



**XVI EURALEX International Congress:
The User in Focus**

**15 – 19 July 2014
Bolzano/Bozen**

BOOK OF ABSTRACTS



Plenary Lectures	7
From lexicography to Terminology: a Cline, not a Dichotomy	9
<i>Thierry Fontenelle</i>	
Natural Language Processing Techniques for Improved User-Friendliness of Electronic Dictionaries	9
<i>Ulrich Heid</i>	
Using Mobile Bilingual Dictionaries in an EFL Class	10
<i>Carla Marelllo</i>	
Meanings, Ideologies, and Learners' Dictionaries	11
<i>Rosamund Moon</i>	
The Dictionary-Making Process	13
The Making of a Large English-Arabic/Arabic-English Dictionary: the Oxford Arabic Dictionary	15
<i>Tressy Arts</i>	
Simple and Effective User Interface of the Dictionary Writing System	15
<i>Kamil Barbierik, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Miroslav Vírúš</i>	
Simple and effective user interface of our new DWS	15
<i>Kamil Barbierik, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Miroslav Vírúš</i>	
The Dictionary Writing System for the Academic Dictionary of the Contemporary Czech	16
<i>Kamil Barbierik, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Miroslav Vírúš</i>	
Totalitarian Dictionary of Czech	16
<i>František Čermák</i>	
Dictionary of Abbreviations in Linguistics: Towards Defining Cognitive Aspects as Structural Elements of the Entry	16
<i>Ivo Fabijanić</i>	
La definizione delle relazioni intra- e interlinguistiche nella costruzione dell'ontologia IMAGACT	17
<i>Gloria Gagliardi</i>	
I neologismi pianificati: esempi dal ladino fassano nell'ultima edizione del DILF.	17
<i>Michela Giovannini</i>	
DPS – single source production system	18
<i>Holger Hvelplund</i>	
A Dictionary of Old Norse Prose and its Users — Paper vs. Web-based Edition	18
<i>Ellert Thor Johannsson, Simonetta Battista</i>	
Making a Learner's Dictionary of Academic English	18
<i>Diana Lea</i>	
The Danish Thesaurus: Problems and Perspectives	19
<i>Sanni Nimb, Lars Trap-Jensen, Henrik Lorentzen</i>	
From a Dialect Dictionary to an Etymological One	19
<i>Vilja Oja, Iris Metsmägi</i>	
Research on Dictionary Use	21
On the presentation and placement of collocations in monolingual English learners' dictionaries	23
<i>Anna Dziemianko</i>	
Wörterbuchbenutzung: Ergebnisse einer Umfrage bei italienischen DaF-Lernern	23
<i>Carolina Flinz</i>	
Translation, cultural adaptation and preliminary psychometric evaluation of the English version of "Strategy Inventory for Dictionary Use" (S.I.D.U)	23
<i>Zoe Gavriilidou</i>	
The Authentic Voices of Dictionary Users – Viewing Comments on an Online Learner's Dictionary Before and After Revision	24
<i>Ann-Kristin Hult</i>	
Mobile Lexicography: A Survey of the Mobile User Situation	24
<i>Henrik Köhler Simonsen</i>	
Kollokationsübungen für den Unterricht als Wörterbuchergänzung	24
<i>Stefanie Meier, Alessandra Widmer</i>	
Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht	25
<i>Martina Nied Curcio</i>	
Do menus provide added value to signposts in monolingual dictionary entries?	25
<i>Bartosz Ptasznik, Robert Lew</i>	
Dictionary users do look up frequent and socially relevant words. Two log file analyses	26
<i>Sascha Wolfer, Alexander Kopleinig, Peter Meyer, Carolin Müller-Spitzer</i>	
La performance dell'utente apprendente di italiano LS/L2 e la microstruttura dei dizionari: sussidi per lo sviluppo della Lessicografia Pedagogica	26
<i>Angela Maria Tenório Zucchi</i>	
Lexicography and Language Technologies	27
ALIQUEOT – Atlante della Lingua Italiana QUOTidiana	29
<i>Michele Castellarin, Fabio Tosques</i>	

Applying a Word-sense Induction System to the Automatic Extraction of Diverse Dictionary Examples	29
<i>Paul Cook, Michael Rundell, Jey Han Lau, Timothy Baldwin</i>	
Cross-linking Austrian dialectal Dictionaries through formalized Meanings.....	30
<i>Thierry Declerck, Eveline Wandl-Vogt</i>	
Nutzung des DWDS-Wortprofils beim Aufbau eines lexikalischen Informationssystems zu deutschen Stützverbgefügen	30
<i>Jörg Didakowski, Nadja Radtke</i>	
Automation of Lexicographic Work Using General and Specialized Corpora: Two Case Studies.....	30
<i>Iztok Kosem, Polona Gantar, Nataša Logar, Simon Krek</i>	
A Corpus-based Dictionary of Polish Sign Language (PJM).....	31
<i>Jadwiga Linde-Usiekniewicz, Małgorzata Czajkowska-Kisil, Joanna Łacheta, Paweł Rutkowski</i>	
Laying the Foundations for a Diachronic Dictionary of Tunis Arabic: a First Glance at an Evolving New Language Resource	31
<i>Karlheinz Mörth, Stephan Procházka, Ines Dallaji</i>	
BabelNet meets Lexicography: the case of an automatically-built multilingual encyclopedic dictionary	31
<i>Roberto Navigli</i>	
Tracing Sentiments: Syntactic and Semantic Features in a Subjectivity Lexicon.....	32
<i>Jana Šindlerová, Kateřina Veselovská, Jan Hajič, jr.</i>	
A Simple Platform for Defining Idiom Variation Matching Rules	32
<i>Koichi Takeuchi, Ulrich Apel, Ray Miyata, Ryo Murayama, Ryoko Adachi, Wolfgang Fanderl, Iris Vogel, Kyo Kageura</i>	
Lexicography and Corpus Linguistics	33
Compatible Sketch Grammars for Comparable Corpora	35
<i>Vladimír Benko</i>	
From GLÀFF to PsychoGLÀFF: a large psycholinguistics-oriented French lexical resource	35
<i>Basilio Calderone, Nabil Hathout, Franck Sajous</i>	
The basic vocabulary of Italian from 1980 to the present time: methods and results	35
<i>Tullio De Mauro, Isabella Chiari</i>	
RIDIRE. Corpus and Tools for the Acquisition of Italian L2.....	36
<i>Alessandro Panunzi, Emanuela Cresti, Lorenzo Gregori</i>	
Empirical approaches to German paronyms.....	36
<i>Petra Storjohann, Ulrich Schnörch</i>	
Pragmatic Meaning in Lexicographical Description: Semantic Prosody on the Go	36
<i>Mojca Šorli</i>	
Bi- and Multilingual Lexicography	37
Linking a dictionary to other open data – Better access to more specific information for the users	39
<i>Ulrich Apel</i>	
Bilingual word sketches: the <i>translate</i> button	39
<i>Vít Baisa, Miloš Jakubíček, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý</i>	
Creating a Bilingual Italian-English Dictionary of Collocations	39
<i>Barbara Berti, Laura Pinnavaia</i>	
La valencia del adjetivo en diccionarios bilingües alemán-español-alemán	40
<i>Andreu Castell, Natàlia Català, Maria Bargalló</i>	
Esame storico dei “realia” nei dizionari bilingui italiano/ungheresi.....	40
<i>Zsuzsanna Fábíán</i>	
Quello che i dizionari possono fare: l’esempio dei Dizionari di Tedesco (Giacoma/Kolb –Zanichelli/Klett).....	41
<i>Luisa Giacoma</i>	
Bilingual Dictionary Drafting. The example of German-Basque, a medium-density language pair	41
<i>David Lindemann, Iker Manterola, Rogelio Nazar, Iñaki San Vicente, Xabier Saralegi</i>	
Illustrative Examples and the Aspect of Culture: The Perspective of a Tshivenda Bilingual Dictionary.....	42
<i>Munzhedzi James Mafela</i>	
Corpus, Parallélisme et Lexicographie Bilingue	42
<i>Adriana Zavaglia, Gisele Galafacci</i>	
Lexicography for Specialised Languages, Technology and Terminography	43
Contrastive Collocational Networks for Compiling Specialized Dictionaries for Learners of Spanish for Specific Purposes.....	45
<i>Araceli Alonso Campo</i>	
EcoLexicon.....	45
<i>Pamela Faber, Miriam Buendía Castro</i>	
Experts and terminologists: exchanging roles in the elaboration of the Terminological Dictionary of the Brenner Base Tunnel (BBT).....	45
<i>Elena Chiocchetti, Natascia Ralli</i>	
Cloud Terminology Services Facilitate Specialised Lexicography Work.....	46
<i>Tatiana Gornostay, Andrejs Vasiljevs</i>	

Good Contexts for Translators—A First Account of the Cristal Project.....	46
<i>Amélie Josselin-Leray, Cécile Fabre, Josette Rebeyrolle, Aurélie Picton, Emmanuel Planas</i>	
Kontextbasierte lexikalische Kontrolle von Anforderungsdokumenten	47
<i>Jennifer Krisch</i>	
From term dynamics to concept dynamics: term variation and multidimensionality in the psychiatric domain	47
<i>Pilar León-Araúz, Arianne Reimerink</i>	
<i>Bon usage</i> vs. Fachliches: Fachsprache in der Geschichte der französischen Sprachpflege und Lexikographie.....	48
<i>Martina Mayer</i>	
EU-Terminologie in den einsprachigen Wörterbüchern des Deutschen	48
<i>Diana Stantcheva</i>	
Text boxes as lexicographic device in LSP dictionaries.....	48
<i>Elsabé Taljard, Danie J. Prinsloo, Rufus H. Gouws</i>	
Station Sensunique: Architecture générale d'une plateforme web paramétrable, modulaire et évolutive d'acquisition assistée de ressources	49
<i>Izabella Thomas, Blandine Plaisantin Alecu, Bérenger Germain, Marie-Laure Betbeder</i>	
Station Sensunique: une plateforme Web modulaire, collaborative et évolutive d'acquisition assistée de ressources terminologiques et non terminologiques (orientée Langues Contrôlées)	49
<i>Izabella Thomas, Blandine Plaisantin Alecu, Bérenger Germain, Marie-Laure Betbeder</i>	
Lexicography of Lesser Used Languages.....	51
Towards an Integrated E-Dictionary Application – The Case of an English to Zulu Dictionary of Possessives	53
<i>Sonja Bosch, Gertrud Faaß</i>	
Zur (Vor-)Geschichte der saamischen Lexikografie: ein lateinisch-saamisches Wörterverzeichnis aus dem 17. Jahrhundert	53
<i>Eino Koponen</i>	
Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses	54
Dearcadh na nDéise – Representations of Gaeltacht na nDéise in Dineen's Bilingual Irish-English Dictionary (1927)	54
<i>Chris Mulhall, Seamus ó'Diollúin</i>	
The eLexicon Mediae et Infimae Latinitatis Polonorum. The Electronic Dictionary of Polish Medieval Latin.....	55
<i>Krzysztof Nowak</i>	
From DANTE to Dictionary: The New English-Irish Dictionary.....	55
<i>Pádraig Ó Mianáin, Cathal Convery</i>	
User Support in e-Dictionaries for Complex Grammatical Structures in the Bantu Languages.....	55
<i>Danie J. Prinsloo, Theo J.D. Bothma, Ulrich Heid</i>	
Concerning the Treatment of Co-existent Synonyms in Estonian Monolingual and Bilingual Dictionaries	56
<i>Enn Veldi</i>	
Phraseology and Collocation.....	57
Unusual phrases in English MLDs: increasing user friendliness	59
<i>Stephen Coffey</i>	
Harvesting from One's Own Field: A Study in Collocational Resonance.....	59
<i>Janet DeCesaris, Geoffrey Williams</i>	
The use of corpora in bilingual phraseography	59
<i>Dmitrij Dobrovol'skij</i>	
Comparing Phraseologisms: Building a Corpus-Based Lexicographic Resource for Translators	60
<i>Laura Giacomini</i>	
Lexical Variation within Phraseological Units	60
<i>Tarja Riitta Heinonen</i>	
<i>Prendere il toro per le corna o lasciare una bocca amara?</i> – The Treatment of Tripartite Italian Idioms in Monolingual Italian and Bilingual Italian-English Dictionaries	61
<i>Chris Mulhall</i>	
Especialización y Prototipicidad en Binomios N y N	61
<i>Ignacio Rodríguez Sánchez</i>	
Syntax and Semantics vs. Statistics for Italian Multiword Expressions: Empirical Prototypes and Extraction Strategies	61
<i>Luigi Squillante</i>	
Historical Lexicography and Etymology	63
Il DiVo (<i>Dizionario dei Volgarezzamenti</i>). Un archivio digitale integrato per lo studio del lessico di traduzione nell'italiano.....	65
<i>Diego Dotto</i>	
Informatiser le <i>Französisches etymologisches Wörterbuch</i> : la nécessaire prise en compte de l'utilisateur.....	65
<i>Pascale Renders, Esther Baiwir</i>	
A morphological historical root dictionary for Portuguese.....	65
<i>João Paulo Silvestre, Alina Villalva</i>	

Lexicological Issues of Lexicographical Relevance	67
What can lexicography gain from studies of loanword perception and adaptation?	69
<i>Miroslaw Bańko, Milena Hebal-Jezierska</i>	
Pejorative Language Use in the Satirical Journal "Die Fackel" as documented in the "Dictionary of Insults and Invectives"	69
<i>Hanno Biber</i>	
The Presence of Gender Issues in Spanish Dictionaries	69
<i>Ana Costa Pérez</i>	
Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes	70
<i>Václava Kettnerová, Markéta Lopatková</i>	
Polysemous Models of Words and Their Representation in a Dictionary Entry	71
<i>Tinatin Margalitadze</i>	
One lexicological theory, two lexicographical models and the pragmatemes	71
<i>Lena Papadopoulou</i>	
Analyzing Specialized Verbs in a French-Italian-English Medical Corpus: A Frame-based Methodology	72
<i>Anna Riccio</i>	
Neoclassical Formatives in Dictionaries.....	72
<i>Pius ten Hacken, Renáta Panocová</i>	
Reports on Lexicographical and Lexicological Projects.....	73
Revision and digitisation of the early volumes of <i>Norsk Ordbok</i> : Lexicographical challenges	75
<i>Surla Berg-Olsen, Åse Wetås</i>	
A dictionary guide for Web users	75
<i>Valeria Caruso, Anna De Meo</i>	
What a multilingual loanword dictionary can be used for: searching the Dizionario di italianismi in francese, inglese, Tedesco (DIFIT)	76
<i>Matthias Heinz, Anne-Kathrin Gärtig</i>	
The Basic Estonian Dictionary: the first monolingual L2 learner's dictionary of Estonian	76
<i>Jelena Kallas, Maria Tuulik, Margit Langemets</i>	
Creating a German-Basque electronic dictionary	76
<i>David Lindemann</i>	
Die fremdsprachige Produktionssituation im Fokus eines onomasiologisch-konzeptuell orientierten, zweisprachig-bilateralen Wörterbuches für das Sprachenpaar Deutsch - Spanisch:Theoretische und methodologische Grundlagen von DICONALE	77
<i>Meike Meliss</i>	
Graph-Based Representation of Borrowing Chains in a Web Portal for Loanword Dictionaries	77
<i>Peter Meyer</i>	
The Problem of Scale. Handling (More) Data with the <i>Wiki Lexicographica</i>	78
<i>Bruno Bon, Krzysztof Nowak</i>	
At the Beginning of a Compilation of a New Monolingual Dictionary of Czech (A Report on a New Lexicographic Project).....	78
<i>Pavla Kochová, Zdeňka Opavská, Martina Holcová Habrová</i>	
Frame Semantics and Learner's Dictionaries: Frame Example Sections as a New Dictionary Feature.....	78
<i>Carolin Ostermann</i>	
Translating action verbs using a dictionary of images: the IMAGACT ontology	79
<i>Alessandro Panunzi, Irene De Felice, Lorenzo Gregori, Stefano Jacoviello, Monica Monachini, Massimo Moneglia, Valeria Quochi, Irene Russo</i>	
Degrees of Synonymity as the Basis of a Network for German Communication Verbs in the Online Reference Work <i>Kommunikationsverben</i> in OWID.....	79
<i>Kristel Proost, Carolin Müller-Spitzer</i>	
Job-hunting in Italy: Building a glossary of "English-inspired" job titles.....	80
<i>Virginia Pulcini, Angela Andreani</i>	
A small dictionary of life under communist totalitarian rule (Czechoslovakia 1948-1989)	80
<i>Věra Schmiedtová</i>	
A Frequency Dictionary of Dutch.....	80
<i>Carole Tiberius, Tanneke Schoonheim, Adam Kilgarrieff</i>	
The Corpus of the Croatian Church Slavonic Texts and the Current State of Affairs Concerning the Dictionary of the Croatian Redaction of Church Slavonic Compiling	81
<i>Vida Vukoja</i>	
Others.....	83
Considerations about Gender Symmetry in the Dictionary of Bavarian Dialects in Austria	85
<i>Isabella Flucher, Eveline Wandl-Vogt, Thierry Declerck</i>	
Advancing Search in the Algemeen Nederlands Woordenboek.....	85
<i>Carole Tiberius, Jan Niestadt, Lut Colman, Boudewijn van den Berg</i>	
List of authors	87

Plenary Lectures

From lexicography to Terminology: a Cline, not a Dichotomy

Thierry Fontenelle

Translation Centre for the Bodies of the European Union

Thierry.fontenelle@cdt.europa.eu

Abstract

In a paper presented at the Euralex 2012 conference, ten Hacken (2012) discusses the OED's problematic claim to be the "definitive record of the English language". He argues that what distinguishes the OED from other dictionaries is the information it provides about English words and the range of problems this information can be used to solve. Dictionaries are not descriptions of a language, he claims, but tools with which users of the dictionary solve problems of a particular type. The nature of the dictionary therefore determines which types of problems it can solve.

