found in that same collocation. For baby and bathwater we find throw the baby out with the bathwater.

The third, 'multi level tokenization', allows intelligent handling of items like in front of, which are, arguably, best treated as a single token, so lets us find its collocates: mirror, camera, crowd.

While the methods have been tested and exemplified with English, we believe they will work well for many languages.

Yesterday's idioms today: a corpus linguistic analysis of Bible idioms

Laura Pinnavaia

Keywords: *idioms, corpus linguistics, pragmatic meaning*.

Many of the idioms used in English stem from the Bible. There they were originally coined and used to announce God's word, to facilitate the understanding of it, and to capture the ineffable and unsaid. Nowadays with newly derived and synchronic

meanings, they can be employed in a similar fashion in contexts that are not just religious. It is the simultaneous existence of the two metaphoric readings – the historic and the synchronic – that makes Bible idioms particularly rich and fascinating linguistic tools worthy of study. This article analyses a series of twenty-five Bible idioms in contemporary English, as represented by the *British National Corpus*. While the examination provides data as to the frequency and distribution of the idioms in different texts, particular attention is placed upon their communicative functions in discourse in order to try and individuate three pragmatic types of Bible idiom.

Idioms beyond their dictionary borders: how figurative meaning functions in texts

Ekaterina Lukyanova

Keywords: *idioms, figurative meaning, literal meaning, text, experiential domains*.

This paper discusses semantic approaches to idiomatic meaning and their implications for lexicographic practice, focusing on idioms, whose meaning is described as 'figurative' and is generally thought to be non-compositional. It is argued that figurative meaning is a function of so-called 'literal' meaning, and can only exist on the basis of compositional semantic structures. Idioms are approached as expressions that employ culturally prominent source domain scenarios in a figurative way with the purpose of projecting a clear evaluation onto a complex target situation. This hypothesis is supported by an analysis of how two idioms - 'carry the ball' and 'carry the can' - function in a number of texts.

Creating a phraseme matrix based on a Tertium Comparationis

Cerstin Mahlow

Keywords: *tertium comparationis, meta-index, phrasemes*.

Diachronic exploration of linguistic resources like collections and dictionaries from different time periods allows researchers to get first impressions on language change and define specific research questions to investigate further, for example by integrating empirical data. However, manual inspection of large collections is exhausting and error prone. Automatic extraction and comparison of the keywords of dictionary entries from several dictionaries can be used to create a combined index, allowing to easily access respective dictionary entries to extract related information. As a case in point we consider information on German phrasemes in dictionaries and collections from the 18th to the 21st century. We use a concept-driven semi-automatic approach to create a matrix based on a Tertium Comparationis to allow users to easily look up related phrasemes.

Creating a bilingual learner's dictionary of Italian and German collocations: strategies and methods for searching, selecting and representing collocations on the basis of a learner-oriented, semantic-conceptual approach.

Erica Autelli, Christine Konecny & Martina Bradl-Albrich

Keywords: *collocations*, *didactics*, *learner's dictionary*, *second language learning*, *semantics*.

Collocations are commonly used expressions which, from the point of view of a narrow conception based primarily on semantic-conceptual and learner-oriented criteria, can be defined as semi-fixed word combinations situated on the continuum between free combinations and idioms. While collocations are seen as an entirely 'normal' phenomenon and intuitively used correctly by native speakers, for second language learners they can be very tricky because they often vary in different languages, especially due to the different 'conceptualisations' used by the speaking communities, that is the different cognitive approaches to actual situations of the extralinguistic reality. A learner of Italian, for instance, needs to know that in this language a drawn number or lot is literally 'fished' (pescare un numero / un biglietto), that if classes in school have been cancelled, the lessons are literally 'jumping' (le lezioni saltano), or that a free phone number is called a 'green number' (numero verde).

As far as Italian linguistics and lexicography is concerned, collocations have only recently become a focus of interest and thus no specific collocational dictionary for L2 learners exists yet. Hence, our aim is to create a bilingual (Italian-German) learner's dictionary of collocations, connecting our lexicographic approach to didactic and semantic research. One of the innovative aspects of our dictionary is that we will insert various drawings made by pupils in order to visualise the conceptualisations of Italian collocations and to facilitate in this way the process of learning and remembering them. The dictionary is mainly aimed at German speakers wanting to learn Italian, but it can also be used the other way round (Italian-German). Its target groups are primarily L1 German and Italian pupils, but it will be equally useful for students, translators and interpreters as well as for German and Italian speakers in general who are learning the other language. The collocations listed in the dictionary will belong to four specific morphosyntactic categories, namely "subject + verb", "verb + direct object", "verb + prepositional phrase" and "noun + adjective or prepositional phrase".

In our paper we will illustrate which strategies and methods we use to find and select our data. Moreover, we will show on the basis of which criteria we decide what word combinations are to be classified as collocations and thus to be included in our dictionary. Finally, we will provide the sample entry of the lemma "dente" ('tooth').

Idioms as a Microstructural Component: A History of Bilingual Italian-English Dictionaries (1749-2009)

Chris Mulhall

Keywords: *idioms*, *microstructure*, *bilingual Italian-English dictionaries*, *phraseology*.

The purpose of this paper is to look at the evolution of microstructural design in bilingual Italian-English Dictionaries, with particular emphasis on the positioning on idioms, from the period 1749-2009. Idioms, which can be described as phraseological units whose overall meaning is greater than the sum of their individual semantic parts, pose a variety of difficulties for lexicographers. Probably the greatest challenge comes in the form of lemmatisation, which requires a lexicographer to choose a suitable headword under which to insert an idiom. An equally important consideration is their positioning within the entry as this can enhance or impinge on the dictionary user's ability to access the desired information. Although the past 150 years have witnessed an evolution in the design of entries in Bilingual Italian-English dictionaries, some reference works in this category remain deficient and inconsistent in their methods of recording and positioning idioms. This paper charts the development of the microstructure component of bilingual Italian-English dictionaries since 1749 and details their diverse approach to dealing with idioms, while also trying to reconcile their unique semantic and lexical features.

Lemma selection

Treatment of Complex Prepositions in Czech and English Dictionaries

Renata Novotná

Keywords: *complex prepositions*, *representative corpora*, *monolingual and bilingual dictionaries*.

The paper studies complex prepositions, such as *within the bounds of*, *on the basis of*, their frequency in the 100-million corpora (Czech corpora SYN2000, SYN2005 and SYN2010, English corpus BNC) and their treatment in dictionaries. In the Czech monolingual dictionary Dictionary of Literary Czech (Slovník spisovné češtiny) the complex prepositions are stated under the lemmas of abstract nouns, such as *hledisko* (viewpoint) - *z hlediska* (from the point of view of), in the Great Czech-English Dictionary (Velký česko-anglický slovník) by J. Fronek states as prepositions only part of them, the rest of complex prepositions is given in collocations, such as *v rámci zákona* - *within the bounds of law*. In the Collins COBUILD English Dictionary the prepositions *out of* and *according to* are stated in separate entry, while the rest of the prepositions are stated within another entry. In New Oxford Dictionary of English most of the complex prepositions are stated within the phrases given at the end of each entry, the exceptions are prepositions *according to* and *rather than*. The author proposes to state all the complex prepositions as separate entries, i. e. on the same level as single or one-word prepositions.

Lexicographic interchange between a specialized and a general language dictionary

Marie-Claude Demers, Ilan Kernerman & Marie-Claude L'Homme

Keywords: *general language dictionary*, *terminological database*, *specialized meaning*, *term*, *wordlist*.

One of the important issues lexicographers need to address concerns the desired coverage of a dictionary's wordlist. This paper addresses the issue from a practical angle. We propose a method for comparing the contents of two resources and evaluating to what extent each can contribute to increase and improve the coverage of the other. Concretely, the project consists of comparing the contents of the English version of DiCoInfo (a dictionary of computing and Internet terms) with the appropriate entries of the *Random House Webster's College Dictionary* (RHWCD). The entries missing in one resource are considered for inclusion in the other, and vice versa. The approach proves

beneficial for both resources. Approximately 100 entries were added to DiCoInfo and over 500 lexical items or meanings are being included in the RHWCD.

Reports on lexicographical projects

From paper dictionary to an elaborate electronic lexicographical database

Mark Van Mol

Keywords: *Arabic database driven lexicography, Arabic tagset development, online dictionaries.*

At the 2000 Euralex conference we presented a paper on the development of a new learner's dictionary for Modern Standard Arabic, based on a corpus linguistic approach. In 2001 this dictionary was published in two volumes: a Dutch-Arabic volume and an Arabic-Dutch one. After the publication of the two dictionaries, we started new projects to work on both the existing corpus on which the dictionary was based (at that time 3,000,000 words) and the internal extension of the lexicographical database. We did not limit ourselves to additional lexical information and expressions, but included very detailed grammatical information. In recent years, the evolution of language technology has led to increased possibilities for lexicographical exploration of databases, especially in Arabic. In this paper we present the elements that we added to the contents of the lexicographical database: new words and expressions, 646 detailed POS tags, the technological changes it underwent (for example, the transformation from 4th Dimension (4D) to Mysql). This resulted this year in the development of the first full online consultable Arabic-Dutch/Dutch-Arabic dictionary. This Arabic dictionary is the first of its kind, not limiting itself to mere lexical information, but allowing a much greater variety of searches for all kinds of grammatical information. In this paper we offer an overview of some of the possible searches. One of the next challenges is the production of an online dictionary with a clear layout in order not to be forced to skip much of its detail and accuracy.

The Comprehensive English-Georgian Online Dictionary: Methods, Principles, Modern Technologies

Tina Margalitadze

Keywords: *English-Georgian dictionary, structure content, software*.

The aim of the paper is to present the lexicographic project completed in I. Javakhishvili Tbilisi State University, namely the Comprehensive English-Georgian Online Dictionary.

Conceived back in the 1960s of the previous century at the Chair of English Philology of the University, the dictionary project has gone through many difficult stages: erroneous decisions about principles of compilation of dictionary entries,

incorrect sources chosen for the dictionary, lack of experience of lexicographic work at an educational institution, no financing, etc. In the 1980s a small team of editors embarked on thorough revision of the dictionary material and started publication of the dictionary in fascicles, on a letter-by-letter basis. The online version of the dictionary, posted in the Internet in 2010, is based on the mentioned fascicles.

The paper discusses the macrostructure of the dictionary, considerations behind the selection of the word-list for the dictionary; principles of presentation of homonyms, converted forms, polysemy; exemplification policy, as illustrative phrases and sentences constitute a very important component of dictionary entries. The paper pays special attention to the treatment of semantic asymmetry between genetically unrelated and systemically completely different languages as is the case with the Georgian and English languages.

The paper elucidates grammatical, as well as other types of labels employed in the dictionary: temporal (archaic, obsolete), regional (American English, Australian English, etc); formal and informal, spoken words, sociolects and connoted vocabulary are also marked by respective labels (formal, informal, colloquial, vulgar, slang, derogatory, contemptuous, pejorative, etc); specialized terminology has subject-specific labels (anatomy, architecture, astronomy, biology, geography, geology, economics, medicine, metallurgy, philosophy, finance, technical, zoology, etc).

The Comprehensive English-Georgian Online Dictionary is a web-application developed in accordance with modern standards and requirements. The engine of the dictionary is written in PHP scripting language. The dictionary vocabulary and systemic bases are located in MySQL database. Interfaces use some JavaScript. The web-application comprises user, administration and billing functions and interfaces, thus creating an integrated and dynamic resource which provides a unique opportunity to simultaneously use, maintain and administer the dictionary.

False Italianisms in British and American English: A Meta-Lexicographic Analysis

Cristiano Furiassi

Keywords: *false Italianisms*, *metalexicography*, *English dictionaries*.

Inspired by the existing literature on Italianisms, this work aims to investigate the presence of selected false Italianisms (or pseudo-Italianisms), that is *alfresco*, *bimbo*, *bologna*, *bravura*, *confetti*, *dildo*, *gondola*, *gonzo*, *inferno*, *latte*, *pepperoni*, *politico*, *presto*, *stiletto*, *studio*, *tutti-frutti*, and *vendetta*, in the English language through a meta-lexicographic analysis of the *OED* and the *Merriam-Webster*, authoritative dictionaries considered to be representative of British English and American English respectively. False Italianisms – which most English speakers believe to be purely Italian – are created when genuine lexical borrowings from Italian are so reinterpreted by a recipient language, English in this case, that native speakers of Italian would not recognize them as part of their own lexical inventory and would neither understand nor use. The creation of false Italianisms yields to new insights into the covert prestige attributed to the supposed donor language and culture.

Setting up for Corpus Lexicography

Adam Kilgarriff, Jan Pomikalek, Miloš Jakubíček & Pete Whitelock

Keywords: *corpora*, *corpus lexicography*, *web crawling*, *dependency parsing*.

There are many benefits to using corpora. In order to reap those rewards, how should someone who is setting up a dictionary project proceed? We describe a practical experience of such 'setting up' for a new Portuguese-English, English-Portuguese dictionary being written at Oxford University Press. We focus on the Portuguese side, as OUP did not have Portuguese resources prior to the project. We collected a very large (3.5 billion word) corpus from the web, including removing all unwanted material and duplicates. We then identified the best tools for Portuguese for lemmatizing and parsing, and undertook the very large task of parsing it. We then used the dependency parses, as output by the parser, to create word sketches (one page summaries of a word's grammatical and collocational behavior). We plan to customize an existing system for automatically identifying good candidate dictionary examples, to Portuguese, and add salient information about regional words to the word sketches. All of the data and associated support tools for lexicography are available to the lexicographer in the Sketch Engine corpus query system.

Digital and traditional resources for the second edition of the *Deutsches Wörterbuch*

Elke Gehweiler & Christiane Unger

Keywords: *diachronic lexicography*, *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm*, *digital resources*, *historical corpus*, *quotation archive*.

The paper gives a short overview of the selection of sources for the first and second editions of the Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm, from which the quotations for the dictionary entries are drawn. We will introduce the 2DWB quotation archive, which forms the basis of the lexicographical work on the second edition of the Deutsches Wörterbuch (2DWB) and which today is complemented by digital resources. We will assess a number of freely available digital collections of text according to their suitability for diachronic lexicography. We will look at size, selection of texts, verifiability of search results, quality of full texts and scans, presentation of search results and search functions. It will turn out that none of the resources can (yet) substitute 2DWB archive. We will further suggest that from the point of view of diachronic lexicography in some areas the examples from the "intelligent" quotation archive are superior to automatically retrieved examples from digital corpora.

Other topics

Word formation in electronic language resources: state of the art analysis and requirements for the future

Janina Radtke & Ulrich Heid

Keywords: *word formation, morphology, electronic dictionaries, user needs*.

We report on a state of the art survey on electronic lexical resources for word formation; these include online specialized dictionaries for interactive use, a grammar information system, as well as a few online tools for morphological analysis. Our comparison is inspired by the Function Theory of Lexicography (e.g. Tarp 2008), and by a definition of needs of users in different communicative situations. Our survey is part of plans towards electronic dictionaries for word formation, and we thus formulate requirements that such dictionaries should ideally fulfil.

Filling the gap: The Pharos semi-bilingual *English Dictionary for South Africa*

Wanda Smith-Muller & Ilan Kernerman

Keywords: *Afrikaans, bilingual, English, monolingual, semi-bilingual*.

For more than a decade Pharos Dictionaries specialised in the publication of bilingual Afrikaans-English dictionaries, and Afrikaans monolingual dictionaries. The need for an English learner's dictionary that distinguished itself from the existing ones on the South African market became a growing urgency. With its small editorial capacity Pharos had to work and plan strategically to maintain its competitiveness in a fierce local market. For this reason, it entered into an agreement with K Dictionaries whereby it could utilise the data from the Kernerman Semi-Bilingual Dictionaries series to compile a dictionary that was uniquely suited to the South African market. The end product did not only fill the gap on Pharos's backlist. It also distinguished itself from the existing monolingual English learners' dictionaries on the market through its semi-bilingual character, as well as its uniquely South African flavour on both micro- and macrostructural level.

Onomastic lexicography

Patrick Hanks & Richard Coates

Keywords: *surnames, family names, FaNUK, etymology, medieval evidence, geodemographic analysis, onomastic database*.

Modern scholarship and techniques of data analysis have shown that even the best current dictionaries of surnames in Britain are full of errors, oversights, guesswork, fudges, and omissions. In this paper we present a new project designed to rectify this situation. A database has been compiled containing entries for all the family names in Britain. Entries in this database for family names with more than 100 bearers — and for many less frequent names, too — are being systematically compared with data on medieval surnames and with a geodemographic analysis of the 1881 census. The associations between surnames and localities are explored systematically, with results that quite often have a profound effect on our understanding of the origins and etymology. The English, Celtic, French, and Scandinavian etymologies of native names are investigated using the best techniques of historical linguistic scholarship. The national identity of recent immigrant names is explained.

Foregrounding the Development of an Online Dictionary for Intermediate-level Learners of Brazilian Portuguese as an Additional Language: Initial Contributions

Tanara Zingano Kuhn

Keywords: *corpus research, defining vocabulary, dictionary for language learners, lexicography, Portuguese as an additional language*.

The present PhD project intends to collaborate with the designing of a monolingual online dictionary for intermediate-level learners of Brazilian Portuguese as an additional language. Considering that the development of such a reference work involves the investigation of a series of theoretical-methodological aspects, this research will be narrowed down to one specific issue: the use of simplified Portuguese language patterns in the writing of the definitions. Therefore, the steps to be taken entail a thorough bibliographical review on lexicographical definitions for monolingual learners' dictionaries and the use of defining vocabulary for their writing; Brazilian Portuguese corpus research in order to compile a defining vocabulary list (DVL); and tests with learners to verify which kind of definitions – those which were written with or without the use of DVL – is better for the user. Since pedagogical (meta)lexicography regarding Brazilian Portuguese as an Additional Language (BPAL) is to a fairly large degree still incipient, especially when compared to what has been done in the area of English as a

Foreign Language (EFL), this project is expected to give substantial contribution to new knowledge.