In this paper, I would like to extend the parallel made by ten Hacken between general dictionaries, learners' dictionaries and historical dictionaries such as the OED to what is traditionally perceived as a dichotomy, namely the distinction between dictionaries and terminological databases. Instead of viewing term bases as a totally distinct type of linguistic product, I would like to argue that they should rather be seen as a specific kind of tool which provides information that specific users will use in order to solve specific linguistic problems, usually related to translation. In the course of their careers, translators will indeed need to make use of a whole range of dictionaries, starting from learners' dictionaries when they learn foreign languages, to monolingual dictionaries and bilingual dictionaries to learn translation techniques, to term bases as soon as they start translating highly specialized and technical texts. We will focus on terminology databases such as IATE, the European Union's interinstitutional term base, which is the natural tool to which they turn to obtain information about technical terms in the medical, legal, environmental, chemical fields, to cite only a few domains covered by this resource. With 8.7 million terms covering the 24 official languages of the European Union, including 1.4 million English terms and half a million abbreviations, this database is a highly popular tool in the translation world (44 million queries in 2013).

In addition to general term bases such as IATE, we will also discuss other specialized EU terminological databases such as ECHA-Term, a term base compiled for the European Chemicals Agency (ECHA) to help industry comply with the legal requirements of the REACH Directive and of the regulation on the classification, labeling and packaging of chemicals. We will show how user requirements have been taken into account to meet the needs of the users of the database, who resort to ECHA-Term to obtain reliable, coherent and up-to-date multilingual terminology in the chemicals field, a *sine qua non* for clear specialized communication.

The description of these databases will make it clear that the distinction between 'traditional' lexicography and terminology is more a cline than a dichotomy, insofar as the types of linguistic information included in the respective products created by both disciplines all correspond to the specific needs of their users.

Language of Presentation: English

Sat. 19/7/2014 - AUDITORIUM

Natural Language Processing Techniques for Improved User-Friendliness of Electronic Dictionaries

Ulrich Heid

Universität Hildesheim
heidul@uni-hildesheim.de

Abstract

There is no doubt that electronic dictionaries should be user-friendly. Lexicographic approaches inspired, for example, by the Function Theory of Lexicography (Tarp 2008, etc.) place the user and his/her needs in the center of their discussions about dictionary design.

In this talk, we aim at reviewing a number of proven to work techniques from Natural Language Processing, with respect to their impact on the user friendliness of electronic dictionaries. Our impression is that the potential of NLP tools used as components of online lexical information systems has not yet fully been exploited. We thus intended to discuss possibilities for combining dictionary resources with processing tools, to enhance the user friendliness of the integrated product.

We start from a brief recall of basic assumptions of the Lexicographic Function Theory, using a simplified model of reception and production oriented tasks as a background for the subsequent discussion of techniques from NLP. We then focus on aspects of data representation, and on comparatively simple, mainly symbolic tools and procedures that support users in text reception or text production tasks, as well as in bilingual tasks.

As dictionaries are seen by most users as authoritative reference works, we do not address approximative techniques, but concentrate on those approaches which can produce an assessment of the quality of their output. In this sense, reliability is a major issue, and the tools should either produce correct results or (at least) warn the user about the unverified status of their output (as it is done, for example, within the word formation analyzer available for German on the website of canoo.com).

An important prerequisite for language processing tools is an appropriate representation of the lexical data they use: this includes in particular a fine-grained categorization of the different types of lexic(ographic)al data. Similarly, also lexicographers interested in providing appropriate data for a given dictionary function can profit massively from a fine-grained classification and description of different types of lexicographical data. This view has gained more wide-spread acceptance over the last few years. We will discuss it in the framework of a model of the function-theoretical approach which assumes a central lexicographical data collection and a set of filters to select, order and present lexicographical data for different target user groups and/or dictionary functions. In this context, we

will report about the experience of publishers who use similar concepts for dictionary production.

We then address techniques for text reception dictionaries. Such NLP techniques mainly serve as access tools for the user; a prime example is inflectional morphology, which allows the user to insert (or, in an interactive dictionary related with text reading tools, to highlight) an inflected item and to get access to the appropriate lemma entry. We also discuss tools for word formation analysis and possibilities to include them into dictionary access devices. Another type of access tool that allows users to efficiently query the dictionary are devices that can identify idioms: if a user point this mouse to an item that is a part of an idiom, the tool verifies the idiom status of the full expression and displays only the dictionary entry of the idiom, and not that of the compositional meaning of the targeted item (cf. Seretan/Wehrli 2013). We assess to what extent such functions can be extended beyond idioms.

In terms of support for text production, inflection paradigms and/or inflection generation are relatively standard; we also think that links between dictionary and corpus, as they have been much discussed over the last few years, are very useful for text production, if combined with NLP techniques, for example to select appropriate examples for syntactic constructions of a given type. In a similar way, example selection from bilingual corpora may be a way forward towards tailor-made support for the verification of users' translation hypotheses.

This overview may help to identifying domains where the cooperation between lexicographers and NLP researchers and/or language technologists may lead to considerable improvements in the user friendliness of electronic dictionaries.

Language of Presentation: English

Tue, 15/7/2014 – AUDITORIUM

Using Mobile Bilingual Dictionaries in an EFL Class

Carla Marellò
Università degli Studi di Torino
carla.marellò@unito.it

Abstract

Are reference skills of native digital EFL students developed enough to take advantage of having a big bilingual dictionary on their smartphones? What would happen if they were not used to looking up words in a bilingual dictionary or they were not trained to?

To carry out this research a class from an Italian technical high school was observed. Students aged 17 were split into three groups of users and were given different versions of the same bilingual dictionary, the Ragazzini Italian and English dictionary (Zanichelli Bologna last edition 2013). The first group was allowed to use the Android app, the second was given access to the online version on the web portal ubidictionary.zanichelli.it, the third group received paper copies of the same dictionary.

First students were asked to answer some questions about their (un)familiarity with Italian monolingual dictionaries, English monolingual learner dictionaries, bilingual dictionaries. Then the three groups of users faced the same activities during a two hour-English-lesson. Students' self-confidence in what they thought to be right was encouraged in order to avoid too much test-induced dictionary referral.

Participants were observed while performing the activities and finally interviewed about their experience.

The case study will report similarities and differences in their performances, trying to check whether different supports and access to digital versions was determinant for success and rapidity in answering questions. Linguistic proficiency proved more determinant.

Students were also asked to report on the main difficulties they had to overcome when they looked words up in the dictionaries and were invited to suggest possible improvements in the way pieces of information were displayed on the mobile digital version.

Language of Presentation: English

Fri, 18/7/2014 - AUDITORIUM

Meanings, Ideologies, and Learners' Dictionaries

Rosamund Moon
University of Birmingham
r.e.moon@bham.ac.uk

Abstract

Learners' dictionaries aim to give clear indications of meaning, but not all aspects of meaning can be neatly packaged up in definitions – even without the constraints of a restricted defining vocabulary. It is difficult, for example, to deal effectively with connotation and culture-bound meanings; to show evaluations clearly and consistently; or to explain even-handedly the ideological undercurrents of items, particularly where these relate to sensitive issues such as race/ethnicity, gender, sexual orientation, religion, etc. Learners' dictionaries over the last fifty years or so have made strenuous efforts to ensure that they treat such items appropriately and with care. But the changing mores and attitudes of the culture in which these dictionaries are written – specifically here the community of native-speakers of British English – may be at odds with the mores and attitudes of the markets and readerships to which the texts are presented: an interesting tension between language, lexicography, and receiver.

The aim of this paper, then, is to explore ideological meanings in learners' dictionaries. Racism and ethnocentricity, sexism and homophobia are obvious areas to discuss. Less discussed, however, is age and ageism: something evident not just in adjectives and nouns which reference age directly, but also many other items which embed or entail notions of age covertly or indirectly. This will be the particular focus of my paper, where, as a case study, I will be examining the lexicographical treatment of ways in which younger and older people are represented linguistically and lexicographically, not only in conventionally written dictionaries – the British big five – but in crowd-sourced texts too.

Language of Presentation: English

Wed, 16/7/2014 – AUDITORIUM

The Dictionary-Making Process

The Making of a Large English-Arabic/Arabic-English Dictionary: the Oxford Arabic Dictionary

Tressy Arts
Oxford Arabic Dictionary
tressy.arts@gmail.com

Abstract

In this presentation, we will illustrate the process of making our brand-new Arabic-English/English-Arabic dictionary, which is due for publication in print and online in August 2014. It is intended for speakers of both English and Arabic. It contains over 26,000 entries on each side, including many up-to-the-minute words and expressions. Collocations and examples are an important feature. The dictionary has been compiled using dictionary writing software that enables editors to work and communicate with one another regardless of their location. It will be available in both print and online.

We show the entire process of making an Arabic dictionary, from finding a reliable framework in both languages, to developing a unique online functionality. We show the difficulties lexicographers face when compiling an Arabic dictionary, and the ways in which we dealt with those. In addition, the Oxford Arabic Dictionary has quite a few features that are entirely new to Arabic dictionaries, and we illustrate how we went about developing those.

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

Simple and Effective User Interface of the Dictionary Writing System

Kamil Barbierik, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Miroslav Virius
Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i.
kamil.barbierik@jfifi.cvut.cz, dengeova@ujc.cas.cz, holcova@ujc.cas.cz,
vladimir.jary@foxcom.cz, tomas.liska@foxcom.cz, liskova@ujc.cas.cz, miroslav.virius@jfifi.cvut.cz

Abstract

A new monolingual dictionary of the contemporary Czech language is being prepared at the Institute of the Czech language of the Academy of Sciences of the Czech Republic, v.v.i. As part of this project supported by the Ministry of Culture of the Czech Republic grant within the National and Cultural Identity (NAKI) applied research program a dictionary writing software is being developed. We shortly present the overall architecture of the software and then we focus on its user interface and one new module – the editorial tool..

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

Simple and effective user interface of our new DWS

Kamil Barbierik, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Miroslav Virius
Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i.
kamil.barbierik@foxcom.cz, dengeova@ujc.cas.cz, vladimir.jary@foxcom.cz, tomas.liska@foxcom.cz,
liskova@ujc.cas.cz, miroslav.virius@jfifi.cvut.cz

Abstract

The Department of Contemporary Lexicology and Lexicography of the Institute of the Czech Language of the Academy of Sciences of the Czech Republic, v. v. i., is preparing a new monolingual dictionary of contemporary Czech since 2012. Its working title is Akademický slovník současné češtiny (The Academic Dictionary of Contemporary Czech). It is a medium-sized dictionary with the expected number of 120,000–150,000 lexical units. As a support for this project, a new Dictionary Writing System (DWS) is developed. On this poster we introduce the user interface of the new DWS for the first time.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

The Dictionary Writing System for the Academic Dictionary of the Contemporary Czech

Kamil Barbierik, Zuzana Děngeová, Vladimír Jarý, Tomáš Liška, Michaela Lišková, Miroslav Virius
Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i.
kamil.barbierik@foxcom.cz, dengeova@ujc.cas.cz, vladimir.jary@foxcom.cz, tomas.liska@foxcom.cz,
liskova@ujc.cas.cz, miroslav.virius@fjfi.cvut.cz

Abstract

We demonstrate the dictionary writing system, which is developed for the new monolingual Academic Dictionary of Contemporary Czech. The demonstration covers main parts of the system – a list of entries, an entry detail and related editor's module as well as other important integrated features.

Language of Presentation: English

Tue 15, Wed 16, Thu 17/7/2014 – Room 8

Totalitarian Dictionary of Czech

František Čermák
Institute of the Czech National Corpus, Charles University
frantisek.cermak@ff.cuni.cz

Abstract

Due to a decades-long experience with the Communist totalitarian regime, an attempt was made to capture main features of its language, and thus the substance of the time, by a corpus-based dictionary. The corpus, based on large samples of Communist newspapers and booklets, which is now accessible on the web, served as basis for a frequency-based dictionary of the time; it has been provided with statistics according to time periods, with pertinent collocations, appearance in various genres and, most of all, with comparative figures with today's standard as it is to be found in a large contemporary corpus of Czech. This, by far the first systematic description of the period in any language, is complemented by a detailed study of the lexis of the time, many typical excerpts, etc.

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

Dictionary of Abbreviations in Linguistics: Towards Defining Cognitive Aspects as Structural Elements of the Entry

Ivo Fabijanić
English Department - University of Zadar, Croatia
ivo.fabijanic@unizd.hr

Abstract

The first part of the article deals with general information about the project of compiling a dictionary of abbreviations in linguistics. It also contains a short overview of past research together with their main results. So far, some specific theoretical and practical solutions were proposed. Theoretical solutions refer to the Multi-level approach in collecting data for submorphemic word-formations, which consists of three aspects: 1) *Structure and Modes of Production*, 2) *Cognitive Aspects*, and 3) *Functional Aspects*. Practical solutions for the structure and modes of production have already been recommended with the results properly substantiated by the examples of abbreviations. The second part presents results of the analysis for the cognitive aspect of the multi-level approach. The semantic part reveals the relationship between an abbreviation type and the semantic (sub-)field it might be assigned to. Semiotic part is achieved by designating a specific interpretation of a sign's meaning, while the analysis of lexicalization and institutionalization confirms their contingent addition to this scientific lexicon. The motivational aspect takes two conceptual perspectives in consideration: the narrower and broader senses of the formation, and various semantic relationship patterns which led to the classification of fully and partially motivated abbreviations.

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

La definizione delle relazioni intra- e interlinguistiche nella costruzione dell'ontologia IMAGACT

Gloria Gagliardi
Università degli Studi di Firenze
gloria.gagliardi@unifi.it

Abstract

IMAGACT è un'ontologia interlinguistica che rende esplicito il *range* di variazione pragmatica associata ai predicati azionali a media ed alta frequenza in italiano ed inglese. Le classi di azione che rappresentano le entità di riferimento dei concetti linguistici, indotte da *corpora* di parlato da linguisti madrelingua, sono rappresentate in tale risorsa lessicale nella forma di scene prototipiche (Rosch 1978). Tale metodologia sfrutta la capacità dell'utente di trovare somiglianze tra immagini diverse indipendentemente dal linguaggio, sostituendo alla tradizionale definizione semantica, spesso sottodeterminata e linguo-specifica, il riconoscimento e l'identificazione dei tipi azionali. L'articolo illustra i criteri generali che hanno ispirato il *mapping* inter-/intra- linguistico dei dati derivati da *corpora* per la formazione dell'ontologia, le questioni di natura teorica e tecnica poste dalla costruzione della risorsa e le soluzioni adottate. Vengono descritte le tipologie e la natura delle relazioni tra le entità del database nella sua versione 1.0, e le modalità generali con cui i materiali linguistici annotati sono stati organizzati in una struttura dati coerente.

Language of Presentation: Italian

Thu, 17/7/2014 – Corridor seminar rooms

I neologismi pianificati: esempi dal ladino fassano nell'ultima edizione del DILF

Michela Giovannini
Università di Trento
michela.giovannini@lett.unitn.it

Abstract

Il ladino fassano è una varietà del ladino centrale parlato nella valle di Fassa in provincia di Trento e raggruppa tre sottovarietà: *cazet* (alta valle di Fassa), *brach* (centro valle), *moenat* (bassa valle). In questo *poster* verrà presentato il lessico neologico della nuova edizione del *Dizionario italiano-ladino fassano/Dizionèr talian-ladin fassician* (2013) e verrà focalizzata l'attenzione sui prestiti (adattati e non adattati) e sui calchi (semantici e sintattici) con l'obiettivo di valutare la vitalità di questa varietà di minoranza. Benché l'obiettivo primario del lessico neologico del DILF puntasse al recupero di parole del lessico patrimoniale ladino, si constaterà come l'acquisizione di forestierismi, prestiti o calchi sia risultata imprescindibile, in particolare per l'accettabilità, la fortuna della forma e la settorialità dei termini. L'elenco di nuovi lemmi evidenzierà inoltre il fenomeno della convivenza tra due forme che solo il cambiamento linguistico *in fieri* selezionerà nel tempo. La spinta conservativa di un lemma costruito con elementi lessicali propri del ladino fassano (ad esempio *chech/banca de dac/douc/elaborator eletronich/desciarièr*) e la spinta innovativa della forma concorrente dello stesso concetto, per lo più un prestito non adattato dall'inglese o dal francese (rispettivamente: *chef/database/dessert/computer/download*), produrranno una convivenza disequilibrata tra le due forme (alcune già producono: ad esempio *database* – nel lessico spontaneo –, *banca de dac* – nel lessico pianificato –).

Language of Presentation: Italian

Thu, 17/7/2014 – Corridor seminar rooms

DPS – single source production system

Holger Hvelplund
IDM, Paris, France
hvelplund@idm.fr

Abstract

DPS CMS is a single source production platform for producing and publishing content in print and digital products.

To produce and publish content products for many delivery channels an efficient single source production and publishing system combined with automated content delivery processes is a must. In DPS CMS the content source is neutral to formats used by delivery channels - it means that from planning and all the way up to publishing the content is treated in a delivery channel neutral way. The content is the same no matter if it should be published in print, e-book or online format.

DPS Process is powerful tool for: centralized management of data processing scripts; automatic and secure formatting and export of content from DPS CMS (via DPS CMS REST API) to delivery channel specific formats; and import of legacy content to DPS CMS (DPS CMS REST API). As new delivery channels are emerging continuously, the strict separation of content in DPS CMS from delivery and formatting of the content in delivery channel specific formats handled by DPS Process makes your content versatile enough to publish to multiple delivery channels and enables new delivery channels to be added with ease.

Language of Presentation: English

Tue 15, Wed 16/7/2014 – Room 8

A Dictionary of Old Norse Prose and its Users — Paper vs. Web-based Edition

Ellert Thor Johannsson, Simonetta Battista
A Dictionary of Old Norse Prose, University of Copenhagen
nk950@hum.ku.dk, sb@hum.ku.dk

Abstract

Ordbog over det norrøne prosasprog/A Dictionary of Old Norse Prose (ONP) is a historical dictionary project at the Department of Scandinavian Research at the University of Copenhagen and covers medieval Old Norse material found in prose texts from the oldest written documents to early modern times (Old Icelandic 1150-1540 and Old Norwegian 1150-1370). After the publication of the first four volumes of an intended thirteen volume printed edition, a decision was made to change the format of the dictionary to a digital edition available online. This new medium has had to accommodate the already published printed material as well as unprinted and not fully edited material. In this article, we discuss how the change in ONP's format has provided the user with some benefits in working with the dictionary material, but also some new challenges. We compare the features of the printed volumes to the features of the online version and in doing so address some key questions that relate to the perspective of the user and how the digital version can be further improved.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Making a Learner's Dictionary of Academic English

Diana Lea
Oxford University Press
diana.lea@oup.com

Abstract

This paper gives an account of the development of the *Oxford Learner's Dictionary of Academic English*, a dictionary for non-native-English-speaking students who are studying academic subjects at tertiary level through the medium of English. First, a corpus of academic English was created, using high-quality texts from a broad range of disciplines, maintaining a balance between textbooks, containing typical student reading material, and journal articles, modelling expert academic writing. Drawing on the corpus, and on previous research in the field, a core headword list of "general academic vocabulary" was drawn up. This list was continually supplemented with necessary defining words, complements, collocates, synonyms and antonyms of these words as the work of compiling the dictionary entries progressed. In compiling the entries, particular challenges were encountered in reconciling the academic and pedagogic requirements of the dictionary. This can be seen, for example, in decisions about sense division and the wording of definitions and also in the selection of example sentences from the corpus. Editors had to find ways to represent academic language faithfully whilst making it accessible to learners. The result is a genuinely academic learner's dictionary that should offer real help to learners with their academic writing.

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

The Danish Thesaurus: Problems and Perspectives

Sanni Nimb, Lars Trap-Jensen, Henrik Lorentzen
Society for Danish Language and Literature
sn@dsl.dk, ltj@dsl.dk, hl@dsl.dk

Abstract

In this paper, we present a new thesaurus for Danish and discuss some of the problems and decisions that the compilers have been faced with. The thesaurus is compared to other thesauri: Roget, Dornseiff and particularly to its smaller Danish predecessor Andersen. The different steps in the compilation process are outlined, with special attention being devoted to word ordering at the lowest level (alphabetical vs. semantic) and to the use of stylistic labels. In a comprehensive thesaurus, the conclusion is that semantic ordering is more useful for the user and that stylistic labels are necessary.

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

From a Dialect Dictionary to an Etymological One

Vilja Oja, Iris Metsmägi
Institute of the Estonian Language
viljaoja@eki.ee, iris.metsmagi@eki.ee

Abstract

Typically, loanwords from different sources are presented in different entries of a dialect dictionary, but the etymologies of the dialect words are often obscure. An etymologist, on the other hand, has to consider the phonetic shape and developments of a stem, its areal distribution and meanings in dialects to find out the etymology. What are the cooperative prospects of etymologists and dialectologists? The paper compares the presentation of dialect words in two dictionaries currently compiled on the web, namely, in the Estonian dialect dictionary and in the Estonian etymological dictionary. Despite the different specifics of the two dictionaries a number of similar problems have cropped up. Also, comparisons of the material have yielded essential information enabling solutions for both sides. Across dialects, a loanword often displays numerous phonetic variants, while a variant may easily sound untypical of the concrete dialect. The possible donor can be traced considering the occurrence of the dialect word in the traditional area of loanwords of a certain origin and the semantic relationship of the word with the presumed donor language. Cooperation between etymologists and dialectologists has contributed a lot to making a distinction between homonymy and polysemy, to identification of folk etymologies (words as well as semantic nuances), to distinguishing between separate loanwords and derivatives of the same stem, etc. A shared electronic environment enables bilateral specification of the linguistic material, if necessary.

Language of Presentation: English

Fri, 18/7/2014 – Room 2-3

Research on Dictionary Use

On the presentation and placement of collocations in monolingual English learners' dictionaries

Anna Dziemianko
Adam Mickiewicz University
danna@wa.amu.edu.pl

Abstract

The study examines the role of presentation and positioning of collocations in encoding and retention. Two factors, the place of collocations (entry-initial and entry-final) and their presentation (box, bold before examples, bold within examples), were systematically manipulated in an empirical investigation in which purpose-built monolingual dictionary entries were employed. The research reveals that both variables significantly influence the use of collocations. The retention of collocations, in turn, remains largely unaffected by their place in the microstructure, but proves to be dependent on the adopted method of presentation. Recommendations for optimizing the treatment of collocations in learners' dictionaries conclude the paper.

Language of Presentation: English

Thu, 17/7/2014 - Room 1

Wörterbuchbenutzung: Ergebnisse einer Umfrage bei italienischen DaF-Lernern

Carolina Flinz
Universität Pisa
c.flinz@ec.unipi.it

Abstract

Die vorliegende empirische Untersuchung befasst sich mit einer Umfrage zur Wörterbuchbenutzung bei 41 Studentinnen und Studenten des Dipartimento di Filologia, Letteratura e Linguistica der Universität Pisa, dasselbe Department, an dem auch das deutsch-italienische sprachwissenschaftliche Online-Wörterbuch DIL erarbeitet worden ist (vgl. Flinz: 2011). Die schriftliche Umfrage wurde in Anlehnung an Hartmanns 5. Hypothese „An analysis of users' needs should precede dictionary design“ (1989) durchgeführt. Die wichtigsten Ergebnisse waren von großer Bedeutung für die Gestaltung der makro- und mikrostrukturellen Eigenschaften des Fachwörterbuches. Die Ergebnisse der Untersuchung und die daraus folgenden Reflektionen werden in thematischen Kernblöcken vorgestellt.

Language of Presentation: German, Italian

Thu, 17/7/2014 - Room 1

Translation, cultural adaptation and preliminary psychometric evaluation of the English version of “Strategy Inventory for Dictionary Use” (S.I.D.U)

Zoe Gavriilidou
Democritus University of Thrace
zoegab@otenet.gr

Abstract

The present paper reports results regarding the adaptation in English of the Strategy Inventory for Dictionary Use (S.I.D.U.) and the preliminary psychometric evaluation of the English version. S.I.D.U is a 36-item self-report questionnaire for assessing dictionary use strategies specifying four main areas of interest: a) dictionary use awareness skills, b) strategies for dictionary selection and acquaintance with dictionary conventions, c) lemmatization strategies, and finally d) look-up strategies. The original scale was translated from Greek into English, back-translated and reviewed. Cross-cultural adaptation included the experts' revision followed by its administration to 52 participants. Its internal consistency was .89. Similarly all four subscales showed good to excellent internal consistency (dictionary use awareness: $\alpha=.83$, dictionary selection and acquaintance: $\alpha=.76$, lemmatization: $\alpha=.86$ and dictionary search: $\alpha=.78$). Test-retest reliability ranged from fair to good for the total scale and its six-subcales.

Language of Presentation: English

Thu, 17/7/2014 - Room 1

The Authentic Voices of Dictionary Users – Viewing Comments on an Online Learner’s Dictionary Before and After Revision

Ann-Kristin Hult
Department of Swedish, University of Gothenburg
ann-kristin.hult@svenska.gu.se

Abstract

This paper deals with comment field data from two web questionnaire surveys, performed in 2007 and 2011, about the use and users of the *Swedish Lexin Dictionary* (SLD), an online monolingual learner’s dictionary. The SLD underwent comprehensive revision in between the two studies. In order to evaluate the update I compare the respondents’ comments on SLD before and after the revision. The two sets of comment field data are categorised by the types of comments expressed and by the information categories mentioned. As it turns out, respondents do seem happier with the new version of SLD. Generally, there are more positive comments after the revision than before and the information categories mentioned are more often set in positive contexts in the latter data. In addition, the majority of respondents in 2011 belong to the dictionary’s target group, which was not the case in 2007. Although respondents seem more satisfied with the new version of SLD, the two sets of comment field data contain largely the same kind of criticism, as for example in comments concerning (usually the lack of) lemmas and examples.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Mobile Lexicography: A Survey of the Mobile User Situation

Henrik Køhler Simonsen
Copenhagen Business School
hks.ibr@cbs.dk

Abstract

Users are already mobile, but the question is to which extent knowledge-based dictionary apps are designed for the mobile user situation. The objective of this article is to analyse the characteristics of the mobile user situation and to look further into the stationary user situation and the mobile user situation. The analysis is based on an empirical survey involving ten medical doctors and a monolingual app designed to support cognitive lexicographic functions, cf. (Tarp 2006:61-64). In test A the doctors looked up five medical terms while sitting down at a desk and in test B the doctors looked up the same five medical terms while walking around a hospital bed. The data collected during the two tests include external and internal recordings, think-aloud data and interview data. The data were analysed by means of the information scientific star model, cf. (Simonsen 2011:565), and it was found that the information access success of the mobile user situation is lower than that of the stationary user situation, primarily because users navigate in the physical world and in the mobile device at the same time. The data also suggest that the mobile user situation is not fully compatible with for example knowledge acquisition.

Language of Presentation: English

Thu, 17/7/2014 - Room 1

Kollokationsübungen für den Unterricht als Wörterbuchergänzung

Stefanie Meier, Alessandra Widmer
Universität Basel
stefanie.meier@unibas.ch, alessandra.widmer@stud.unibas.ch

Abstract

Im Deutschunterricht sind der Begriff und das Konzept der Kollokation noch wenig bekannt. Dies steht im Gegensatz zum englischen Sprachraum, wo feste Wortverbindungen/Kollokationen unter dem Begriff *collocations* als wichtiger Teil des Wortschatzes anerkannt sind und im Unterricht auch gelehrt werden. Es ist ein Desiderat der Phraseologieforschung und -didaktik, dass dieser Missstand aufgehoben wird. Das an der Universität Basel erarbeitete und 2014 erscheinende Wörterbuch „Feste Wortverbindungen des Deutschen. Kollokationswörterbuch für den Alltag“ mit einer stark nutzungsfreundlichen Ausrichtung versucht, diese Lücke zu schliessen. Das Wörterbuch beschränkt sich auf einen Grundwortschatz von rund 2'000 Wörtern und umfasst über 90'000 Kollokationen. Die Auswahl der Kollokationen erfolgte durch die eingehende lexikographische Bearbeitung von korpusbasierten Kookkurrenzlisten.

Damit Kollokationen auch tatsächlich Eingang in den Unterricht finden, wurde ein Begleitheft mit Kollokationsübungen erarbeitet, das sich an Muttersprachler_innen und DaF-Lernende richtet. Das Poster stellt dieses Beiheft zur neuen Publikation vor. Viele der aktuellen Lehrmaterialien haben keinen Fokus auf Kollokationen oder erwähnen sie allenfalls zusammen mit Routineformeln. Die Informationstexte und Übungen im Beiheft sorgen für mehr Klarheit, was Begriff und Konzept der Kollokation angeht und unterstützen Lernende und Lehrende beim Einüben und Nachschlagen von Kollokationen. Das Beiheft für Selbststudium und

Unterricht soll besonders auch für Lehrpersonen als nützliches Hilfsmittel dienen. Die abgebildeten Übungen können in den Sprachunterricht integriert werden und können auch als Grundlage dienen, eigene, ergänzende Übungen mit Fokus auf Kollokationen zu erstellen.

Language of Presentation: German

Thu, 17/7/2014 – Corridor seminar rooms

Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht

Martina Nied Curcio
 Università degli Studi Roma Tre
 martina.nied@uniroma3.it

Abstract

Die Benutzung von Smartphones, iPhones, iPads und Tablets im Fremdsprachenunterricht scheint den Studierenden den schnellen und unmittelbaren Gebrauch von Online-Wörterbüchern zu garantieren, sowie unbegrenzte Recherchemöglichkeiten zu bieten, so dass sprachliche Schwierigkeiten direkt überwunden werden können. Doch sieht es danach aus, als würden die Fremdsprachenstudierenden das Potential nicht ausnützen und sich auf zweisprachige Wörterbücher und Übersetzungsprogramme konzentrieren. Auch ihr Benutzerverhalten in Bezug auf Online-Wörterbücher scheint dem der Verwendung von Print-Wörterbüchern ähnlich zu sein. Eine empirische Untersuchung mit italienischen Studierenden der Germanistik zur Smartphone-Benutzung zeigt, wie und wofür sie hinsichtlich lexikalischer Fragen das Smartphone verwenden, welches Benutzerverhalten sie generell, aber auch in spezifischen Übersetzungsaufgaben, an den Tag legen und wie sie selbst über ihre Benutzung mit dem Smartphone bei bestimmten Schwierigkeiten reflektieren. Die Ergebnisse vermitteln erste Eindrücke und zeigen Trends auf, die Ausgangspunkt für weitere umfangreichere Forschungen sein können, die ihrerseits der Fremdsprachendidaktik und der lexikographischen Praxis wichtige Impulse geben können. Interessant in diesem Zusammenhang sind die „ricerche incrociate“ („cross research“) und die Konsultation von multilingualen Online-Wörterbüchern mit der Konsultation des Englischen als „Sandwichsprache“.