What belongs in a dictionary? The Example of Negation in Czech

Dominika Kovarikova, Lucie Chlumska & Vaclav Cvrcek

Keywords: *negation*, *lexicography*, *grammatical category*, *frequency*, *lemmatization*.

In this paper, the authors try to answer the basic lexicographical question: how do we know whether a particular word is a mere word form, or a new lexeme that should thus be assigned an individual entry in a dictionary? The issue of negation in Czech (namely negative forms of nouns, adjectives, adverbs and verbs) serves them as a perfect example. They introduce two criteria for the choice of dictionary entries, the frequency criterion and the grammatical category criterion, and show how the negation of the parts of speech examined differs and what the implications are for lexicographers.

Defining and Structuring Saussure's Terminology

Nilda Ruimy, Silvia Piccini & Emiliano Giovannetti

Keywords: *Saussure's terminology*, *computational lexicon*, *ontology*, *semantic relations*.

In the framework of the Italian project '*For a digital edition of Ferdinand de Saussure's manuscripts*', an electronic thesaurus of Saussure's terminology is being built, which includes new terms extracted from recently found manuscripts. The lexical model on which it is grounded is a customized version of the SIMPLE model. In this paper, an overview of the customization process is provided, with a special focus on the steps taken for designing a domain-specific ontology as well as on the creation of additional semantic relations and features. Lexical entries are illustrated and the potential of a structured organization of semantic knowledge for gaining a wider understanding of the overall domain terminology is highlighted.

In what sense is the OED the definitive record of the English language?

Pius ten Hacken

Keywords: *OED*, *language*, *usage notes*, *dictionaries of record*, *dictionary use*.

OED (2011) presents itself as "Oxford English Dictionary | The definitive record of the English language". Superficially, this claim may seem a marketing slogan, but Simpson's (2000) preface to the third edition shows that it is a reflection of the editors' understanding of their dictionary, what may be called their 'lexicographic ideology'. In this paper, I consider the claim from three perspectives. Section 1 presents the foundations of the claim as formulated in the preface. Section 2 analyses the claim with regard to some relevant insights gained in linguistic theory since work on the first edition of the OED started. Section 3 discusses some of the practical reflections of the ideology of recording as opposed to prescribing. Finally, section 4 formulates some general conclusions.

Using social media to find English lexical blends

Paul Cook

Keywords: *lexical blends*, *neologisms*, *computational lexicography*, *social media*, *Twitter*.

We present a method for identifying English lexical blends — words such as *complisult* (*compliment* + *insult*) and *globesity* (*global* + *obesity*) — from social media, specifically Twitter. Our method is based on observations about words and phrases that are commonly used to introduce new words and corpus patterns that are often used to describe the meaning of lexical blends, and leverages the massive volume of data that is readily-available for analysis through Twitter. We run our method for 5 weeks and identify 976 candidate lexical blends; analysis of a sample of these candidates indicates that approximately 57% are blends. We further discuss a small number of blends identified by our method that are in regular usage on Twitter but which are not recorded in any of a number of dictionaries searched.

Software demonstrations

Léacslann: A platform for building dictionary writing systems

Michal Boleslav Měchura

Keywords: *dictionary writing systems*, *terminology management systems*, *e-lexicography*, *databases*.

The purpose of this demo is to introduce *Léacslann*, a new platform for building dictionary writing systems (DWS) and terminology management systems (TMS) as well as other lexicographic and reference applications. Léacslann can be used without any knowledge of programming to create a basic lexical database with an arbitrary structure. This will be demonstrated in the first half of the demo, while the second half will show how a software developer can customize Léacslann for more demanding applications.

Gentyll English-Spanish non sexist on-line glossaries

José Simón

Keywords: *non-sexist professional titles*, *gender-aware glossaries*, *database online query system*.

The purpose of this paper is to introduce the Gentyll online glossaries: non-sexist bilingual English-Spanish glossaries of terms relating to human naming in various subject areas (agentives). Our team has being working for several years in the study of the penetration of non-sexist language policies in the language used in various sectors of social and professional activity. After close inspection of a good number of printed and online lexicographic and terminological resources we came to the conclusion that almost all of them ignore non-sexist language policies and recommendations, both in their structure and in their actual data. Being aware of this lack of gender-aware resources we have tackled the publication of a series of non-sexist glossaries which cover a number of fields of activity. We do not intend to devise neither a new theory nor a new methodology. Our objective, far more modest, is to unveil a new, gender-aware, perspective which, in our view, should inspire lexicography and terminology in the 21st century.

In this presentation we will start by summarising the principles underlying our glossaries together with the pre-requisites we bore in mind in their conception. Later we will detail the contents of the databases together with the methodology adopted in their compilation, to end with a sketchy enumeration of the main features of the query system we have developed in order to streamline online searches.

We firmly believe our glossaries are a contribution, modest indeed, to a new perspective that wishfully will inspire more ambitious works to come.

Compiling medical dictionaries and spellcheckers for the Dutch language

Arnoud van den Eerenbeemt

Keywords: *medical lexicography*, *innovation on methodology*, *automation of lexicographical tasks*.

This article outlines general aspects of medical monolingual lexicography in Dutch as based on the author's personal experience since 1995, converting the typesetting file of the Dutch *Pinkhof* monolingual medical dictionary into a dictionary database, editing the database since then, updating spelling and definitions and compiling specialised pocket dictionaries, spellcheckers and even medical dictates from the lexical content.

Colidioms: An Online Software for Phraseography and Paremiography

Elena Berthemet

Keywords: *collaborative*, *idiom*, *notion*, *semantics*, *translation*.

This paper investigates the possibility of building a multilingual phraseological database. It presents the framework of a privately-funded online project called Colidioms. The goal of the Colidioms project is to build a public collaborative database. The software is designed for the full perception and reproduction of phrasemes. Combining tradition and innovation, Colidioms is based on recent technological advances. It is a web application that supports English, French, German and Russian and enables multi-directional search of phraseological equivalents in any of these four languages. Two types of search are available: semasiological and onomasiological. The central organizing principle of the software is based on the concept of 'notions'. Notions allow to create a bridge between phrasemes in different languages. It has been demonstrated that notions make it possible to carry out cross-lingual comparisons. Notions link all parts of the database and homogenize the corpus and are compatible with all studied languages.

Cypriot Greek Lexicography: An online lexical database

Charalambos Themistocleous, Marianna Katsoyannou, Spyros Armosti & Kyriaki Christodoulou

Keywords: *web-service*, *Cypriot Greek*, *dialectal lexicography*, *text to speech*.

This article presents an online dictionary environment, with enhanced sorting and searching functionalities and a text to speech feature, for hearing the pronunciation of the

words. The online dictionary environment has been developed as part of the ‘Syntychies’ research program. ‘Syntychies’ online environment is a pioneering web-service for Greek dialectal lexicography and it is the first of its kind for Cypriot Greek.

Posters

The earliest days of Estonian lexicography

Madis Jürviste

Keywords: *historical lexicography, 17th century dictionaries, beginnings of Estonian lexicography*.

The first Estonian dictionaries were compiled by German pastors in the 17th century, at a time when bilingual lexicography was already rather widely spread in Western Europe. Why were these dictionaries compiled and for whom? What are the main characteristics of these dictionaries? In the article, an overview is given of the aspects relevant to historical lexicography of three authors: Heinrich Stahl and his *Anführung zu der Esthnischen Sprach* (1637), Johannes Gutslaff's *Observationes grammaticae circa linguam esthonicam* (1648), as well as Heinrich Göseken's *Manuductio ad Linguam Oesthonicam* (1660). These bilingual German-Estonian dictionaries were not independent works, but were published as appendixes of German- and Latin-based grammars for Estonian. Their authors were native speakers of German, outstanding members of the local clergy, and at the same time the first Estonian lexicographers. Regardless of the limited number of entries and several inconsistencies in presenting the information about target language equivalents, in addition to evident mistakes in the choice of certain equivalents, the importance of these works should not be underestimated: not only were these grammars and dictionaries the first such publications in the region, but they also helped to fix the orthographic standards of written Estonian. Even if the fact that current Estonian language has extensive German influences both in vocabulary and syntax is most probably not due to a direct impact of these grammars and dictionaries, these three works were influential in their own time (partially due to the importance of their authors in the local church hierarchy) and had a role in the development of the Estonian language.

The Swedish Dialect Dictionary – a presentation

Annika Karlholm & Eva Thelin

Keywords: *Swedish dialects, word selection, geographical distribution*.

The most recent dictionary covering all dialects of Swedish, was published in1867 by Johan Ernst Rietz. Hence, the need for a modern dialect dictionary is considerable and in 2003 preparations for the *Swedish Dialect Dictionary* (*Svenskt dialektlexikon* or *SDL*) were initiated at the Department of Dialectology and Folklore Research in Uppsala. The *SDL* will be directed to the general public and the overall aim, apart from providing information about dialect words, is to stimulate people's interest in dialects.

The *SDL* is to be published as a one-volume dictionary comprising app. 500 to 600 pages. It will be based on the dialect collections kept at the department, which comprise more than 7.3 million paper slips, each describing a single word from a single parish. The *SDL* will only include a small proportion of the dialect words in these collections and in the preparatory work, the key issue was therefore to establish inclusion policies. A very strict selection of words is essential and there was a need for clear guidelines in order to speed up the selection process and to prevent a purely subjective choice of words. Based on the presumed needs of our target audience, the most important aspects were found to be the degree of 'dialectness' and the geographical distribution of the words, the number of examples and the age of the source material.

Digital Dictionary Development for Torwali, a less-studied language: Process and Challenges

Inam Ullah, Gull Feroz, Mahwish Bano & Sarmad Hussain.

Keywords: *lexicography, endangered languages, language technology.*

Torwali is an endangered and less-studied language spoken in the north of Pakistan. Recently, the community celebrated publication of the first ever Torwali dictionary both in print and an online version. This paper discusses issues and challenges regarding lexicography of a previously non-written language; from data collection by the native speakers having no set goals and training or institutional support, to organization and presentation of the data for producing multiple versions of the dictionary. The first section describes the process of developing the database using the methods of wordlists and semantic domains. The proceeding sections describe the technical development of its printed and online versions in detail, and discuss orthographical, technical, computational and social concerns of the project. The paper concludes with recommendations for future dimensions of the present work and for similar projects with special consideration to lexicographical work on non written languages.

The Romanian-Latin-Hungarian-German Lexicon - The Lexicon of Buda (1825). Informatics Challenges for an Emended and On-Line Ready Edition

Daniel-Corneliu Leucuta, Bogdan Harhata, Lilla Marta Vremir & Maria Aldea

Keywords: *informatics challenges, multilingual, old lexicon.*

The Lexicon of Buda or *the Romanian-Latin-Hungarian-German Lexicon*, published in 1825 in Buda, is the first etymological and explicative, quadrilingual Romanian dictionary. The roughly 13,000 entries / 771 pages lexicon are an important cultural

heritage, representing the collective cultural memory of those times, offering a testimony in the life, circulation and evolution of many words. The aim of this paper is to present the informatics challenges in the creation of an emended and on-line ready edition of *the Lexicon of Buda*.

Lexical relations in dialects and place names: Landscape terms

Vilja Oja & Marja Kallasmaa

Keywords: *dialect words, place names, concept 'field', Estonian, Finnic*.

The occurrence of landscape terms in Estonian dialects and place names is compared. Material from cognate languages is also used. Analysing the areal distribution and function of appellative nouns in dialects vs. place names, their lexical differences and semantic relations are discussed. The terms *nurm, põld* and *väli* occur throughout the Estonian area and in several cognate languages. In North Estonian dialects and Northern Finnic languages *nurm* means 'grassland, meadow', while in South Estonian dialects and the Livonian and Votic languages it stands for 'field'. An analogous semantic boundary runs through toponyms. In the Islands and Western dialects the common generic term in field names is *põld*, while *väli* is used in the North Estonian dialect east of the area. The meanings differ across dialects. Transferred names and recent farm names taken from standard Estonian stand out from the local dialectal background. Sometimes, homonymy may cause semantic confusion.

Extracting and Annotating Extended Lexical Units of Culinary Terms for Korean Culinary Manuscripts of *Joseon* Period

Kil-Im Nam, Hyeon-Ju Song, Jun Choi & Young-Hee Hyun

Keywords: *ELUs (Extended Lexical Units), culinary terms, terminological lexicography, semantic annotation*.

This is a follow up study of the previously reported project conducted from 2007 to 2009. In the previous project, a corpus of culinary manuscripts was constructed with rich morphological and semantic annotations. However, the morpheme based annotation was not sufficient for extracting traditional culinary terms since many terms are in the form of so-called 'extended lexical units (ELUs).' To tackle the limitations of the original annotations, This research attempted to apply phrase level semantic annotation. By extracting ELUs of culinary terms, firstly, richer information of the expressions could be obtained. Secondly, more accurate annotation has been achieved in the current research. Lastly, the products attained from this study can be applied to compile domain-specific dictionaries (in this case, culinary domain) and contribute to extend lemma status to multi-word items.

Old and New User Study Methods Combined – Linking Web Questionnaires with Log Files from the *Swedish Lexin Dictionary*

Ann-Kristin Hult

Keywords: *dictionary use*, *web questionnaire survey*, *log file analysis*.

The *Swedish Lexin Dictionary*, SLD, is an online Swedish monolingual learner's dictionary for immigrants at beginner's level (http://lexin.nada.kth.se/lexin/). The dictionary has a well-defined target group and an explicit purpose. Thus the intended use is very clear. Also worth noting is that the dictionary is frequently used. For these reasons, it is most interesting to examine the actual use of this dictionary. The SLD has recently been revised and the online search functions have been improved. The initial part of the paper briefly describes the SLD and the updating project. The main part of the paper reports on an ongoing study of the use and users of the SLD. The study has combined two methods: web questionnaire survey and log file analysis. Thanks to the linking between the questionnaire data and the log file data, issues concerning, for example, whether users really do what they say they do can be verified with greater certainty. The paper demonstrates an example of analysis, where the questionnaire answers of one user are compared with the same user´s actual searches in the SLD. This analysis is but one example of many hundred possible analyses. Apart from the results of the user study, it will also be of great interest to evaluate the procedure of combining two methods within the same study in this way, as the combination has a chance to yield more reliable and valid results on dictionary users and user behaviour.

A Dictionary of Spoken Danish

Carsten Hansen & Martin H. Hansen

Keywords: *lexicography*, *speech corpus*, *pragmatics*, *conversation analysis*.

The purpose of this project is to establish a dictionary of spoken Danish, titled Ordbog over Dansk Talesprog (ODT). Through the use of extensive empirical data, it is the aim of the project to convey the latest knowledge of spoken language to the broad public. ODT combines existing and new research based primarily on qualitative methods with the quantitative analysis of a corpus of spoken language. The result of this combined method will be made available to the public through the development of a web-based dictionary of spoken Danish.

ODT is a project of the Centre for Language Change in Real Time (LANCHART) at the University of Copenhagen. Building on a large corpus of spoken language consisting primarily of sociolinguistic interviews, recorded from 1978 – 2010 and consisting of almost 7 million transcribed tokens, we are working on a dictionary portal. We inscribe the project into a tradition of significant national dictionaries, namely the Dictionary of the Danish Language (1918 – 1956) and The Danish Dictionary (2003

– 2005). Both were published by the Society for Danish Language and Literature, which is one of our foremost institutional cooperating partners along with the Danish Language Council.

The ODT project pursues two spheres of action. One lets the editors conduct research of their own, both in the field of spoken-language research in line with the other activities at the LANCHART Centre, and in the new field of spoken-language lexicography. In this way the editors, future dissertation writers, and Ph.D. students working on the project will produce new knowledge. The other sphere of action concerns conveying this knowledge to the public. We see it as our job not only to promote and expose the research activities of the editors themselves and the other LANCHART researchers, but also to pass on knowledge and research on spoken language gained outside of the Centre.

The user segment of ODT consists of two groups. The primary recipient is the linguistically curious layperson interested in spoken language; the secondary recipient is the research oriented user. Both groups will benefit from a web portal which allows fast access, is segmentally differentiated (i.e., relevant), has a high level of service, is free of advertising, and is free to use.

ODT is designed as a web-based dictionary portal with a possibility for parallel comparable searches in a corpus of written Danish (KorpusDK) and in a dictionary mainly based on written Danish (The Danish Dictionary).

Theoretical work on ODT consists in elaborating on well-established lexicographic methods and exploring the possibilities for transferring them into a dictionary of spoken language. The practical work consists of actual dictionary compilation: searching, editing, storing, and presenting the corpus data.

The BRO-project, a bridge in the wild, Norwegian linguistic landscape

Ruth Vatvedt Fjeld & Petter Henriksen

Keywords: *national dictionary*, *xml*, *collaboration*.

The Norwegian language falls into two main variants - bokmål and nynorsk. The majority's variant is bokmål, used by over 90 % of the population. Historically, bokmål again falls into several sub-variants, but now the two main sub-variants- riksmål and bokmål proper - are practically united in one common norm. This norm is being documented in the national dictionary project bearing the symbolically significant name BRO ('bridge'). The article presents the background for the BRO collaboration, and sketches a concrete and feasible plan for the lexicographical documentation of the common norm. A challenge lies in the choice of lemma sign form and the presentation of bokmål's wide variety of optional forms, where also style nuances play a role. The same applies to the choice of examples and collocations and other multi-word lemmas. Both challenges arise from the need for freedom of expression within the norm, which is typical of Norwegians' preference to mark identity through language.

Ease of access in the new Afrikaans–Nederlands/ Nederlands–Afrikaans dictionary (2011) in the Dutch L2 classroom – a case study

Nerina Bosman

Keywords: *dictionary use, ease of access, bilingual Afrikaans Dutch dictionary, amalgamated lemma list, empirical observation*.

In 2011 a new bilingual Dutch Afrikaans dictionary, popularly known by the acronym ANNA, was published. The dictionary stands out mainly because of its unique macrostructure - it has one amalgamated lemma list.

Ease of access has become an important criterion for a user-friendly dictionary which must meet certain information needs. The aim of this research is to ascertain to what extent users found the access process in ANNA easy and satisfying while completing a task set to them in the L2 (Dutch) classroom. Questions that the research will attempt to find an answer to are:

- What was the search word / expression?
- What was the search time?
- What was the search route that was followed?
- How many search steps were necessary to find the information?
- Was the task completed within a time acceptable to the user?