Language of Presentation: German

Thu, 17/7/2014 - Room 1

Do menus provide added value to signposts in monolingual dictionary entries?

Bartosz Ptasznik, Robert Lew
 Faculty of English, Adam Mickiewicz University in Poznań Affiliations
 rlew@amu.edu.pl

Abstract

Previous studies indicate that sense-guiding devices in the form of signposts and menus can assist users in selecting the relevant sense in a polysemous entry. Although two studies (Lew 2010; Nesi and Tan 2011) have found signposts introducing individual senses to be more effective than entry-initial menus, no study so far has assessed a combination of signposts and menus within the same entry. The present study compares entry consultation time and accuracy of sense selection in entries supplied with a combination of signposts and menus against signposts alone. Experimental data were collected from 118 intermediate Polish students of English. A linear mixed model analysis with planned comparisons was carried out on the data. Results indicate that adding menus to signposts brings no benefits and that signposts alone work best.

Language of Presentation: English

Thu, 17/7/2014 - Room 1

Dictionary users do look up frequent and socially relevant words. Two log file analyses.

Sascha Wolfer, Alexander Koplenig, Peter Meyer, Carolin Müller-Spitzer
Institut für Deutsche Sprache, Mannheim, Germany
wolfer@ids-mannheim.de

Abstract

We start by trying to answer a question that has already been asked by de Schryver et al. (2006): Do dictionary users (frequently) look up words that are frequent in a corpus. Contrary to their results, our results that are based on the analysis of log files from two different online dictionaries indicate that users indeed look up frequent words *frequently*. When combining frequency information from the Mannheim German Reference Corpus and information about the number of visits in the Digital Dictionary of the German Language as well as the German language edition of Wiktionary, a clear connection between corpus and look-up frequencies can be observed. In a follow-up study, we show that another important factor for the look-up frequency of a word is its temporal social relevance. To make this effect visible, we propose a de-trending method where we control both frequency effects and overall look-up trends.

Language of Presentation: English

Thu, 17/7/2014 - Room 1

La performance dell'utente apprendente di italiano LS/L2 e la microstruttura dei dizionari: sussidi per lo sviluppo della Lessicografia Pedagogica

Angela Maria Tenório Zucchi
Università di San Paolo, Brasile
angelazucchi@usp.br; angelatz@gmail.com

Abstract

Questo *paper* presenta dei risultati quantitativi e qualitativi di una ricerca empirica sulla comprensione di unità lessicali contestualizzate e l'uso dei dizionari, svolta con studenti brasiliani di lingua italiana come lingua straniera presso la Facoltà di Lettere dell'Università di San Paolo. Con una metodologia innovativa nell'usare immagini come risposta nelle alternative a scelta multipla e nell'organizzare i dati raccolti in particolari schede, questo esperimento ha reso possibile verificare statisticamente l'efficacia dell'uso dei dizionari, sia bilingue o monolingue, per la comprensione di parole di uso comune, ma a volte non presenti nei manuali didattici, presentate in testi scritti in italiano. Attraverso la scheda denominata FLICULAD, si è potuto controllare in modo sistematico quali sono state le informazioni contenenti la microstruttura di ogni unità lessicale analizzate che hanno contribuito o meno, secondo l'opinione dello studente, ad arrivare alla comprensione del significato di quella unità in quel contesto.

Language of Presentation: Italian

Thu, 17/7/2014 - Room 1

Lexicography and Language Technologies

ALiquot – Atlante della Lingua Italiana QUOTidiana

Michele Castellarin, Fabio Tosques
Humboldt-Universität zu Berlin
michele_castellarin@yahoo.it, ftosques@gmail.com

Abstract

Der folgende Beitrag erklärt und stellt die Methoden und Resultate dar, die im Projekt ALIQUOT (Atlante della Lingua Italiana QUOTidiana – Atlas der italienischen Alltagssprache) verwendet bzw. gewonnen wurden, um die italienische Alltagssprache zu untersuchen. Das Ziel des Projekts besteht darin, geolinguistische Karten zu erzeugen, die die enorme Vielfalt der regionalen und lokalen Varianten der italienischen Sprache im täglichen Gebrauch sichtbar machen. Die Daten für ALIQUOT werden auf der Basis der indirekten Methode mit einem Online-Fragebogen anonym erhoben. Nur wenige persönliche Daten (Alter, Geschlecht, Wohnort, Beruf, Bildungsgrad usw.) werden abgefragt. Mit Hilfe dieser Daten sollen spätere soziolinguistische Analysen ermöglicht werden. Sämtliche geolinguistische Karten werden online publiziert und sind für jeden frei zugänglich. Wir möchten der Forschung und Lehre mit ALIQUOT ein nützliches Tool in die Hand geben, welches sprachliche Phänomene und Variationen des Italienischen präzise darstellt. ALIQUOT ist damit der erste Sprachatlas, der exakte und umfangreiche Karten der italienischen Alltagssprache anschaulich präsentiert.

Language of Presentation: German

Wed, 16/7/2014 - Room 2-3

Applying a Word-sense Induction System to the Automatic Extraction of Diverse Dictionary Examples

Paul Cook¹, Michael Rundell², Jey Han Lau³, Timothy Baldwin¹

¹Department of Computing and Information Systems, The University of Melbourne

²Lexicography Masterclass and Macmillan Dictionaries

³Department of Philosophy, King's College London

paulcook@unimelb.edu.au, michael.rundell@lexmasterclass.com,
jeyhan.lau@gmail.com, tb@ldwin.net

Abstract

There have been many recent efforts to automate or semi-automate parts of the process of compiling a dictionary, including building headword lists and identifying collocations. The result of these efforts has been both to make lexicographers' work more efficient, and to improve dictionaries by introducing more systematicity into the process of their construction. One task that has already been semi-automated is that of finding good dictionary examples, and a system for this, GDEX, is readily available in the Sketch Engine. An ideal system, however, would be able to automatically retrieve candidate examples of a particular sense of a word, which is beyond the current scope of GDEX. In this paper, as a step towards this ambitious goal, we propose and evaluate a method for applying a 'word-sense induction' system to automatically extract examples that exhibit a greater diversity of usages of a target word than can currently be obtained through GDEX. We then discuss the future prospects for systems that are able to automatically select candidate dictionary examples for a particular word sense.

Language of Presentation: English

Sat, 19/7/2014 - Room 2-3

Cross-linking Austrian dialectal Dictionaries through formalized Meanings

Thierry Declerck, Eveline Wandl-Vogt
DFKI GmbH, Language Technology Lab, Germany;
Austrian Academy of Sciences, ICLTT, Austria
declerck@dfki.de, eveline.wandl-vogt@oeaw.ac.at

Abstract

This paper deals with the formalization of aspects of definitions used in dialectal dictionaries. We focus on the way “meanings” are encoded in such dictionaries, an essential position for many users and lexicographers, and describe how such an encoding can be re-used for cross-linking entries across on-line (dialectal) dictionaries. In this contribution we describe in some details experiments we made in this respect with two Austrian dialectal dictionaries: The dictionary of Bavarian dialects of Austria (“Wörterbuch der bairischen Mundarten in Österreich”, WBÖ) and the dictionary of the Viennese dialect (“Wörterbuch der Wiener Mundart”, WWM), which we ported into the SKOS and *lemon* models, in order to publish them in the Linguistic Linked Open Data cloud. We show how this approach is not only appropriate for supporting the automation of the cross-linking of dialectal dictionaries, but also for linking entries of the dialectal dictionaries to other types of lexical and encyclopaedic resources in the web.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

Nutzung des DWDS-Wortprofils beim Aufbau eines lexikalischen Informationssystems zu deutschen Stützverbgefügen

Jörg Didakowski, Nadja Radtke
Berlin-Brandenburgische Akademie der Wissenschaften,
Technische Universität Dortmund
didakowski@bbaw.de, nadja.radtke@tu-dortmund.de

Abstract

Im Hinblick auf eine Zuarbeit für die lexikographische Arbeit an einem lexikalischen Informationssystem zu deutschen Stützverbgefügen wird das DWDS-Wortprofil vorgeschlagen. Mithilfe dieses Werkzeugs kann eine zeitintensive und mühsame Recherchearbeit über eine Textsuchmaschine vermieden werden, indem auf Basis eines ausgewogenen Korpus potenzielle Stützverbgefüge bereitgestellt werden. Des Weiteren wird die Einbeziehung von Assoziationsmaßen vorgeschlagen, um die Menge des zu sichtenden Materials für die lexikographische Arbeit weiter reduzieren zu können.

Language of Presentation: German

Fri, 18/7/2014 – Room 7

Automation of Lexicographic Work Using General and Specialized Corpora: Two Case Studies

Iztok Kosem¹, Polona Gantar², Nataša Logar³, Simon Krek⁴
¹Trojina, Institute for Applied Slovene Studies
²Fran Ramovš Institute for the Slovene Language, ZRC SAZU, Ljubljana, Slovenia
³Faculty of Social Sciences, University of Ljubljana
⁴Jožef Stefan Institute, Ljubljana
iztok.kosem@trojina.si, apolonija.gantar@guest.arnes.si,
natasa.logar@fdv.uni-lj.si, simon.krek@guest.arnes.si

Abstract

Due to increasingly large amounts of authentic data to analyse, lexicographers are nowadays looking to language technologies to provide them with not only the tools to analyse the data, but also with tools and methods that ease and speed up the data analysis. One of the most promising avenues of research has been the automation of early stages of the corpus data analysis, with the aim to summarize, and consequently reduce, the amount of corpus data that the lexicographers need to examine. However, most of this research deals with general lexicography; terminology is yet to extensively test these methods. This paper attempts to address this gap by presenting two separate Slovene research projects, one lexicographic (Slovene Lexical Database) and the other terminological (Termis), that used the same method of automatic extraction of corpus data (presented in Kosem et al. 2013). After describing the projects and the corpora use, similarities and differences in the parameter settings and the quality of extracted data in the two projects are presented. We conclude with discussing the further potential of automation in both general and specialised lexicography.

Language of Presentation: English

Sat, 19/7/2014 – Room 1

A Corpus-based Dictionary of Polish Sign Language (PJM)

Jadwiga Linde-Usiekiewicz, Małgorzata Czajkowska-Kisil, Joanna Łacheta, Paweł Rutkowski
University of Warsaw
jlinde@uw.edu.pl, maczajkis@wp.pl,
j.lacheta@uw.edu.pl, p.rutkowski@uw.edu.pl

Abstract

The aim of this paper is to give a general overview of an on-going lexicographic project devoted to Polish Sign Language (PJM), a natural language used by the Deaf community in Poland. The project in question will result in the first-ever PJM dictionary based on extensive corpus data (encompassing more than 300 hours of video material recently collected by the Section for Sign Linguistics of the University of Warsaw). The present article discusses the most important assumptions and methodological foundations of the current PJM dictionary project and confronts them with previous work on PJM and Signed Polish glossaries, as well as with international standards in contemporary sign language lexicography. The design of the new PJM dictionary is discussed in detail, including the most problematic issues, such as lemmatization, sense division and sense description principles. Sample entries are given as illustration. It is important to note that apart from filling an important gap in the availability of sign language teaching and learning materials in Poland, the PJM dictionary outlined in this paper is also likely to further the recognition of PJM as a full-fledged natural language.

Language of Presentation: English

Thu, 17/7/2014 – Room 7

Laying the Foundations for a Diachronic Dictionary of Tunis Arabic: a First Glance at an Evolving New Language Resource

Karlheinz Mörth¹, Stephan Procházka², Ines Dallaji²

¹Institute of Corpus Linguistics and Text Technology (Austrian Academy of Sciences)

²Department of Oriental Studies (University of Vienna)

Karlheinz.Moerth@oeaw.ac.at, stephan.prochazka@univie.ac.at, ines.dallaji@univie.ac.at

Abstract

Arabic lexicography has a long tradition. However, at the time of writing this report, there exist only a very few digital products, let alone products documenting Arabic dialects. Our paper presents the TUNICO project (Linguistic Dynamics in the Greater Tunis Area) and digital language resources which are being produced as part of the project. The TUNICO working group is working on a digital diachronic dictionary of Tunis Arabic which is being compiled as part of a larger linguistic endeavour to document the variety of the Tunisian capital. One of the interesting features of the project is that it draws on a number of heterogeneous sources: text books, grammatical descriptions and a corpus of spoken youth language which is currently being compiled. In this project, the dictionary is used as an analytical tool, as a research instrument, by integrating the various sources into one new coherent language resource thus allowing researchers to gain unprecedented insights in material that partly has been available for quite some time.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

BabelNet meets Lexicography: the case of an automatically-built multilingual encyclopedic dictionary

Roberto Navigli
Sapienza University of Rome
navigli@di.uniroma1.it

Abstract

In this paper we provide a first study of the lexicographic quality of BabelNet, a very large automatically-created multilingual encyclopedic dictionary. BabelNet 2.0, available online at <http://babelnet.org>, covers 50 languages and provides both lexicographic and encyclopedic knowledge for all the open-class parts of speech. It is obtained from the automatic integration of several language resources, namely: WordNet, Open Multilingual WordNet, Wikipedia and OmegaWiki. Here we present a first analysis of the dictionary entries in terms of their coverage of English and Italian word tokens in a large corpus and in comparison to existing, well-established dictionaries, namely the Oxford Dictionary of English and the Treccani Italian dictionary. We observe that BabelNet contains most meanings of the frequent words under analysis and provides additional, often domain-specific meanings and their textual definitions unavailable in traditional dictionaries, as well as encyclopaedic coverage for those words.

Language of Presentation: English

Sat, 19/7/2014 – Room 1

Tracing Sentiments: Syntactic and Semantic Features in a Subjectivity Lexicon

Jana Šindlerová, Kateřina Veselovská, Jan Hajič, jr.
Charles University in Prague, Faculty of Mathematics and Physics
{sindlerova,veselovska,hajicj}@ufal.mff.cuni.cz

Abstract

In this paper we present a syntactic and semantic analysis of verbal entries in the Czech subjectivity lexicon Czech SubLex 1.0 concerning their semantic and valency properties with respect to the roots and degree of subjectivity and evaluativeness. We demonstrate that evaluative verbs share certain abstract syntactic patterns with valency positions encoding the position of the source and target of evaluative stance. These patterns then roughly correspond to semantic properties of the verbs. For example, verbs propagating sentiments to the Actor position usually describe events of destruction (negative sentiments), or progress (positive sentiments), or events of direct experiencing emotional states, whereas verbs propagating sentiments to the Addressee or Patient position usually describe events of taking and communicating a stance, stopping or eliminating, or praising. The analysis represents the first step towards enhancing the lexical database with syntactic and semantic features in order to suit the lexicon for the task of the detection of targets (and sources) within evaluative stances.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

A Simple Platform for Defining Idiom Variation Matching Rules

Koichi Takeuchi[#], Ulrich Apel[§], Ray Miyata[¶], Ryo Murayama[¶], Ryoko Adachi[¶], Wolfgang Fanderl[§], Iris Vogel[§], Kyo Kageura[¶]
[#]Okayama University, Japan
[§]Tübingen Eberhard Karls University, Germany
[¶]University of Tokyo, Japan
koichi@cl.cs.okayama-u.ac.jp, ulrich.apel@uni-tuebingen.de, ray.miyata@gmail.com,
utatanenohibi@gmail.com, littlemarmalade@gmail.com,
fanderl@japanologie.uni-tuebingen.de,
iris@fruchtfledermaus.de, kyo@p.u-tokyo.ac.jp

Abstract

In this demonstration, we present a system which enables people who are learning languages to define idiom variation matching rules, with special reference to variations created by insertion. The system is fully operational, currently providing Japanese idiom entries taken from Japanese-English and Japanese-German dictionaries, and is used by both graduate and undergraduate university students who are studying Japanese and Japanese native speakers.