This paper will report on an empirical observation of dictionary use. The participants (5 – 8 Afrikaans students in the Dutch class) will be asked to produce five Dutch lexical items in a vocabulary test. The look-up behaviour of the respondents while they complete the task (one at a time) will be directly observed and monitored. The Think Aloud Protocol (TAP) will be used; that is, the students will be asked to vocalise their own mental processes by "thinking out loud" during the search process. An audio recorder will be used and the researcher will also make notes.

For the analysis use will be made of the terminology proposed by Bergenholtz & Gouws (2010) such as search route, search step and search time.

A Swedish Academic Word List: Methods and Data

Håkan Jansson, Sofie Johansson Kokkinakis, Judy Ribeck & Emma Sköldberg

Keywords: *Swedish language, language learning and teaching, academic vocabulary, corpus-based dictionary, corpus compilation*.

Academic language often presents a challenge to students, both language learners and native speakers. Therefore there is a need for educational language tools such as academic vocabulary resources. To date, resources developed have mainly focussed on learners of English; similar support is not yet available for Swedish. This paper reports

on three different approaches to compiling a corpus of authentic academic text material used in academic settings. The purpose is to compose an empirical basis for the construction of a Swedish academic word list which can be used in language teaching. Because we have chosen to follow the method used for the creation of *The Academic Word List* (Coxhead 2000), the corpus content is crucial to the final content of our word list.

The principles behind the drafting of the Onomatopoeic Dictionary of the Lithuanian Language

Jolanta Zabarskaitė

Keywords: *iconicity*, *onomatopoeia*, *dictionary*.

The Lithuanian language has preserved a vast layer of iconic lexis. This type of lexis is interesting from the perspective of onomasiologic definition, pragmatics, and the history of language. Of all the language parts, it functionally covers onomatopoeic words. Elements of iconism are also typical of some words from other parts of language, as defined by the level of pragmatism, such as emotives and expressives, some interjections (and invocations in particular), as the emotional – expressive element embedded in their semantic structure is also the basis of their existence in the language. Iconic nature can also be a definitive feature of words of child-speak and any other lexical periphery: riddle words, some of the euphemisms, refrains, etc. Word formation can be iconic as well. The phonetic structure of an iconic word has references of phonetic or phonosemantic motivation that can be described linguistically, actualised language sounds and sound complexes with articulative and acoustic properties. Such properties, when transformed from the psychophysiological audible stimulus to phonologically described acoustic and articulative properties of phonemes are one of the key principles of describing the iconic lexis in lexicographic resources.

In language, expressive words serve the function of conveying the impression or emotion of the speaker/writer so that the target can experience/feel it too. M. Grammont has indicated that the ability of language sounds to give connotation to the meaning of a word is often potential and that it emerges in the process of the act of speaking (with the exception of the "pure" onomatopoeias that have a phonetic motivation). For instance, the connotational qualities of sounds of language make the narrator choose a member of the phonosemantic opposition (or triad) of synonymous onomatopoeias – *kaukšt*: *taukšt*: *paukšt*; *bumpt* : *bliumpt*; *kapt* : *knapt*; *pliaukšt* : *paukšt*; *čiaukšt* : *taukšt*, etc. – to be able to disclose the specifics and details of the image, sound or sense/experience being described (imitated) better. However, the choice of pragmatic situations is unlimited and therefore, when presented on its basis, the lexical meaning might be very inaccurate. Thus, the drafting of a lexicographic inventory has to begin with identifying the type of descriptive imitatives, i.e. the impression (visual, acoustic, sensual/experiential) they carry. Describing the semantic system of specific imitative requires the identification of certain tools, i.e. the phonic (or formative) instruments that are used to create the correlation of meaning and expression.

Instead of employing the conventional classification of onomatopoeias, in the Onomatopoeic Dictionary such words are categorised by the method of imitation, forming a total of four groups: onomatopoeias (construed as only those onomatopoeic words that imitate real-life sounds using linguistic tools, turning them into words), imitatives, mimemes and verbal onomatopoeic words (which are not considered iconic words). The idea here is to demonstrate the versatility of their iconic features and the variety of pragmatics, and therefore a systematic approach to presentation under the behaviourist scheme has been adopted. For instance, onomatopoeic words of a punch are presented systematically, and their lexicological articles are broken down against other attributes, i.e. a punch to a soft/hard surface or a vertical/horizontal punch and so on. The unique phonetic structure of onomatopoeic words is considered, describing some of the features, like frequent replication and consonantal variations of onomatopoeic endings, like *plept*:*plep*:*ple*, etc. as universal in the inventory. The systematic relationships are presented in several languages, which makes the Onomatopoeic Dictionary more accessible for linguists and semiotic scientists from other countries. The type of the lexicographic work being presented is an ideographical dictionary.

Encyclopedic Dictionary as a Crossroad between Place Names and Antroponyms: A Project of a New Type

Olga Karpova

Keywords: *culture*, *dictionary*, *encyclopedic*, *Florence*, *megastructure*, *macrostructure*, *microstructure*.

The paper is devoted to the description of the encyclopedic dictionary project *Florence in the Works of World Famous People: A Dictionary for Guides and Tourists* supported by Italian Cultural Foundation Romualdo Del Bianco. Main steps of the dictionary making process are carefully analysed as well as mega-, macro- and microstructure of the reference book based on *the Genius of the Place* principle. The paper is focused on outstanding foreigners with special reference to writers, artists, musicians and other public figures who worked and lived in Florence in different historical periods since the XVth c. up to the present day that have become the object of the Dictionary. Special attention is given to dictionary microstructure including four reference sections: *Biography*, *Creative Work, Florentine Influence*, and *Learn More*. The model of the Dictionary is supposed to become a sample for future reference books describing famous visitors to other cultural cities: London, Moscow, Paris, Oslo, etc.

Dictionary Use and Language Games: Getting to Know the Dictionary as Part of the Game

Tanneke Schoonheim, Carole Tiberius, Jan Niestadt & Rob Tempelaars

Keywords: *dictionary use*, *language game*, *log files*.

Most electronic dictionaries promise dynamic, proactive search via multiple criteria and via diverse access routes, but, often, they do not realise their full potential and their search options are still limited to the traditional search from word to meaning. The ANW (Algemeen Nederlands Woordenboek) - a free online scholarly dictionary of contemporary standard Dutch, which is currently being compiled at the Instituut voor Nederlandse Lexicologie (INL) - is different. It offers a range of search strategies, helping the user both with encoding and decoding tasks.

In December 2009, a demo version of the dictionary was launched. The dictionary is updated on a regular basis with an average of 500 to 750 new entries each time. An analysis of the log files shows that since its launch the average use of the dictionary is fairly stable, except for November 2010, when it almost tripled as a result of a language game, Het Verloren Woord ('The Lost Word') that INL launched. During a period of 6 weeks, participants received every week one or more cryptic descriptions or instructions in order to find the 'lost' word. Each description and/or instruction gave part of the word away and after solving all cryptic descriptions, the lost word, could be found in the ANW. The game attracted almost 2,000 players, who for several weeks explored the ANW thoroughly, using all the search facilities that are offered.

We will discuss the effect of this language game on the use of the ANW dictionary. In addition, we will show how a language game can play an educational role in familiarising users with the new possibilities that online dictionaries offer.

Finding Proverbs in the Venda Dictionary: Tshivenḓa - English

Munzhedzi James Mafela

Keywords: *dictionary*, *proverb*, *headword*, *illustrative example*, *user's style guide*.

Since Tshivenḓa was reduced to writing by the Berlin Missionaries in 1872, a dictionary of proverbs has yet to be produced. Only now has the Tshivenḓa National Lexicography Unit begun working on such a dictionary. The only dictionary, although not specifically of proverbs, that has included these in its definition of headwords is the Venḓa Dictionary: Tshivenḓa – English. The proverbs provided in this dictionary have been included as part of its illustrative examples. Only when headwords happen to be key words in proverbs have the latter been provided. Illustrative examples occur at the end of the definition of a headword in many dictionaries. It is often difficult for dictionary users to find specific or relevant proverbs because they do not recognise the order of their arrangement. This is partly because of the absence of information on how to find proverbs in the user's style guide. The proverbs in this particular dictionary are listed under their key words. A dictionary user must therefore identify the key word in the proverb and look for this word in the dictionary. Information regarding how to find the proverbs in this dictionary could be valuable to dictionary users. The purpose of this paper is to provide important directions to dictionary users to assist them in finding proverbs, and to discuss the importance of finding proverbs in dictionaries such as the Venda Dictionary: Tshivenḓa – English.

The application of corpus-based approach in the Bulgarian new-word lexicography

Sia Kolkovska, Diana Blagoeva & Atanaska Atanasova

Keywords: *new-word lexicography*, *corpus-based approach*, *Bulgarian lexicography*.

The paper focuses on the specific directions of application of the corpus-based approach in the new-word lexicography. The research deals with the main application of corpus-based techniques in the elaboration of the latest academic neological Bulgarian dictionary. The usefulness of applying corpus-based techniques at the various stages of compiling this neological dictionary is described in more details. The usages of these techniques make compiling of the Dictionary easier at the following stages: working out the list of new units, included as head words in the Dictionary; determining of the degree of establishment of the new units in Bulgarian; determining of the representative variant among some graphic, phonetic or morphological variants of a new word; determining of the most typical collocations of a given head word.

Single-clause when-definitions: Take three

Robert Lew & Anna Dziemianko

Keywords: *definition*, *folk defining*, *syntactic information*, *learner's dictionary*, *definition format*.

In our EURALEX 2006 contribution (Dziemianko and Lew 2006), we focused on the practice of defining certain abstract nouns by means of a *when*-clause, which seems to have gained much popularity in recent years in some major monolingual English learner's dictionaries

. We tested the hypothesis that a definition of this format would fare worse than the classic analytical definition in terms of conveying information on the syntactic class of the lemma. Experiments with Polish high-intermediate and advanced learners of English provided strong empirical support for this hypothesis. However, the testing instruments employed in the 2006 study used a relatively restricted microstructure, with just headwords and definitions. In the present follow-up study, we attempt to verify the results using a more complete microstructure to assess the strength of the effect of single-clause *when*-definitions on syntactic class identification in the presence of other potential indicators of syntactic class. Below we summarize the findings of the whole series of studies of this contentious defining format.

Underlying principles of *Gentyll* English-Spanish non-sexist glossaries: A response to a need

Mercedes Bengoechea & María Rosa Cabellos

Keywords: *sexist lexicography, feminist guidelines, non-sexist occupational titles, Spanish sexed terms, neutral English, gender-aware glossaries*.

Our research team has elaborated a series of Spanish/English glossaries of specialized terms for man and woman in various subject fields which can be consulted online at http://gentyll.uah.es/glossaries.html. The glossaries aim to challenge traditional sexist practices in terminology and lexicography, and follow the recommendations for non-sexist usage issued by various institutions, agencies and scholars. It is a project still in progress which aims to be expanded into more subject fields and languages.

The aim of this paper is twofold: on the one hand, to highlight the necessity of gender aware alternatives to existing terminology databases and dictionaries, and, on the other, to facilitate an understanding of the principles that we have adopted in our glossaries –principles consistent with our criticism to existing lexicographical and terminological resources.

Growing naturally: The DicSci Organic E-Advanced Learner's Dictionary of Verbs in Science

Geoffrey Williams, Chrystel Millon & Araceli Alonso

Keywords: *collocational networks, verbal patterns, learner's dictionary, specialised dictionary, organic dictionary, phraseology*.

In this paper we illustrate the principles and building methodology of the E-Advanced Learner's Dictionary of Verbs in Science (DicSci), paying special attention to the methodology being developed for its compilation which is based on the application of collocational networks and the adaptation of Corpus Pattern Analysis (Hanks 2004, forthcoming) to specialised language environments. DicSci focuses on showing specialised usage patterns commonly associated with certain verbs used in specialised contexts by means of collocational networks (Williams 1998). The different steps to create the dictionary, its present state and plans for its completion and future are explained.

The evolution of the Romanian digitalized lexicography. The essential Romanian lexicographic corpus

Elena Tamba Dănilă, Marius-Radu Clim, Mădălin Pătraşcu & Ana Catană-Spenchiu

Keywords: *Romanian lexicography, computerized lexicography, linguistic resources, computerized lexicographic instruments.*

The aim of this paper is to highlight the present stage of the digitalized lexicographic research from Romania and the importance of creating a Romanian Essential Lexicography Corpus. In the last years there have been taken measures for creating electronic instruments and resources that are necessary for supporting the Romanian language and culture on a transnational level, in the general context of the computerization of the fundamental academic research.

The Romanian academic specialists in linguistics and applied informatics, as well as in computational linguistics fields, have initiated research projects by which they want to valorise the non-digitized resources by acquiring them in electronic formats and to create new resources and instruments for the automatic processing of the language.

The project presented in this paper has as purpose the valorization of certain results from the complex project eDTLR, by using, as reference text for the alignment, the Thesaurus Dictionary in electronic format and creating a Romanian lexicographic corpus. This project's aims are: the realization of a scanned corpus, with the reference dictionaries of DLR (taking into account the present legislation regarding copyright); scanning and processing of these dictionaries (by OCR – optical character recognition – the conversion from image to text; parsing the text at entry); realizing an on-line interface for validating/correcting of the parsing (= automatic identification of the entries from previously scanned and converted dictionaries), as well as validating the alignment between the text of the Romanian Language Thesaurus Dictionary (in electronic format, from eDTLR project) and the reference dictionaries from DLR Bibliography. The final database will include an important number of essential Romanian language dictionaries (100 dictionaries from the 16th century to present day) aligned at entry level, fact that will offer Romanian specialists an excellent working instrument and will set basis for future research.

Exploring semantic change with lexical sets

Karin Cavallin

Keywords: *lexical sets, semantic change, language technology*.

Many areas of linguistics which use corpora as their main data have benefited from research in natural language processing, NLP. Apart from a few recent studies such as Sagi et al. (2009), Rohrdantz et al. (2011) and the GoogleNgram-viewer (Michel et al. 2011), the field of semantic change seems to have received little attention in NLP. This

paper describes some first steps in viewing semantic change in terms of distributional semantics with a computational and linguistically motivated approach. By parsing, adding lemmatization and part of speech information, a method is developed to describe semantic behavior and to track semantic change over time. In distributional semantics, meaning is characterized with respect to the context. This idea is developed from Firth (1957) and is formulated according to 'the distributional hypothesis' of Harris (1968). Whereas most approaches to statistical semantics uses some kind of vector analysis based on ngrams. Distribution here is presented as the statistically ranked lists of verb-object constructions, that is 'lexical sets'. A lexical set is more focused than ngrams and can be seen as essential minimal co-occurrence information for a given word, which facilitates manual analysis.

Dictionary of valencies meets corpus annotation: A case of Russian FrameBank

Olga Lyashevskaya

Keywords: *frame semantics*, *FrameNet*, *Construction Grammar*, *Russian*

Abstract

The Russian FrameBank project aims at the development of a hybrid lexical resource that links a dictionary of valencies and an annotated corpus. Two types of data present generalized lexical constructions (LexCxs) and their realizations in contemporary written texts (1950-present).

The predicate-argument structure for verbs, nominalizations, adjectives, adverbs, and other lexical units in Russian is mostly encoded in case and prepositional marking while word alignment is determined by information structure. This means that an argument can be found in any part of the sentence and the window for argument detection is infinitely wide. Russian predicates reveal more than 1000 typical morphosyntactic patterns; the number of shallow realizations under certain grammatical and discourse constraints is even greater.

Morphosyntactic patterns are not fully predictable by semantics (Apresjan 1967), and, hence, we can speak here about lexical constructions. The patterns with lexical slots evoked by two or more target lexemes (e.g. idiomatic phrases like *vzjal i <uexal>* 'he suddenly (lit. took and) <went away>') are also treated as LexCxs. As experiments on unsupervised LexCx retrieval have shown (Toldova et al. 2008, Lashevskaja and Mitrofanova 2009), there is a great need for an open data pool annotated manually for lexical frames. In a wider perspective, the project on tagging the form and meaning pairings is of great significance for lexical and syntactic research, lexicography, and IR tasks.

The dictionary of lexical constructions matches frames evoked by a particular target word into morphosyntactic patterns. The relevant dataset here is semantic explications (roles), lexico-semantic constrains (e.g. human, emotion, etc.), morphosyntactic constraints on the elements, their syntactic ranks.

FrameBank is an offspring project of the Russian National Corpus (http://www.ruscorpora.ru) and involves a large illustrative sample taken from the corpus. The goal of framenet-like corpus annotation is to reveal the diverse realizations of a certain LexCxs in the running text and to mark the elements that correspond to constructional arguments and adjuncts. The corpus part of FrameBank details morphological and syntactic mismatches, violation of lexical and semantic constraints, and focuses on the grammatical constructions that introduce or license the use of elements within a given construction. This is a report on work in progress, which can be followed at http://framebank.ru.