Language of Presentation: English

Wed 16, Thu 17/7/2014 – Room 8

Lexicography and Corpus Linguistics

Compatible Sketch Grammars for Comparable Corpora

Vladimír Benko

Slovak Academy of Sciences, L. Štúr Institute of Linguistics
Comenius University in Bratislava, UNESCO Chair in Translation Studies
vladob@juls.savba.sk

Abstract

Our paper describes an on-going experiment aimed at creating a family of billion-token web corpora that could to a large extent deserve the designation “comparable”: corpora are of the same size, data gathered by crawling the web at (approximately) the same time, containing similar web-specific domains, genres and registers, further pre-processed, filtered and deduplicated by the same tools, morphologically annotated by (possibly) the same tagger and made available via Sketch Engine. To overcome the problem of great differences in the existing sketch grammars for the respective languages, a set of “compatible” sketch grammars have been written that will aid contrastive linguistic research and bilingual lexicographic projects. The sketch grammars use a uniform set of rules for all word categories (parts of speech) and the resulting set of tables is displayed in a fixed order in all languages.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

From GLÀFF to PsychoGLÀFF: a large psycholinguistics-oriented French lexical resource

Basilio Calderone, Nabil Hathout, Franck Sajous
CLLE-ERSS (CNRS&Université de Toulouse-Le Mirail)
{basilio.calderone, nabil.hathout, franck.sajous}@univ-tlse2.fr

Abstract

In this paper, we present two French lexical resources, GLÀFF and PsychoGLÀFF. The former, automatically extracted from the collaborative online dictionary Wiktionary, is a large-scale versatile lexicon exploitable in Natural Language Processing applications and linguistic studies. The latter, based on GLÀFF, is a lexicon specifically designed for psycholinguistic research. GLÀFF, counting more than 1.4 million entries, features an unprecedented size. It reports lemmas, main syntactic categories, inflectional features and phonemic transcriptions. PsychoGLÀFF contains additional information related to formal aspects of the lexicon and its distribution. It contains about 340,000 entries (120,000 lemmas) that are corpora-attested. We explain how the resources have been created and compare them to other known resources in terms of coverage and quality. Regarding PsychoGLÀFF, the comparison shows that it has an exceptionally large repertoire while having a comparable quality.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

The basic vocabulary of Italian from 1980 to the present time: methods and results

Tullio De Mauro, Isabella Chiari
Sapienza Università di Roma
tullio.demauro@uniroma1.it, isabella.chiari@uniroma1.it

Abstract

An overview of the main features and innovations introduced in the building of the New Basic Vocabulary of Italian will be given, stressing the methodological novelties on the background of existing resources describing core vocabularies. We will present some of the central choices, features and problems encountered in corpus building and processing, and illustrate some of the most significant linguistic aspects that the new version of the basic vocabulary presents. We will provide a rapid summary of significant changes occurring the fundamental vocabulary of Italian after thirty years of its first versioni is given, accounting for new entries, shifts in lexicon layers, and entries that have disappeared from the core of Italian lexicon and connecting these changes to general transformations that have occurred in Italian society and in all aspects of language in the last fifty years.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

RIDIRE. Corpus and Tools for the Acquisition of Italian L2

Alessandro Panunzi, Emanuela Cresti, Lorenzo Gregori
University of Florence
alessandro.panunzi@unifi.it, elicresti@unifi.it, lorenzo.gregori@unifi.it

Abstract

This paper introduces the RIDIRE corpus, built by means of an open source tool (RIDIRE-CPI) for creating specifically designed web corpora through a targeted crawling strategy. The RIDIRE-CPI architecture combines existing open source tools with specifically developed modules, comprising a robust crawler, a user friendly web interface, several conversion and cleaning tools, an anti-duplicate filter, a language guesser, and a PoS-tagger. The RIDIRE corpus is a balanced Italian web corpus (1.5 billion tokens) designed for enhancing the study of Italian as a second language, while also being exploitable for lexicographic purposes. The targeted crawling was performed through content selection, metadata assignment, and validation procedures. These features allowed the construction of a large corpus with a specific design, covering a variety of language usage domains (News, Business, Administration and Legislation, Literature, Fiction, Design, Cookery, Sport, Tourism, Religion, Fine Arts, Cinema, Music). The RIDIRE query system allows research to be carried out on the whole corpus itself or on the sub-corpora. Specifically, available queries comprehend all the functions usually exploited in corpus-based lexicography: frequency lists, concordances and patterns, collocations, Sketches, and Sketch Differences.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

Empirical approaches to German paronyms

Petra Storjohann, Ulrich Schnörch
Institut für Deutsche Sprache Mannheim
storjohann@ids-mannheim.de, schnoerch@ids-mannheim.de

Abstract

German lexical items with similar or related morphological roots and similar meaning potential are easily confused by native speakers and language learners. These include so-called paronyms such as *effektiv/effizient*, *sensitive/sensibel*, *formell/formal/förmlich*. Although these are generally not regarded as synonyms, empirical studies suggest that in some cases items of a paronym set have undergone meaning change and developed synonymous notions. In other cases, they remain similar in meaning, but show subtle differences in definition and restrictions of usage. Whereas the treatment of synonyms has received attention from corpus-linguists (cf. Partington 1998; Taylor 2003), the subject of paronyms has not been revisited with empirical, data-driven methods neither in terms of semantic theory nor in terms of practical lexicography. As a consequence, we also need to search for suitable corpus methods for detailed semantic investigation. Lexicographically, some German paronyms have been documented in printed dictionaries (e.g. Müller 1973; Pollmann & Wolk 2010). However, there is no corpus-assisted reference guide describing paronyms empirically and enabling readers to find the correct contemporary usage. Therefore, solutions to some lexicographic challenges are required.

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

Pragmatic Meaning in Lexicographical Description: Semantic Prosody on the Go

Mojca Šorli
Trojina - Institute for Applied Slovene Studies
mojca.sorli@trojina.si

Abstract

The present paper focuses on ways in which the pragmatic (functional) meaning, known in corpus linguistics as semantic prosody, is treated in monolingual and also bilingual dictionaries. We have analysed a selection of lexicographical descriptions, as they are represented in the *Slovene Lexical Database* (SLD), comparing them to a number of English and Slovene lexical sources, and demonstrated how corpus-derived pragmatic meaning can become an integral part of dictionary definitions. This is particularly important for the treatment of phraseology and idiomatics, where greater involvement of pragmatics is demonstrated. A tentative typology was compiled for the purpose of this analysis in order to categorise lexical units according to their inner semantic-pragmatic relations, with a view to examining the implications for the monolingual dictionary treatment of individual items, as well as any potential strategies that could be applied, on the basis of the posited categories, to their translation. We have also pointed out the treatment of individual lexical units in the selected bilingual dictionaries.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

Bi-and Multilingual Lexicography

Linking a dictionary to other open data – Better access to more specific information for the users

Ulrich Apel
Eberhard Karls University Tübingen
ulrich.apel@uni-tuebingen.de

Abstract

The project WaDokuJT [2013] has developed into the most comprehensive Japanese-German dictionary in regard to covered lemmata and translation equivalents. The dictionary is used by many users from German speaking countries and from Japan. The project was designed concentrating on German speaking users who want to read and translate Japanese texts. So, priority is given to a rather complete coverage of orthographic variations of Japanese headwords on the one side and German translation equivalents on the other side. Definitions are given only when they are really necessary and only in German language; further, they are as short as possible.

This means that the dictionary project may have certain shortcomings for text production, for Japanese users or for users who need deeper encyclopedic explanations connected to a certain dictionary headword.

This paper presents an approach how to alleviate such deficiencies by providing hyperlinks or other references to data of other dictionaries and encyclopedia projects which may contain the information, users are looking for. For this aim, WaDokuJT data refers especially to open source projects, dictionaries, which aren't protected by copyright anymore, and collaborating projects, the data of which can be accessed easily at least in parts.

Language of Presentation: English

Fri, 18/7/2014 – Room 1

Bilingual word sketches: the *translate* button

Vít Baisa, Miloš Jakubíček, Adam Kilgarriff, Vojtěch Kovář, Pavel Rychlý
Lexical Computing Ltd, UK; Faculty of Informatics, Masaryk University, Czech Republic
{xbaisa,jak,xkovar3,pary}@fi.muni.cz, adam@lexmasterclass.com

Abstract

We present bilingual word sketches: automatic, corpus based summaries of the grammatical and collocational behaviour of a word in one language and its translation equivalent in another. We explore, with examples, various ways that this can be done, using parallel corpora, comparable corpora and bilingual dictionaries. We present the formalism for specifying equivalences between gramrels in the two languages. We show how bilingual word sketches can be useful for dictionary-making and we present additional functionality to make them more useful. We state the language pairs for which bilingual word sketches are currently available, and our plans for adding more pairs.

Language of Presentation: English

Fri, 18/7/2014 – Room 1

Creating a Bilingual Italian-English Dictionary of Collocations

Barbara Berti, Laura Pinnavaia
Università degli Studi dell'Insubria, Università Statale di Milano
barbara.berti@uninsubria.it, laura.pinnavaia@unimi.it

Abstract

Collocations are among the most challenging issues that learners of a foreign language have to face. The main obstacle for Italian learners of English is encountered in active tasks, and especially, when translating or writing essays in English. The encoding of collocations in fact often results in the selection of wrong terms. There is no doubt that the production and translation of Italian collocations would be facilitated were students to have at their disposal a bilingual Italian-English Dictionary of Collocations. The aim of this work is to endorse the importance of the existence of such a tool and to show the beginnings of its compilation.

Language of Presentation: English

Fri, 18/7/2014 – Room 1

La valencia del adjetivo en diccionarios bilingües alemán-español-alemán

Andreu Castell, Natàlia Català, María Bargalló
Universitat Rovira i Virgili – Tarragona (España)
andreu.castell.@urv.cat, natalia.catala@urv.cat, maria@bargallo@urv.cat

Abstract

El verbo constituye el elemento fundamental en la teoría de valencias, por lo que es normal que se le haya prestado siempre una especial atención. Menos atención se le ha dedicado hasta ahora a la valencia del nombre y mucho menos aún a la del adjetivo. El estudio que se presenta forma parte de la fase inicial de un proyecto que se está desarrollando con el fin de establecer las bases teóricas y metodológicas para la elaboración de un diccionario sintáctico del adjetivo en alemán y en español. Se analiza la información sintáctica que ofrecen diversos diccionarios bilingües alemán-español-alemán sobre la complementación de ocho adjetivos que, en principio, poseen los mismos complementos argumentales en ambas lenguas. Se estudia, en concreto, qué información ofrecen respecto de una serie de criterios que desde el punto de vista de la gramática de valencias se deberían considerar fundamentales, así como hasta qué punto la información ofrecida, independientemente de los presupuestos de la teoría de valencias, resulta realmente útil al usuario. También se analiza el grado de coherencia interna que presentan los diccionarios estudiados en las diversas entradas.

Language of Presentation: Spanish

Fri, 18/7/2014 – Room 1

Esame storico dei “realia” nei dizionari bilingui italiano/ungheresi

Zsuzsanna Fábíán
Università “Eötvös Loránd” di Budapest, Istituto di Romanistica
fabian.zsuzsanna@btk.elte.hu

Abstract

Lo scopo del presente lavoro (di carattere lessico-contrastivo ma anche storico-lessicografico) è quello di analizzare l'evoluzione del trattamento dei “realia” (parole culturo-specifiche, *lexical/referential gaps*) in otto dizionari bilingui italiano-ungheresi e in sei ungherese-italiani. Per l'esame sono stati scelti 16 esponenti italiani e altrettanti ungheresi appartenenti al lessico della gastronomia (i lemmi sono presentati nelle due Appendici). Le singole equivalenze registrate nei dizionari sono state assegnate alle varie categorie delle possibili strategie traduttive quali l'uso di una parola di origine straniera (forestierismi, prestiti o calchi), quello di un equivalente (sintagmatico) esplicativo (distinto ev. anche graficamente, p.es. col corsivo o con la collocazione tra parentesi, da un “vero” equivalente), o di un equivalente generalizzante (iperonimo), l'adattamento o la sostituzione con un equivalente approssimativo/parziale di L2, ecc. Nella parte conclusiva dell'analisi sono state indagate le problematiche seguenti: Quali sono le strategie traduttive più frequenti nei dizionari bilingui tra italiano e ungherese? E' possibile individuare, nei singoli autori dei dizionari analizzati, una preferenza verso certe strategie traduttive? Nel caso di eventuali errori si possono registrare “copiature” e/o “correzioni”? Sussiste una connessione tra le strategie dominanti nei dizionari bilingui e la direzione delle lingue/tipo del vocabolario (attivo o passivo)?

Language of Presentation: Italian

Fri, 18/7/2014 – Room 1

Quello che i dizionari possono fare: l'esempio dei Dizionari di Tedesco (Giacoma/Kolb – Zanichelli/Klett)

Luisa Giacoma
Università di Torino
luisa.giacoma@unito.it

Abstract

Partendo dall'affermazione di Hausmann (1993) che dichiara la non apprendibilità del lessico e che vede nei dizionari l'unica possibilità di salvezza per l'apprendente, si avverte l'esigenza di indagare su cosa possa fare la lessicografia per far fronte a questa gravosa responsabilità. Da una parte bisogna educare il lettore ad un uso proficuo del dizionario, mentre dall'altra è necessario innovare i dizionari, affinché possano aiutare l'apprendente anche nell'uso attivo della lingua. Per potersi esprimere adeguatamente in L2 egli deve infatti saper padroneggiare i legami semantici, sintattici e pragmatici esistenti tra le parole ed essere consapevole dei notevoli vincoli che questi pongono alla libertà combinatoria del parlante (Lo Cascio 1997). Se teoricamente il parlante può abbinare liberamente le parole, in realtà solo poche combinazioni sono possibili, poiché esse sottostanno a limitazioni sintattiche, semantiche, enciclopediche e combinatorie dettate dall'uso e dal contesto. Un'analisi dei principali problemi rimasti irrisolti nella lessicografia tradizionale, orientata per lo più verso l'uso passivo della lingua, ha portato allo sviluppo di un nuovo modello lessicografico alla base dei dizionari di Tedesco a cura di Luisa Giacoma e Susanne Kolb e pubblicati dalla casa editrice italiana Zanichelli e tedesca Klett, che costituiscono una risposta concreta alle nuove esigenze degli apprendenti.

Language of Presentation: Italian

Fri, 18/7/2014 – Room 1

Bilingual Dictionary Drafting. The example of German-Basque, a medium-density language pair

David Lindemann, Iker Manterola, Rogelio Nazar, Iñaki San Vicente, Xabier Saralegi
UPV-EHU University of the Basque Country, Elhuyar Foundation, Pontificia Universidad Católica de Valparaíso,
Elhuyar Foundation, Elhuyar Foundation
david.lindemann@ehu.es, i.manterola@elhuyar.com, rogelio.nazar@ucv.cl, i.sanvicente@elhuyar.com,
x.saralegi@elhuyar.com

Abstract

This paper presents a set of Bilingual Dictionary Drafting (BDD) methods including manual extraction from existing lexical databases and corpus based NLP tools, as well as their evaluation on the example of German-Basque as language pair. Our aim is twofold: to give support to a German-Basque bilingual dictionary project by providing draft Bilingual Glossaries and to provide lexicographers with insight into how useful BDD methods are. Results show that the analysed methods can greatly assist on bilingual dictionary writing, in the context of medium-density language pairs.