Index

List of authors

Abel, Andrea. European Academy Bolzano/Bozen. andrea.abel@eurac.edu
Aldea, Maria. "Babes-Bolyai" University, Cluj-Napoca. aldea_maria@yahoo.com
Alonso, Araceli. Universitat Pompeu Fabra; Université de Bretagne-Sud. aalonsocampo@gmail.com
Armosti, Spyros. University of Cyprus. armostis@cantab.net
Atanasova, Atanaska. Institute for Bulgarian Language, Bulgarian Academy of Sciences. nassi_n@abv.bg
Autelli, Erica. University of Innsbruck, Department of Romanistics. a.Autelli@uibk.ac.at
Baisa, Vít. Masaryk University; Lexical Computing Ltd. vit.baisa@gmail.com
Bano, Mahwish. Center for Language Engineering (CLE). mahwish.bano@kics.edu.pk
Battaner, Paz. Universitat Pompeu Fabra. paz.battaner@upf.edu
Berthemet, Elena. Université de Bretagne Occidentale. elena.berthemet@univ-brest.fr
Bejček, Eduard. Charles University, Prague. bejcek@ufal.mff.cuni.cz
Bengoechea, Mercedes. Universidad de Alcalá. mercedes.bengoechea@uah.es
Benko, Vladimír. Slovak Academy of Sciences, Ľ. Štúr Institute of Linguistics. vladob@juls.savba.sk
Berg-Olsen, Sturla. University of Oslo. sturla.berg-olsen@iln.uio.no
Blagoeva, Diana. Institute for Bulgarian Language, Bulgarian Academy of Sciences. diablag@mail.bg
Blancafort, Helena. Syllabs. blancafort@syllabs.com
Bondi Johannessen, Janne. The Text Laboratory, ILN, University of Oslo. j.b.johannessen@iln.uio.no
Bosch, Sonja. Department of African Languages, University of South Africa. boschse@unisa.ac.za
Bosman, Nerina. University of Pretoria. nerina.bosman@up.ac.za
Bradl-Albrich, Martina. University of Innsbruck, Department of Romanistics. MartinaAlbrich@gmx.at
Budykina, Vera. Chelyabinsk State University. vbudykina@gmail.com
Buendía Castro, Miriam. Universidad de Granada. mbuendia@ugr.es
Cabellos, María Rosa. Universidad de Alcalá. rosa.cabellos@uah.es

Catana-Spenchiu, Ana. Institute of Romanian Philology, The Romanian Academy. anaspenchiu@gmail.com
Cavallin, Karin. University of Gothenburg. karin.cavallin@gu.se
Chan, Alice Yin Wa. City University of Hong Kong. enalice@cityu.edu.hk
Chlumska, Lucie. Institute of the Czech National Corpus, Charles University. lucie.chlumska@ff.cuni.cz.
Choi, Jun. Kyungpook National University. c-juni@hanmail.net
Christodoulou, Kyriaki. University of Cyprus. kyriaki_ptf@yahoo.gr
Cinková, Silvie. Institute of Formal and Applied Linguistics, Charles University. cinkova@ufal.mff.cuni.cz
Clim, Marius-Radu. Institute of Romanian Philology, The Romanian Academy. marius.clim@gmail.com
Coates, Richard. UWE. Richard.Coates@uwe.ac.uk
Cook, Paul. The University of Melbourne, Department of Computer and Information Systems. paulcook@unimelb.edu.au
Cvrcek, Vaclav. Institute of the Czech National Corpus, Charles University. vaclav.cvrcek@ff.cuni.cz
Davidsdottir, Rosa Elin. Paris-Sorbonne University (Paris IV). rosaelin@gmail.com
DeCesaris, Janet. Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. janet.decesaris@upf.edu
Demers, Marie-Claude. OLST, Université de Montréal. marie-claude.demers.2@umontreal.ca
Didakowski, Jörg. Berlin-Brandenburgische Akademie der Wissenschaften. didakowski@bbaw.de
Dziemianko, Anna. Adam Mickiewicz University. danna@ifa.amu.edu.pl
van den Eerenbeemt, Arnoud. Bohn stafleu van Loghum. a.vandeneerenbeemt@bsl.nl
Eklund, Ann-Marie. Språkbanken, Department of Swedish Language, University of Gothenburg. ann-marie.eklund@gu.se
Eskonsipo, Berit Nystad. University of Tromsø. berit.nystad.eskonsipo@uit.no
Feroz, Gull. Center for Language Engineering (CLE). gullferoz6@gmail.com
Ferrara-Léturgie, Alice. Université de Cergy-Pontoise. alice.ferrara@gmail.com
Fjeld, Ruth Vatvedt. University of Oslo. r.e.v.fjeld@iln.uio.no

Fotopoulou, Aggeliki. Institute for Language and Speech Processing, R.C. "Athena". afotop@ilsp.athena-innovation.gr
Friberg, Håkan. Blackebergs Gymnasium. hakan.friberg@stockholm.se
Friberg Heppin, Karin. University of Gothenburg. karin.friberg@svenska.gu.se
Furiassi, Cristiano. Università degli Studi di Torino. cristiano.furiassi@unito.it
Gao, Yongwei. College of Foreign Languages & Literature, Fudan University. ywgao@fudan.edu.cn
Gehweiler, Elke. DWB, Berlin-Brandenburgische Akademie der Wissenschaften. elke.gehweiler@fu-berlin.de
Geyken, Alexander. Berlin-Brandenburgische Akademie der Wissenschaften. geyken@bbaw.de
Giouli, Voula. Institute for Language and Speech Processing, R.C. "Athena". voula@ilsp.gr
Giovannetti, Emiliano. Istituto di Linguistica Computazionale - National Research Council. emiliano.giovannetti@ilc.cnr.it
Glennon, Dominic. Cambridge University Press. dglennon@cambridge.org
Gojun, Anita. University of Stuttgart, gojunaa@ims.uni-stuttgart.de
Goossens, Diane. Université catholique de Louvain. diane.goossens@uclouvain.be
Granger, Sylviane. University of Louvain, Centre for English Corpus Linguistics. sylviane.granger@uclouvain.be
Griškevičienė, Aurelija. Vilnius University. aurelija.griskeviciene@flf.vu.lt
de Groc, Clément. Syllabs. groc@syllabs.com
ten Hacken, Pius. Department of Languages, Translation and Media, Swansea University. p.ten-hacken@swansea.ac.uk
Hanks, Patrick. UWE. patrick.w.hanks@gmail.com
Hansen, Carsten. LANCHART – ODT. carhan@hum.ku.dk
Hansen, Martin H. LANCHART – ODT. martin_h_hansen@yahoo.dk
Harhata, Bogdan. "Sextil Puscariu" Institute of Linguistics and Literary History, the Romanian Academy branch of Cluj-Napoca. bogdan_harhata@yahoo.it
Heid, Ulrich. Universität Hildesheim; University of Stuttgart heid@ims.uni-stuttgart.de
Héja, Enikő. Research Institute for Linguistics, Hungarian Academy of Sciences. eheja@nytud.hu

Henriksen, Birgit. Department of English, Germanic and Romance Studies, Center of Internationalisation and Parallel Language Use, University of Copenhagen. birgit@hum.ku.dk
Henriksen, Petter. Det Norske Akademis store ordbok. petter.henriksen@kunnskapsforlaget.no
Heyvaert, Frans. Instituut voor Nederlandse Lexicologie. Frans.Heyvaert@inl.nl
Holub, Martin. Institute of Formal and Applied Linguistics, Charles University. holub@ufal.mff.cuni.cz
Hult, Ann-Kristin. Department of Swedish, Centre for Lexicography and Lexicology, University of Gothenburg. ann-kristin.hult@svenska.gu.se
Hussain, Sarmad. Center for Language Engineering (CLE). sarmad.hussain@kics.edu.pk
Hyun, Young-Hee. Kyungpook National University. dew840713@hanmail.net
Iversen, Sarah Hoem. University of Oxford. sarah.iversen@gmail.com
Jakubicek, Milos. Lexical Computing Ltd. milos.jakubicek@sketchengine.co.uk
Jansson, Håkan. Dept. of Swedish, University of Gothenburg. hakan.jansson@svenska.gu.se
Jarošová, Alexandra. Slovak Academy of Sciences, Ľ. Štúr Institute of Linguistics. sasaj@juls.savba.sk
Johansson Kokkinakis, Sofie. Dept. of Swedish, University of Gothenburg. sofie.johansson.kokkinakis@svenska.gu.se
Jürviste, Madis. Institute of the Estonian Language, University of Tartu. madis.jyrviste@eki.ee
Kallasmaa, Marja. Institute of the Estonian Language. marja.kallasmaa@eki.ee
Karlholm, Annika. Institute for Language and Folklore, Department of Dialectology and Folklore Research. annika.karlholm@sofi.se
Karpova, Olga. Ivanovo State University. olga.m.karpova@gmail.com
Katsoyannou, Marianna. University of Cyprus. marianna@ucy.ac.cy
Kernerman, Ilan. K Dictionaries. ilan@kdictionaries.com
Kettnerová, Václava. Charles University, Prague. kettnerova@ufal.mff.cuni.cz
Kilgarriff, Adam. Lexical Computing Ltd. adam@lexmasterclass.com
Kinn, Kari. The Text Laboratory, ILN, University of Oslo. kari.kinn@iln.uio.no
Klein, Juliane. Uni Leipzig. julianeklein.trier@googlemail.com

Klosa, Annette. Institut für Deutsche Sprache. klosa@ids-mannheim.de
Knudsen, Rune Lain. University of Oslo. runelk@ifi.uio.no
Kokkinakis, Dimitrios. Språkbanken, Department of Swedish Language, University of Gothenburg. dimitrios.kokkinakis@svenska.gu.se
Kola, Kjersti Wictorsen. University of Oslo. kjerstwk@student.ilos.uio.no
Kolkovska, Sia. Institute for Bulgarian Language, Bulgarian Academy of Sciences. sia_btb@yahoo.com
Kompara, Mojca. Department of Comparative and General Linguistics, University of Ljubljana. mokopt@yahoo.com
Konecny, Christine. University of Innsbruck, Department of Romanistics. Christine.Konecny@uibk.ac.at
Konovalova, Maria. Saint-Petersburg State University. maschako@rambler.ru
Kovarikova, Dominika. Institute of the Czech National Corpus, Charles University. dominika.kovarik@gmail.com
Kríž, Vincent. Institute of Formal and Applied Linguistics, Charles University. vincent.kriz@gmail.com
Lefer, Marie-Aude. University of Louvain, Centre for English Corpus Linguistics. marie-aude.lefer@uclouvain.be
Lemnitzer, Lothar. Berlin-Brandenburgische Akademie der Wissenschaften. lemnitzer@bbaw.de
León Araúz, Pilar. University of Granada. pleon@ugr.es
Léturgie, Arnaud. Université de Cergy-Pontoise. arnaud.leturgie@gmail.com
Leucuta, Daniel-Corneliu. Department of Medical Informatics and Biostatistics, "Iuliu Hatieganu" University of Medicine and Pharmacy, Cluj-Napoca. dleucuta@umfcluj.ro
Lew, Robert. Adam Mickiewicz University. rlew@amu.edu.pl
L'Homme, Marie-Claude. OLST, Université de Montréal. mc.lhomme@umontreal.ca
Lopatková, Markéta. Charles University, Prague. lopatkova@ufal.mff.cuni.cz
Lorentzen, Henrik. Society for Danish Language and Literature. hl@dsl.dk
Lukyanova, Ekaterina. Saint-Petersburg State University. ekaterina_lukyanova@yahoo.com
Lyashevskaya, Olga. National Research University; Higher School of Economics; Vinogradov Institute of Russian Language RAS. olesar@gmail.com

Mafela, Munzhedzi James. University of South Africa. mafelmj@unisa.ac.za
Magga, Ole Henrik. Sámi University College. Ole-Henrik.Magga@samiskhs.no
Mahlow, Cerstin. Department of German, University of Basel. cerstin.mahlow@unibas.ch
Marello, Carla. Università di Torino. Carla.marello@unito.it
Margalitadze, Tina. Iv. Javakhishvili Tbilisi State University. tinatin@margaliti.ge
Marinov, Svetoslav. Findwise AB. svetoslav.marinov@findwise.com
McCarthy, Diana. Lexical Computing Ltd. diana@dianamccarthy.co.uk
Měchura, Michal Boleslav. Fiontar, Dublin City University. mechrm@dcu.ie
Millon, Chrystel. Université de Bretagne-Sud. chrystel.millon@gmail.com
Mulhall, Chris. Waterford Institute of Technology. chrmulhall@gmail.com
Nam, Kil-Im. Kyungpook National University. ki@knu.ac.kr
Nazar, Rogelio. University Institute of Applied Linguistics - Universitat Pompeu Fabra - Barcelona, Spain. rogelio.nazar@upf.edu
Niestadt, Jan. Instituut voor Nederlandse Lexioclogie. jan.niestadt@inl.nl
Novotná, Renata. Institute of the Czech National Corpus, Charles University, Prague. renata.novotna@ff.cuni.cz
Oelke, Daniela. Department of Computer and Information Science, University of Konstanz. oelke@dbvis.inf.uni-konstanz.de
Oja, Vilja. Institute of the Estonian Language. vilja.oja@eki.ee
Ostermann, Carolin. Friedrich-Alexander University Erlangen-Nuremberg. Carolin.Ostermann@angl.phil.uni-erlangen.de
Patrascu, Madalin. Institute of Romanian Philology, The Romanian Academy. madalin.patrascu@gmail.com
Pedersen, Bolette Sanford. University of Copenhagen. bspedersen@hum.ku.dk
Perdek, Magdalena. Adam Mickiewicz University. mperdek@ifa.amu.edu.pl
Pereira, Sandra. Centro de Linguística - Universidade de Lisboa. spereira@clul.ul.pt
Piccini, Silvia. Istituto di Linguistica Computazionale - National Research Council. silvia.piccini@ilc.cnr.it
Pinnavaia, Laura. University of Milan. laura.pinnavaia@unimi.it

Pomikálek, Jan. Faculty of Informatics, Masaryk University; Lexical Computing Ltd. xpomikal@fi.muni.cz
Preite, Chiara. Università di Modena e Reggio Emilia. chiara.preite@unimore.it
Prinsloo, Danie J. Department of African Languages, University of Pretoria. danie.prinsloo@up.ac.za
Pujsza, Julia. Uniwersytet Gdański. j.pujsza@yahoo.de
Pvs, Avinesh. Lexical Computing Ltd. avinesh.pvs@gmail.com
Radtke, Janina Désirée. Universität Hildesheim. janina.radtke@gmx.de
Raïssa, Gillier. Centro de Linguística, Universidade de Lisboa. raissa.gillier@clul.ul.pt
Renau, Irene. University Institute of Applied Linguistics, Universitat Pompeu Fabra. irene.renau@upf.edu
Ribeck, Judy. Dept. of Swedish, University of Gothenburg. carola.ribeck@gu.se
Rica-Peromingo, Juan-Pedro. Complutense University. juanpe@filol.ucm.es
Ruimy, Nilda. Istituto di Linguistica Computazionale - National Research Council. nilda.ruimy@ilc.cnr.it
Rundell, Michael. Macmillan Dictionaries; Lexicography MasterClass. michael.rundell@lexmasterclass.com
Rychly, Pavel. Masaryk University; Lexical Computing Ltd. pary@fi.muni.cz
San Martín, Antonio. University of Granada. pleon@ugr.es
Sanchez Cardenas, Beatriz. Universidad de Granada. bsc@ugr.es
de Santiago, Paula. Department of English Studies, University of Valladolid. padesantiago@hotmail.com
Schnörch, Ulrich. Institut für Deutsche Sprache. schnoerch@ids-mannheim.de
Schoonheim, Tanneke. Instituut voor Nederlandse Lexicologie. tanneke.schoonheim@inl.nl
de Schryver, Gilles-Maurice. Dept. of Languages and Cultures, Ghent University; African Linguistics and Xhosa Dept., University of the Western Cape. GillesMaurice.DeSchryver@ugent.be
Selegey, Vladimir. ABBYY. vladimir_s@abbyy.com
Sharifi, Saghar. Islamic Azad University, Karaj Branch, Karaj-Iran. saghar_sharifi@yahoo.com
Simón, José. Universidad de Alcalá. josef.simon@uah.es
Sköldberg, Emma. Dept. of Swedish, University of Gothenburg. emma.skoeldberg@svenska.gu.se

Smith-Muller, Wanda. Pharos Dictionaries. wanda.smith@pharos.co.za
Song, Hyeon-Ju. Kyungpook National University. songhj@knu.ac.kr
Storjohann, Petra. Institut für Deutsche Sprache. storjohann@ids-mannheim.de
Takács, Dávid. Research Institute for Linguistics, Hungarian Academy of Sciences. takdavid@nytud.hu
Tamba Danila, Elena. Institute of Romanian Philology, The Romanian Academy. isabelle.danila@gmail.com
Tempelaars, Rob. Instituut voor Nederlandse Lexicologie. rob.tempelaars@inl.nl
Theilgaard, Liisa. Society for Danish Language and Literature. lt@dsl.dk
Thelin, Eva. Institute for Language and Folklore, Department of Dialectology and Folklore Research. eva.thelin@sofi.se
Themistocleous, Charalambos. University of Cyprus. themistocleous@gmail.com
Tiberius, Carole. Instituut voor Nederlandse Lexicologie. carole.tiberius@inl.nl
Tolochin, Igor. Saint-Petersburg State University. itfipe@mail.wplus.net
Torjusen, Julie Matilde. University of Oslo. juliemt@student.uio.no
Trosterud, Trond. University of Tromsø. trond.trosterud@uit.no
Ullah, Inam. Center for Language Engineering (CLE). torwalpk@yahoo.com
Unger, Christiane. DWB, Berlin-Brandenburgische Akademie der Wissenschaften.
Van Mol, Marc. Catholic University of Leuven. mark.vanmol@ilt.kuleuven.be
Veisbergs, Andrejs. University of Latvia. anveis@lanet.lv
Veldi, Enn. Dept. of English, University of Tartu. Enn.Veldi@ut.ee
Vojtěch, Kovář. Masaryk University; Lexical Computing Ltd. xkovar3@gmail.com
Vonen, Arnfinn Muruvik. Language Council of Norway. arnfinn.muruvik.vonen@sprakradet.no
Vremir, Lilla Marta. "Sextil Puscariu" Institute of Linguistics and Literary History, the Romanian Academy branch of Cluj-Napoca. vremirm@yahoo.com
Wandl-Vogt, Eveline. Academy of Sciences. eveline.wandl-vogt@oeaw.ac.at
Whitelock, Pete. Oxford University Press. pete.whitelock@oup.com
Williams, Geoffrey. Université de Bretagne-Sud. williams@univ-ubs.fr

Zabarskaite, Jolanta. Institute for Lithuanian Language.
jolanta.zabarskaite@lki.lt
Zimmermann, Jan Timo. Universität Hildesheim.
JT.Zimmermann@gmx.de
Zingano Kuhn, Tanara. Leiden University. tanarazingano@yahoo.com


iFinger
software


OXFORD
UNIVERSITY PRESS