Language of Presentation: English

Fri, 18/7/2014 – Room 1

Illustrative Examples and the Aspect of Culture: The Perspective of a Tshivenda Bilingual Dictionary

Munzhedzi James Mafela
University of South Africa
mafelmj@unisa.ac.za

Abstract

Few Tshivenda dictionaries make use of illustrative examples to define lexical entries. Among those which use illustrative examples to define lexical entries is *Venda Dictionary: Tshivenda – English*, which makes use of simple sentences, idioms, riddles and proverbs. Through the use of illustrative examples lexicographers show the headword in use in typical contexts (Katzaros 2004). In many instances illustrative examples in *Venda Dictionary: Tshivenda -- English* reflect on the culture of the Vhava; information which is valuable to Tshivenda learners. Learners of a foreign language find it difficult to understand meanings of cultural words. Bartholomew (1995:4) writes that a cursory reading of the illustrative sentences not only helps the reader to understand the entry word, but it also gives him a feel for the kind of life led by the speakers. This paper seeks to highlight the importance of illustrative examples in familiarising dictionary users with the culture of the Vhava. Data from *Venda Dictionary: Tshivenda – English* will be used to facilitate the discussion of illustrative examples and the aspect of Tshivenda culture.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Corpus, Parallélisme et Lexicographie Bilingue

Adriana Zavaglia, Gisele Galafacci
Université de São Paulo
zavaglia@usp.br, gisele.galafacci@usp.br

Abstract

Dans le cadre de la lexicographie bilingue brésilienne, surtout sur le statut de la relation entre l'entrée et ses traductions, il n'est pas rare que les dictionnaires proposent à l'utilisateur une liste de traductions possibles pour une entrée sans contextualisation ni rubrique d'usages. L'utilisateur de ses ouvrages se pose fréquemment des questions sur l'usage le plus approprié, sur la frontière entre les différents sens des mots proposés et leur champ notionnel. Les traductions répertoriées et décontextualisées n'aident donc pas à résoudre les ambiguïtés concernées. Pour essayer de collaborer là-dessus, nous proposons une approche différentielle (portugais - français) de la lexicographie bilingue à l'aide de la linguistique de corpus. Composé d'originaux et de traductions, le corpus parallèle utilisé dans ce travail est à la base de la contextualisation authentique de la polysémie des entrées, des définitions et des exemples bilingues. Compte tenu de ces aspects, ce travail fournit des commentaires sur les méthodologies des dictionnaires bilingues consultés en considérant le mot grammatical « bem » du portugais brésilien et présente la méthodologie que nous appelons « parallèle » et les conséquences pour l'utilisateur concernant l'entrée en question, en particulier dans les contextes d'apprentissage du français langue étrangère et de la traduction.

Language of Presentation: French

Fri, 18/7/2014 – Room 1

**Lexicography for Specialised Languages, Technology and
Terminography**

Contrastive Collocational Networks for Compiling Specialized Dictionaries for Learners of Spanish for Specific Purposes

Araceli Alonso Campo
LiCoRN Research Group - Université de Bretagne-Sud
InfoLex Research Group – Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra
araceli.alonso@univ-ubs.fr

Abstract

Most specialised dictionaries have been compiled according to a classical perspective of Terminology and, therefore, have not taken into account the use of the specialised lexical units in a specific context. Such dictionaries are not oriented for codifying, and in many cases, they do not offer much detailed information about the syntagmatic and paradigmatic relations of lexical units. The development of a communicative approach to language learning has brought about the need of new resources which allow the language learner to acquire communicative skills to produce texts in real-life situations. The development of Contrastive Linguistic has shown those aspects of the lexical unit which are not so obvious for a non-native speaker, but which must be taken into account to be represented in a learner's dictionary. The present study focuses on a contrastive analysis of lexical combinatorics from a French-Spanish comparable corpora by applying the methodology of collocational networks (Williams 1998). The final aim is to compile a specialized dictionary for learners of Spanish for specific purposes. The method of collocational networks allows a visual representation of specific lexical combinatorics in a specialised domain and facilitates the learning process to produce correct and natural specialised texts.

Language of Presentation: English

Tue, 15/7/2014 – Room 1

EcoLexicon

Pamela Faber, Miriam Buendía Castro
University of Granada
pfaber@ugr.es, mbuendia@ugr.es

Abstract

EcoLexicon (<http://ecolexicon.ugr.es>) is an environmental knowledge base which is based on the premises of Frame-based Terminology (FBT) (Faber 2009, 2011, 2012). EcoLexicon represents the conceptual structure of the specialized domain of the Environment in the form of a visual thesaurus in which environmental concepts are configured in semantic networks. The various terminological designations for a concept are offered in six languages: Spanish, English, German, French, Russian, and Greek. So far, EcoLexicon contains 3,527 concepts and 18,617 terms. It provides conceptual, linguistic, and administrative information for each entry, and we are now beginning to include phraseological information for terms as well (Buendía 2013; Buendía & Sánchez 2012; Sánchez & Buendía 2012). EcoLexicon is designed to meet the needs of different user types, such as a student of science wishing to acquire specialized knowledge about a certain concept, a translator seeking translation correspondences in a language, or a specialist interested in text production.

Language of Presentation: English

Wed, 16/7/2014 – Room 8

Experts and terminologists: exchanging roles in the elaboration of the Terminological Dictionary of the Brenner Base Tunnel (BBT)

Elena Chiochetti, Natascia Ralli
European Academy of Bolzano/Bozen (EURAC research)
echiochetti@eurac.edu, nralli@eurac.edu

Abstract

This paper presents the Italian-German Terminological Dictionary of the Brenner Base Tunnel (BBT) produced and printed in 2011 by the European Academy of Bolzano. In particular, we describe the role played by domain experts and terminologists during the elaboration of the dictionary, which was different from the one that is usually assigned to them in terminological projects. This switching of roles had several consequences on the structure and content of the dictionary, which we will discuss. We also briefly illustrate the challenges faced and how we approached specific problems, e.g. the structure of the definitions, the managing of synonyms and variants, the varied nature of terms selected for the dictionary and their specific treatment with a view to the needs of the target users. The two main target groups consist of experts and semi-experts from various professions who are confronted daily with terminology pertaining to a large array of specialised domains (from environmental to engineering to legal terminology).

Language of Presentation: English

Wed, 16/7/2014 – Room 1

Cloud Terminology Services Facilitate Specialised Lexicography Work

Tatiana Gornostay, Andrejs Vasiljevs
Tilde
tatiana.gornostay@tilde.lv, andrejs@tilde.lv

Abstract

In this software demonstration paper we present an innovative cloud-based platform TaaS “Terminology as a Service” developed in an EU-funded project.¹ The TaaS platform provides language workers and language applications (human and machine users, accordingly) with the services to foster the creation, validation, harmonisation, sharing, and application of terminology resources. Under language workers we understand language professionals, for example, technical writers, editors and proof-readers, translators and localisers, terminologists and domain specialists, lexicographers and terminographers, and others. Under language applications (or machine users in other words), in the first place we consider computer-assisted translation (CAT) tools and machine translation (MT) systems (also, knowledge organisation systems in library and information science, search engines, and others). TaaS provides the following terminology services: terminology search in various sources, terminology identification in and extraction from user-uploaded documents, terminology visualisation in user-uploaded documents, translation equivalent lookup in and retrieval from various sources, terminology refinement and approval by users, terminology sharing with other users, collaborative working environment, and terminology reuse in other applications. Among other benefits for language workers, TaaS serves the needs of specialised lexicography, or terminography, facilitating user-friendly, collaborative, multilingual, interoperable, portable, and cloud-based specialised terminology work. TaaS fills the gap of innovative environment to speed up the development of specialised dictionaries.

Language of Presentation: English

Tue 15, Wed 16, Thu 17/7/2014 – Room 8

Good Contexts for Translators—A First Account of the Cristal Project

Amélie Josselin-Leray*, Cécile Fabre*, Josette Rebeyrolle*, Aurélie Picton**, Emmanuel Planas***

*CLLE-ERSS, University of Toulouse & CNRS, France;
FTI, University of Geneva, Switzerland; *LINA, University of Nantes, France
josselin@univ-tlse2.fr, cecile.fabre@univ-tlse2.fr, rebeyrol@univ-tlse2.fr,
aurelie.picton@unige.ch, emmanuel.planas@univ-nantes.fr

Abstract

This paper questions the notion of good contexts for translators and describes an experiment which tests the usefulness of two specific kinds of contexts in a translation task, namely (1) contexts that provide conceptual information about a term, and (2) contexts that provide linguistic information about the collocational profile of this term. In the experiment, trainee translators are asked to use several types of resources, including a set of pre-annotated contexts, and to identify the contexts that they consider to be the most relevant for their task. We present the first results of this experiment, which confirm our general assumption about the usefulness of such rich contexts and indicate some differences regarding the use of contexts in the source and target language. This study takes place in the CRISTAL project whose aim is to retrieve from bilingual comparable corpora the contexts that are the most relevant for translation and to provide them to users through a CAT tool.

Language of Presentation: English

Wed, 16/7/2014 – Room 1

¹ The TaaS Beta was officially launched on November 1, 2013 and is publicly available for open Beta testing at <https://demo.taas-project.eu>.

Kontextbasierte lexikalische Kontrolle von Anforderungsdokumenten

Jennifer Krisch
Daimler AG
jennifer.krisch@daimler.com

Abstract

Das Verwalten und Prüfen von Anforderungsdokumenten ist ein sehr wichtiger Bereich in den frühen Phasen von Industrieprojekten. Je früher unpräzise und unvollständige Anforderungen identifiziert werden, desto geringer sind die Folgekosten und die Überarbeitungszeit. Unpräzise und unvollständige Anforderungen resultieren häufig aus der Verwendung von Weak-Words, d.h. von unscharfen Wörtern oder Konstruktionen, die in bestimmten Kontexten Mehrdeutigkeiten auslösen. Um Fehlinterpretationen aufgrund von Weak-Words entgegenzuwirken, müssen diese Wörter oder Konstruktionen identifiziert werden. Eine reine Wortsuche reicht hierbei nicht aus, weil Weak-Words nur in bestimmten Satzkontexten Mehrdeutigkeiten auslösen. Der Beitrag beschreibt eine als Prototyp entwickelte automatisierte Weak-Word-Analysemethodik, durch welche ausgewählte Weak-Words identifiziert werden und bei der aufgrund des Satzkontexts entschieden wird, ob eine Warnung bezüglich eines Mangels an Präzision und Eindeutigkeit des Texts an den Autor zurückgegeben werden soll. Bei der Weak-Word-Analyse kommt ein Lexikon zum Einsatz, in welchem die Weak-Words selbst und deren Kontexte abgelegt sind. Die Evaluation des entwickelten Werkzeugs hat ergeben, dass die Anzahl der Warnungen, die an den Benutzer zurückgegeben werden, durch eine Kontextanalyse im Verhältnis zur wortbasierten Suche erheblich reduziert werden kann, ohne dass problematische Kontexte übersehen werden. Dies stellt eine Verbesserung gegenüber den bisher bekannten Systemen dar.

Language of Presentation: German

Wed, 16/7/2014 – Room 1

From term dynamics to concept dynamics: term variation and multidimensionality in the psychiatric domain

Pilar León-Araúz, Arianne Reimerink
Department of Translation and Interpreting, University of Granada
pleon@ugr.es, arianne@ugr.es

Abstract

Medical terminology is one of the most dynamic terminological domains, and the choice of one term instead of the other is not random, but the result of different perspectives towards reality. VariMed is a research project on medical term variants and its overall objective is to generate a multifunctional resource on the medical domain for linguistic research, translation and technical writing. In this paper, we propose a systematic way of extracting term variants from large corpora within the subdomain of Psychiatry and how to represent them according to cognitive and communicative parameters. Our aim is to discover if different conceptualizations, or different conceptually motivated term variants, of the same concept are preferred in expert or semi-specialized communication. A corpus on Psychiatry was compiled and classified according to user types. A grammar was designed in NooJ (Silberstein, 2003) in order to extract term variants based on the usual lexico-syntactic patterns accompanying synonyms (*also known as; commonly referred to as*, etc.). Corpus analysis results indicate that, from a cognitive perspective, term variants reflect the prototypical dimensions in which psychiatric disorders may be classified. From a communicative perspective, terms and dimensions can also be associated with user-based parameters.

Language of Presentation: English

Wed, 16/7/2014 – Room 1

***Bon usage* vs. Fachliches: Fachsprache in der Geschichte der französischen Sprachpflege und Lexikographie**

Martina Mayer
Universität Innsbruck
contact@martina-mayer.com

Abstract

Frankreich ist seit dem 16. Jahrhundert dafür bekannt, sich besonders intensiv der Pflege seiner Nationalsprache zu widmen. Seit dem 17. Jahrhundert ist die französische Sprachpflege und Sprachnormierung institutionalisiert und wird durch entsprechende lexikographische Aktivitäten gestützt. Allerdings zielten all diese Bemühungen bis zur Mitte des 20. Jahrhunderts in erster Linie auf die Gemeinsprache ab – ein Umstand, der vergessen lässt, dass auch die Pflege der Fachsprachen in Frankreich eine nähere Betrachtung verdient. Dies bezieht sich nicht nur auf die modernen sprachpflegerischen Bemühungen an den Fachsprachen, sondern hat durchaus ab dem 16. bzw. 17. Jahrhundert Gültigkeit. Der vorliegende Beitrag wird sich dem Aspekt der Pflege der Fachsprachen und ihrer Berücksichtigung in der französischen Lexikographie daher aus einer historischen Perspektive widmen. Der Schwerpunkt liegt dabei vor allem auf entsprechenden Aktivitäten im *Grand Siècle*, das mit seinen Sprachpflegern, dem *bon usage des honnête homme*, der Gründung der *Académie française* und seinen lexikographischen Referenzwerken bis in die Gegenwart maßgeblich auf die französische Sprache einwirkte.

Language of Presentation: German

Wed, 16/7/2014 – Room 1

EU-Terminologie in den einsprachigen Wörterbüchern des Deutschen

Diana Stantcheva
American University in Bulgaria
dstantcheva@aubg.bg

Abstract

Die heutige Europäische Union blickt nun schon auf eine über 50 jährige Geschichte zurück. In dieser Zeit ist die europäische Staatengemeinschaft von ursprünglich 6 auf derzeit 28 Länder gewachsen. Deutschland gehörte zu den Gründungsmitgliedern der Europäischen Gemeinschaft für Kohle und Stahl, der Europäischen Wirtschaftsgemeinschaft und der Europäischen Atomgemeinschaft, der Organisationen, die später in die Europäische Union zusammenflossen. Österreich trat 1995 der Europäischen Union bei. Diese Fakten zum Anlass nehmend, untersucht der vorliegende Beitrag die lexikographische Kodifizierung EU-bezogener Terminologie in ausgewählten allgemeinen einsprachigen Wörterbüchern des Deutschen von 1961 bis 2009.

Language of Presentation: German

Wed, 16/7/2014 – Room 1

Text boxes as lexicographic device in LSP dictionaries

Elsabé Taljard*, Danie J. Prinsloo*, Rufus H. Gouws**
*University of Pretoria, **University of Stellenbosch
elsabe.taljard@up.ac.za, danie.prinsloo@up.ac.za, rhg@sun.ac.za

Abstract

It is the duty of the lexicographer to maximally utilise all lexicographic devices at his/her disposal in the compilation of a dictionary. In modern-day dictionaries lexicographers often rely on lexicographic text boxes as a means to present lexicographic data, especially where stronger emphasis or additional data regarding a specific entry is needed. It could be argued that the use of text boxes as a lexicographic device is underutilised in general dictionaries and even more so in LSP dictionaries. They should however not be used in a haphazard way but rather according to a well-devised system. Looking at existing dictionaries one often realises that a random use of text boxes has a detrimental influence on successful data retrieval. Text boxes should be used for data that need to be presented in a position of salience. The aim of this paper is first to give an overview of the current status of the use of text boxes in LSP dictionaries; secondly, to offer some suggestions as to the effective use of text boxes in these dictionaries by drawing on an analysis of the current use of text boxes in LSP dictionaries.

Language of presentation: English

Tue, 15/7/2014 – Room 1

Station Sensunique: Architecture générale d'une plateforme web paramétrable, modulaire et évolutive d'acquisition assistée de ressources

Izabella Thomas¹, Blandine Plaisantin Alecu², Bérenger Germain³, Marie-Laure Betbeder⁴

¹Centre L. Tesnière, Université de Franche-Comté

²Prolipsia, France

³Share and Move Solutions, France

⁴Institut Femto-ST. Université de Franche-Comté

izabella.thomas@univ-fcomte.fr, blandine.alecu@prolipsia.com, berenger.germain@shareandmove.fr,

marie-laure.betbeder@univ-fcomte.fr

Résumé

Dans cet article nous décrivons l'architecture générale d'une plateforme web paramétrable, modulaire, collaborative et évolutive d'acquisition assistée de ressources terminologiques et non-terminologiques : la Station Sensunique. Conçue dans l'objectif initial de faciliter et d'accélérer le processus de constitution du lexique d'une Langue Contrôlée (LC), son champ d'application peut être élargi à l'acquisition de tout type de ressources terminologiques. La Station s'articule autour de deux points de vue du processus d'acquisition de ressources : (1) chronologique (centré processus) : import des textes d'entrées, analyse automatique, analyse manuelle approfondie, validation, et enfin export ; (2) ergonomique (centré utilisateur-analyste) : mise en adéquation de l'analyse selon le corpus et l'application visée, visualisation et gestion des unités lexicales candidates (ULC), recherches complexes en corpus ou dans la liste d'ULC, modification ou enrichissement des descriptions ou relations des ULC, validation progressive, demande de validation par l'expert-métier, etc. La Station dispose de deux interfaces utilisateurs faciles à manipuler ; son utilisation se fait sans aucune contrainte technique ni installation préalable, à partir d'une interface web qui intègre l'ensemble des outils et ressources utilisées.

Langue de présentation: français

Tue, 15/7/2014 – Room 1

Station Sensunique: une plateforme Web modulaire, collaborative et évolutive d'acquisition assistée de ressources terminologiques et non terminologiques (orientée Langues Contrôlées)

Izabella Thomas¹, Blandine Plaisantin Alecu², Bérenger Germain³, Marie-Laure Betbeder⁴

¹Centre L. Tesnière, Université de Franche-Comté

²Prolipsia, France

³Share and Move Solutions, France

⁴Institut Femto-ST. Université de Franche-Comté

izabella.thomas@univ-fcomte.fr, blandine.alecu@prolipsia.com, berenger.germain@shareandmove.fr,

marie-laure.betbeder@univ-fcomte.fr

Résumé

Dans cet article, nous présentons le fonctionnement et les services proposés par la Station Sensunique, une plateforme Web modulaire, collaborative et évolutive d'acquisition assistée de ressources terminologiques et du lexique d'une Langue Contrôlée. La Station prend en charge et facilite l'ensemble de ce processus, à travers une analyse automatique du corpus, puis la possibilité d'approfondir l'analyse manuellement, pour aboutir à la création des fiches terminologiques/lexicales et des ressources exportables. La Station est un outil prêt à l'emploi, ergonomique, facile à prendre en main et à exploiter à travers ses différentes interfaces. Elle allie les avantages d'une analyse automatique (rapidité, coût) avec l'exactitude et la fiabilité d'une analyse humaine.

Langue de présentation : français

Wed 16, Thu 17/7/2014 – Room 8

Lexicography of Lesser Used Languages

Towards an Integrated E-Dictionary Application – The Case of an English to Zulu Dictionary of Possessives

Sonja Bosch*, Gertrud Faaß¹

*University of South Africa, ¹University of Hildesheim
boschse@unisa.ac.za, faassg@uni-hildesheim.de

Abstract

This paper describes a first version of an integrated e-dictionary translating possessive constructions from English to Zulu. Zulu possessive constructions are difficult to learn for non-mother tongue speakers. When translating from English into Zulu, a speaker needs to be acquainted with the nominal classification of nouns indicating possession and possessor. Furthermore, (s)he needs to be informed about the morpho-syntactic rules associated with certain combinations of noun classes. Lastly, knowledge of morpho-phonetic changes is also required, because these influence the orthography of the output word forms. Our approach is a novel one in that we combine e-lexicography and natural language processing by developing a (web) interface supporting learners, as well as other users of the dictionary to produce Zulu possessive constructions. The final dictionary that we intend to develop will contain several thousand nouns which users can combine as they wish. It will also translate single words and frequently used multiword expressions, and allow users to test their own translations. On request, information about the morpho-syntactic and morpho-phonetic rules applied by the system are displayed together with the translation. Our approach follows the function theory: the dictionary supports users in text production, at the same time fulfilling a cognitive function.

Language of Presentation: English

Thu, 17/7/2014 – Room 7

Zur (Vor-)Geschichte der saamischen Lexikografie: ein lateinisch-saamisches Wörterverzeichnis aus dem 17. Jahrhundert

Eino Koponen

Institut für die Landessprachen Finnlands
eino.koponen@kotus.fi

Abstract

Die Saamen (früher Lappen) sind ein nordisches Minoritätsvolk, das auf dem Gebiet von vier Staaten (Norwegen, Schweden, Finnland, Russland) lebt. Die Wurzeln der auf Saamisch gedruckten Literatur reichen ins 17. Jahrhundert zurück, und aus demselben Jahrhundert datiert der erste Versuch eines saamischen Lexikons. Dabei handelt es sich um ein von dem schwedischen Pastor Z. Plantinus verfasstes handschriftliches Wörterverzeichnis von ca. 850 lateinischen Wörtern mit saamischen Äquivalenten, das 1888 in Stockholm wiederentdeckt und später publiziert wurde. Das uns überlieferte Dokument scheint eine korrumpierte Abschrift des Originals zu sein, deren zahlreiche Fehler das Erkennen der saamischen (und bisweilen auch lateinischen) Wörter stellenweise (fast) unmöglich machen. Hier wird das Wörterverzeichnis einer genaueren Analyse unterzogen. Die Arbeit wurde im Rahmen eines größeren Forschungsprojekts, das eine etymologische Datenbank der saamischen Sprachen und der verwandten uralischen Sprachen darstellt, am Institut für die Landessprachen Finnlands durchgeführt. Alle saamischen Wörter des Verzeichnisses sind jetzt in die Datenbank eingespeist. Auch die auf den ersten Blick obskuren Wörter werden dadurch der künftigen Forschung zugänglicher. Bei der Präsentation werden die Resultate der Analyse anhand der Älgu-Datenbank veranschaulicht.

Language of Presentation: German

Thu, 17/7/2014 – Room 7

Compiling a Basic Vocabulary for German Sign Language (DGS) – lexicographic issues with a focus on word senses

Gabriele Langer, Susanne König, Silke Matthes
Institute of German Sign Language and Communications of the Deaf,
University of Hamburg
Gabriele.Langer@sign-lang.uni-hamburg.de, Susanne.Koenig@sign-lang.uni-hamburg.de,
Silke.Matthes@sign-lang.uni-hamburg.de

Abstract

Nowadays lexicographic work such as lemma selection, identification of word senses and usage information are usually based on large corpora. Sign language lexicographers face the same difficulties as their colleagues of other less studied and previously not written minority languages. They cannot rely on written texts and large corpora. In addition sign language linguists have to cope with sign-language specific issues due to the visual-gestural modality, namely the iconicity of signs, a broad utilization of simultaneity of linguistic signals and the integration of lexical material of a spoken language by the way of mouthings.

Sign language specific lexicographic issues are discussed and exemplified with regard to the Basic Vocabulary of German Sign Language (DGS) that is compiled within the larger context of the DGS Corpus Project. The Basic Vocabulary is not corpus-based but based on previously published sign collections that are used as a starting point. The data is reviewed, sign senses are finer split and disambiguated and the data then undergo a validation process by the sign language community. This validation process is conducted through a feedback system especially designed for surveys involving sign language.

Language of Presentation English

Thu, 17/7/2014 – Room 7

Dearcadh na nDéise – Representations of Gaeltacht na nDéise in Dineen's Bilingual Irish-English Dictionary (1927)

Chris Mulhall, Seamus ó'Diollúin
Waterford Institute of Technology
cmulhall@wit.ie, sodiolluin@wit.ie

Abstract

Dictionaries are a written manifestation of the world, but the selection of words is a reflection of society at a particular time from one or more viewpoints. Irish dictionaries are largely uncharted territories in terms of lexicographic research and contain important regional and national narratives that reflect the ever-changing Irish cultural landscape in the late nineteenth and early twentieth century. The creation of the independent Irish state in the early part of the twentieth century and increasing demand for bilingual Irish-English dictionaries at this time heralded a beginning of a new area of activity in European lexicography. One of landmark editions from this period is *An Foclóir Gaedhilge agus Béarla* by Patrick S. Dineen. The aim of this paper is to explore a regional variety of Irish, Gaeltacht na nDéise, found in the South-East of Ireland and examine some of the socio-cultural narratives behind Dineen's choice of entries to depict this regional dialect. Based on a recategorisation of the empirical work of Nyhan (2006), a brief discussion of three prevalent socio-cultural themes, namely; anger, anxiety and poverty, recurring in some Gaeltacht na nDéise-specific entries found in Dineen's work is presented. These offer an interesting window into how the lexis of this particular region was represented and potentially contributed to its regional identity in a national context.

Language of Presentation: English

Thu, 17/7/2014 – Room 7

The eLexicon Mediae et Infimae Latinitatis Polonorum. The Electronic Dictionary of Polish Medieval Latin

Krzysztof Nowak
Institute of Polish Language, Polish Academy of Sciences
krzysztof@ijp-pan.krakow.pl

Abstract

The paper presents goals, methods, and results of the project of the Electronic Dictionary of Polish Medieval Latin. First, a brief history of the paper dictionary, as well as an account of its main features are presented. Second, the main problems of the metalexigraphic analysis and the subsequent XML encoding of the lexicographic content are discussed. The main purpose of both being a fine-grained description of the linguistic resource, it was necessary to make explicit a fair amount of data which were coded only by means of convention. Third, the web interface of the dictionary is treated in more detail. The most important of its design principles include separation of the expert and novice user perspective, system of aids and suggestions, integration with external sources.

Language of Presentation: English

Thu, 17/7/2014 – Room 7

From DANTE to Dictionary: The New English-Irish Dictionary

Pádraig Ó Mianáin, Cathal Convery
Foras na Gaeilge, Dublin
pomianain@forasnagaeilge.ie, cconvery@forasnagaeilge.ie

Abstract

Most major bilingual dictionary projects tend to largely involve adapting an existing bilingual dictionary, by adopting or amending the existing source language material as required and then supplying the necessary target language material. This paper describes the innovative approach followed in the production of the New English-Irish Dictionary (NEID) project. The online version of NEID was launched in 2013 (www.focloir.ie) and the project is to be completed when a printed version is produced in 2016. The English language content of NEID is based on the Database of ANalysed Texts of English (the DANTE database), a ground-breaking corpus-based lexical database developed specifically for the project. Attention is drawn to how the DANTE entry frameworks evolved through the various translation and editing processes to the final entries now available in the online. The paper also discusses the development of the Irish-language material and details some of the challenges, practical and editorial, encountered in the course of the production of the dictionary, and the working solutions that were developed.

Language of Presentation: English

Thu, 17/7/2014 – Room 7

User Support in e-Dictionaries for Complex Grammatical Structures in the Bantu Languages

Danie J. Prinsloo*, Theo J.D. Bothma*, Ulrich Heid**
*University of Pretoria, **University of Hildesheim
danie.prinsloo@up.ac.za, theo.bothma@up.ac.za, heid@uni-hildesheim.de

Abstract

This paper discusses direct user guidance as a mechanism in an e-dictionary to provide user support for complex grammatical structures in the Bantu languages. We present a design study to show that user support through direct user guidance can provide solutions in the case of complex concordial relationships between nouns and pronouns. The compilation of the complex *relative construction* is taken as a case in point. The concept of *user support* appropriately puts the user in focus. Our approach to user support also caters for the casual, on-the-fly user, who is not interested or in a position to learn a foreign language.

Language of Presentation: English

Thu, 17/7/2014 – Room 7

Concerning the Treatment of Co-existent Synonyms in Estonian Monolingual and Bilingual Dictionaries

Enn Veldi
University of Tartu, Estonia
Enn.Veldi@ut.ee

Abstract

Estonian is a lesser used language of Europe (with about one million users) that belongs to the Finnic group of Finno-Ugric language. It has a large proportion of international words, which has resulted in extensive co-existent synonymy between native and international words. However, because of linguistic purism Estonian monolingual dictionaries do not treat the members of such synonym pairs on an equal footing. Usually, they give preference to native words, which in practice implies that international words are provided together with their native-language equivalents, but native words are without any reference to international words. Unfortunately, this kind of asymmetrical treatment is not helpful for language users and compilers of bilingual dictionaries; they need to develop synonym competence, which covers both international and native synonyms. Probably the best solution is symmetrical treatment, that is, both synonyms should be provided side by side on a regular basis, which contributes to their better knowledge. It is argued that the quality of bilingual dictionaries could be significantly improved by providing the synonym pairs together on a systematic basis.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Phraseology and Collocation

Unusual phrases in English MLDS: increasing user friendliness

Stephen Coffey
Università di Pisa
coffey@cli.unipi.it

Abstract

This paper investigates the presentation of compositionally anomalous phrases in English monolingual learners' dictionaries (MLDS). In particular, it argues that it would be pedagogically useful to explain to the dictionary user, where possible, the reason why certain types of anomaly exist. Two types of phrase are discussed: firstly, idiomatic expressions in which the relationship between phrasal meaning and original meaning may not be clear to the learner (e.g. *run the gauntlet*); secondly, phrases which include particularly unusual word forms or word senses. These include lexical fossils (as in *the whys and WHEREFORES*) and phrases partially motivated by phonological characteristics (as in *bits and BOBS*). In order to form an impression of how anomalous phrases are currently treated in MLDS, samples of items were looked for in both print and online editions. It was found that, overall, little attention is paid to the motivation of phrasal composition, and it is suggested that more should be done in this direction. This would involve integrating current description, almost entirely synchronic in nature, with historical data, at least in the case of some types of phrasal unit.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

Harvesting from One's Own Field: A Study in Collocational Resonance

Janet DeCesaris, Geoffrey Williams
Universitat Pompeu Fabra, Université de Bretagne-Sud
janet.decesaris@upf.edu, geoffrey.williams@univ-ubs.fr

Abstract

This paper presents an initial study in the collocational resonance of three words: *field*, *champ*, and *campo*. *Field* and *champ/campo* do not share the same etymology, yet *field* displays sense extension that in some cases is parallel to that displayed by *champ* and *campo*. Collocational resonance posits that meaning associated with one context of use may be activated by speakers in another context, even though the original meaning context may fall into disuse in the language. To determine collocational resonance, the study considers both the historical development of senses as represented in dictionaries and the current behaviour of these words as represented in dictionaries and corpora. It is suggested that an approach to word meaning including prototypes and resonance can improve the representation of polysemy in dictionaries.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

The use of corpora in bilingual phraseography

Dmitrij Dobrovolskij
Russian Academy of Sciences, Russian Language Institute
Austrian Academy of Sciences, AAC-Austrian Academy Corpus
dobrovolskij@gmail.com

Abstract

The present paper discusses issues in the compilation of bilingual dictionaries of idioms based on an analysis of corpus data. The advantages of using corpora consist not only in more detailed and well thought-out illustrations of the expressions being described, but also in the additional possibilities that the corpus materials provide for compiling the idiom list and structuring entries. Thus the corpus allows us to determine the degree of frequency of an expression (at least in the written language). The relevant principles are illustrated by data taken from a new German-Russian dictionary of idioms that is being constructed by an international team of linguists and lexicographers. Fragments of this dictionary are available on the website of the German Language Institute in Mannheim: "Deutsch-russische Idiome online" <http://wvonline.ids-mannheim.de/idiome_russ/index.htm>. Relevant information is also made available via the Europhras homepage on the website <<http://www.europhras.org>>. All examples of idiom usage in this dictionary are taken from the text corpora DeReKo and DWDS, and in individual cases from the German-language Internet. Parallel German-Russian texts from the Russian National Corpus (RNC) are also used.

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

Comparing Phraseologisms: Building a Corpus-Based Lexicographic Resource for Translators

Laura Giacomini
University of Heidelberg
laura.giacomini@iued.uni-heidelberg.de

Abstract

Today there is still a significant need for specific lexicographic resources in digital form, as they can remarkably improve access to data and actively assist the process of text production. This paper describes the way in which corpus data can be explored to retrieve suitable material for the representation of a particular class of culture-specific phraseologisms and similes in a bilingual dictionary for translators.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

Lexical Variation within Phraseological Units

Tarja Riitta Heinonen
Institute for the Languages of Finland
tarja.heinonen@kotus.fi

Abstract

This paper discusses lexical variation in phraseological units from theoretical and lexicographical perspectives. The starting point is the observation that the existence of lexical variation is sometimes disputed in principle. It has been argued that a change in a single word is sufficient to change the meaning of the whole, thus creating a new expression. Another argument is that after allowing variation in one word one cannot but allow it in multiple words, which quickly turns the original expression unrecognizable. In contrast to these views, this paper argues that lexical variation is not arbitrary but follows certain principles. When all contributing factors are taken into account, the variation in phraseological units is often item-specific, and yet it conforms to general patterns. Towards the end of the paper, Petrova's (2011) multi-level model will be introduced, offering a promising view for theoretical analysis. However, in dictionary work it is reasonable to adhere to generally accepted conventions and not to complicate the structure of the entries too much. An ideal entry gives a sound corpus-based description with representative examples of usage.

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

Prendere il toro per le corna o lasciare una bocca amara? – The Treatment of Tripartite Italian Idioms in Monolingual Italian and Bilingual Italian-English Dictionaries

Chris Mulhall
Waterford Institute of Technology
cmulhall@wit.ie

Abstract

This paper undertakes an empirical investigation into the treatment of tripartite Italian idioms in selected monolingual Italian and bilingual Italian-English dictionaries. Tripartite idioms are phrasal constructs typically arranged into one of the three following syntactic forms: V+N+N, V+ADJ+N or V+N+ADJ. From an organisational viewpoint, these idioms are somewhat more problematic for lexicographers due to the presence of a third lexical constituent. Current lexicographical practice adopts a largely subjective approach to dealing with idioms, which for the most part are V+N forms, therefore those with a wider syntactic form require more considered decision making when determining their point(s) of entry in a dictionary. Certain tripartite idioms also collapse into binomial (N+N) or nominal-adjectival forms (ADJ+N/N+ADJ), thus placing a question mark over the necessity to record their verb element or not. Together, these issues contribute to making tripartite idioms one of the most acutely difficult phrasal categories for lexicographers. This paper examines the entry points of 100 tripartite Italian idioms in selected monolingual Italian and bilingual Italian-English dictionaries published within the last decade. Firstly, it discusses relevant theoretical viewpoints relating to the lemmatisation of idioms over the last 30 years. Thereafter, it focuses on the notion of an idiom as a lexicographical entry in consideration of its varied intrinsic features and how these are productive in selecting an appropriate entry point. The presentation of the empirical data in the following section is stratified according to the publishing house of the analysed dictionaries and gives a comparative discussion on their respective approaches to dealing with tripartite idioms. Finally, the last section unifies the theoretical arguments and practical approaches and proposes a theoretical framework model for lemmatising tripartite idioms to offer a more coherent and consistent platform for their organisation in monolingual Italian and bilingual Italian-English dictionaries.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

Especialización y Prototipicidad en Binomios N y N

Ignacio Rodríguez Sánchez
Universidad Autónoma de Querétaro
igrodsan@uaq.mx

Abstract

El trabajo que aquí se presenta es una investigación sobre binomios de estructura *N y N* (dos sustantivos unidos por la conjunción *y*). Esta investigación tiene un carácter exploratorio y se inscribe en la corriente neofirthiana de la lingüística de corpus. A nivel teórico y metodológico partimos de un enfoque guiado por datos, basándonos en los siguientes corpus: Corpus del Español, CORDE, Googlebooks, y EsTenTen11 (Sketchengine). Aquí se abordan asuntos relacionados con la naturaleza de este tipo de colocaciones: su frecuencia, la estadística de la información mutua, la dispersión y el grado de reversibilidad para identificarlas; las relaciones que se establecen entre ambas partes; la especialización del nodo (como primera o segunda parte de los binomios) y, finalmente, la prototipicidad de algunos elementos. Se concluye que el concepto de binomio tal como se entendía a partir de su definición clásica, se diluye, y se propone una visión integrada en un sistema dinámico de interacciones complejas e impredecibles.

Language of Presentation: Spanish

Thu, 17/7/2014 – Room 2-3

Syntax and Semantics vs. Statistics for Italian Multiword Expressions: Empirical Prototypes and Extraction Strategies

Luigi Squillante
Sapienza – Università di Roma
luigi.squillante@uniroma1.it

Abstract

In this work we present an empirical analysis performed on Italian nominal multiword expressions (MWEs) of the form [noun + adjective] that aims at studying quantitatively their syntactic and semantic features in order to improve their automatic identification and collection. Three indices are proposed, which are able to measure syntactic and semantic frozenness of the expressions on empirical basis in a corpus of about 1.8 million words, composed of Italian texts concerning the domain of physics. The combination of the three indices can be used to create a global measure, that we call Prototypicality Index (PI), which appears to be useful in the automatic extraction of terminological MWEs. The performance of PI at extracting true positives out of a candidate list is compared to those of

the well-known statistical association measures Log-likelihood and Pointwise Mutual Information. Our results show how the performance of PI can be comparable to those of association measures, although it does not involve statistical calculations. Thus, PI can be seen as a new option for lexicographers and terminologists to integrate the already available statistical methods when identifying MWEs from texts.

Language of Presentation: English

Thu, 17/7/2014 – Room 2-3

Historical Lexicography and Etymology

Il DiVo (Dizionario dei Volgarizzamenti). Un archivio digitale integrato per lo studio del lessico di traduzione nell'italiano

Diego Dotto
Opera del Vocabolario Italiano - CNR
dotto@ovi.cnr.it

Abstract

Il progetto *Divo (Dizionario dei volgarizzamenti)* si propone uno studio analitico del lessico di traduzione dei volgarizzamenti medievali e allo stesso tempo la costruzione di strumenti *ad hoc* per lo studio di questo lessico. Il progetto si compone di tre fasi: la prima è la compilazione di una bibliografia filologica secondo il modello *TLIOn*; la seconda è la costruzione di un corpus lemmatizzato, in cui ciascun testo volgare è associato paragrafo per paragrafo all'originale latino; il terzo e ultimo punto è lo studio del lessico di traduzione, cioè il lessico identificato come una traduzione diretta dal latino.

Language of Presentation: Italian

Sat, 19/7/2014 – Room 7

Informatiser le *Französisches etymologisches Wörterbuch* : la nécessaire prise en compte de l'utilisateur

Pascale Renders, Esther Baiwir
F.R.S-FNRS/Université de Liège
pascale.renders@ulg.ac.be, ebaiwir@ulg.ac.be

Abstract

« Know your user » (Atkins & Rundell 2008 : 5). Ce conseil ne s'applique pas uniquement à la conception et la rédaction d'un nouveau dictionnaire. La transformation d'un dictionnaire imprimé en dictionnaire électronique peut également bénéficier d'une étude réévaluant les parcours de consultation suivis par les utilisateurs et les difficultés qu'ils rencontrent. Dans une étude préalable à l'informatisation du *Französisches Etymologisches Wörterbuch* de Walther von Wartburg, la prise en compte du point de vue des utilisateurs a mis en évidence deux facettes de l'ouvrage, vu tantôt comme un recueil de monographies, tantôt comme un thesaurus. Après un résumé de l'avancement du projet (qui est maintenant dans sa phase de production), cette communication expose la façon dont les deux visions ont influencé la modélisation informatique du discours lexicographique, permettant ainsi de résoudre une grande partie des problèmes rencontrés par les utilisateurs et ouvrant la voie à une mise à jour de l'ouvrage.

Language of Presentation: French

Fri, 18/7/2014 – Room 2-3

A morphological historical root dictionary for Portuguese

João Paulo Silvestre, Alina Villalva
Centro de Linguística da Universidade de Lisboa
jpsilvestre@fl.ul.pt, alinavillalva@campus.ul.pt

Abstract

The project of a Morphological Historical Root Dictionary (MHRD) for Portuguese aims to build a specialized dictionary containing a critical selection of lexical units: the first stage is devoted to adjectives, namely those that can be found in Figueiredo (1913). Its main goals are the clarification of morphological and semantic issues in the evolution of the lexicon, and the assessment of the communicative adequacy of the words that are registered in current Portuguese dictionaries, particularly by signaling unused or seldom used words.

The methodology we established is three-sided: it first relies on the lexical analysis of the selected words; then, it seeks for lexicographic information in old Portuguese dictionaries (16th to 19th centuries), in Portuguese textual corpora and in etymological dictionaries; finally, it contrasts the Portuguese data with data from other romance languages, searching for lexical and semantic loans. To conclude, we propose a prototype dictionary entry, applying the above-described methodology to the adjective *bravo*.

Language of Presentation: English

Sat, 19/7/2014 – Room 7

Lexicological Issues of Lexicographical Relevance

What can lexicography gain from studies of loanword perception and adaptation?

Mirosław Bańko, Milena Hebal-Jeziarska
Institute of Polish Language, Institute of West and South Slavic Studies, University of Warsaw
m.banko@uw.edu.pl, milena.hebal-jeziarska@uw.edu.pl

Abstract

In normative dictionaries and usage guides, many loanwords are dismissed as ‘unnecessary’ on the grounds that they have a native synonym, which should be favoured instead. If there is no native equivalent of a loanword in the recipient language, one is often invented in order to eradicate the unwanted loan. However, more detailed studies, in particular, those referred to in this paper, suggest that there is no such thing as fully equivalent words: even if two words have the same designative meaning, they differ in other respects, which makes them mutually unexchangeable except for contexts where the difference between them is inessential.

The goal of this paper is to demonstrate that the meaning potential of lexical loans is different from that of their native synonyms, just because their form is different and differently perceived by language users. The different perception of loanwords can in turn affect their semantic development, thus causing a loanword and its native synonym to diverge. The authors of normative dictionaries and language guides should, therefore, give more consideration to lexical borrowings before they condemn them as ‘unnecessary’ or ‘snobbish’.

Language of Presentation: English

Sat, 19/7/2014 – Room 2-3

Pejorative Language Use in the Satirical Journal "Die Fackel" as documented in the "Dictionary of Insults and Invectives"

Hanno Biber
Austrian Academy of Sciences
hanno.biber@oeaw.ac.at

Abstract

Satirical literary texts have certain properties that are highly interesting for the study of pejorative language use. The language of the satirical journal "Die Fackel" published and almost entirely written by Karl Kraus is the text basis for a text-lexicographic exploration into the field of pejorative language and its specific lexicographic units. The "Schimpfwörterbuch zu der von Karl Kraus 1899 bis 1936 herausgegebenen Zeitschrift "Die Fackel". Alphabetisches, Chronologisches, Explikatives" was published in 2008. The three volumes document the usage of invectives in the journal, in alphabetical and in chronological order, and in explicative form explained through the example of the last article of the journal. The alphabetical part consists of 2,775 examples of pejorative phrases and related indices. The chronological part presents 555 of these pejorative phrases arranged in chronological order providing expanded contexts. The third volume contains explicatory texts as well as "Wichtiges von Wichten", the final article of "Die Fackel", where pejorative phrases were marked up and accompanied by commentaries. This source is representing a literary genre that offers a variety of different forms of pejorative language to be studied from various perspectives. The lexicographic insights offered by the text dictionary into the use of pejoration by Karl Kraus will be presented in this paper.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

The Presence of Gender Issues in Spanish Dictionaries

Ana Costa Pérez
Universidad Carlos III de Madrid
Acosta.hum@uc3m.es

Abstract

Dictionaries are ideological creations as they are but a reflection of society itself. A dictionary sets a standard for language; they make an authority, a cultural product, and build a lexical *encyclopaedia* and a social reference. This analysis is built upon the idea of undeniable overlap between ideology and dictionary and the role of the latter as a mechanism to transmit the limited sights of everything around us; talking is a world-defining act, made by individuals, which forces themselves to adapt to a code which is seemingly open and closed at the same time; a code imposed by the society to which they belong and which will be enforced on future generations. Therefore, the present work will highlight the catalogue of definitions that challenge the descriptive neutrality of current lexicographical work, turned into dictionaries which should mirror an equal society, without discrimination.

This requires defining the concept from very different scopes: linguistic, anthropological, sociolinguistic, philosophical or cognitive. We aim at showing how grammar, with its two (and even up to three) *genera*, provides us with the perfect field to focus on sexes, at both biological and social senses, on Nature and Culture, without favouring the existence of two different sexes nor any individual

powers and decisions. The Spanish Language Academy (*hereinafter, the Academy*) notes that the Spanish language foresees the possibility to refer to mixed groups through the grammatical masculine gender, possibility in which there is no discriminatory intent, but the application of the linguistic law based in the expressive economy. Only when the opposition of the sexes is a relevant factor in the context, the Academy considers necessary the explicit presence of both genders.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Reflexive Verbs in a Valency Lexicon: The Case of Czech Reflexive Morphemes

Václava Kettnerová, Markéta Lopatková
Charles University in Prague
kettnerova@ufal.mff.cuni.cz, lopatkova@ufal.mff.cuni.cz

Abstract

In this paper, we deal with Czech reflexive verbs from the lexicographic point of view. We show that the Czech reflexive morphemes *se* and *si* constitute different linguistic meanings: either they are formal means of the word formation process of the so called reflexivization, or they are associated with the syntactic phenomena of reflexivity, reciprocity, and diatheses. All of these processes are associated with changes in the valency structure of verbs. We formulate a proposal for their lexicographic representation for the valency lexicon of Czech verbs, VALLEX. We make use of the division of the lexicon into a data component and a grammar component which represents a part of the overall Czech grammar. The data component stores information on valency structure of verbs in unmarked (active) structures. The grammar component consists of formal rules describing regular changes in the valency structure of verbs; these rules allow for the derivation of valency frames underlying the usages of verbs in marked structures (reflexive, reciprocal, deagentive and dispositional) from the valency frames corresponding to unmarked structures (non-reflexive, unreciprocal, and active). Czech reflexive verbs thus represent an illustrative example of the lexical-grammar interplay: we demonstrate that a close interaction between the lexicon and the grammar is necessary for a representation of these verbs and they both are indispensable if such a representation is to be adequate and economical.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

Polysemous Models of Words and Their Representation in a Dictionary Entry

Tinatini Margalidze
Lexicographic Centre at Ivane Javakhishvili Tbilisi State University
tinatin@margaliti.ge

Abstract

The paper deals with one of the universal models of polysemous adjectives and verbs, namely one-dimensional model and examines the ways of its representation in a dictionary entry. Polysemy is connected with the human perception and cognition of the world. It is determined by the process of perceiving not only particular objects and phenomena, but also the similarities existing between them or seen as such by members of the given language community. It is also connected with the ability of the language to reflect the new, yet un-cognized objects and phenomena by means of their associations and relations with already known, cognized objects and phenomena, i.e. to translate diversity of the world into linguistic unity. This makes polysemy an extremely interesting linguistic phenomenon but also leads to controversies concerning the interpretation of different issues connected with it. The paper touches upon some debatable issues connected with polysemy, such as: boundaries between senses, meaning and context, the role of context in the process of realization of meanings of polysemous words, meanings and sub-meanings, sense-numbering in a dictionary entry, etc. The paper also discusses some peculiarities of lexical meanings of adjectives and verbs.

Language of Presentation: English

Thu, 17/7/2014 – Room 1

One lexicological theory, two lexicographical models and the pragmatemes

Lena Papadopoulou
Hellenic Open University
papadopoulou.lena@gmail.com

Abstract

Generally little attention has been paid to pragmatics in most dictionaries. The present paper focuses on the concept of pragmatemes and their lexicographical treatment within the frame of Explanatory Combinatorial Lexicology. First, necessary preliminary notions closely related with pragmatemes are considered, by briefly reviewing the mel'čukian global model of human linguistic behaviour, linguistic sign and phrasemes typology. Second, the definition of pragmatemes, that is of phrasemes used in given extralinguistic situations, and the central acting part of the conceptual representation of the communicative situation are presented. Following, the structure of Explanatory Combinatorial Dictionary (ECD) and PragmatLex are outlined and an illustration of the Greek pragmatemes *Συγχαρητήρια* 'Congratulations' and *Συλλυπητήρια* 'Condolences' is provided within the simplified versions of ECD - *Dictionary of Collocations* and *Lexique actif du français*- and the lexicographical model for pragmatemes *PragmatLex*. Finally, we conclude our paper with a brief discussion on which lexicographical approach is preferable.

Language of Presentation: English

Sat, 19/7/2014 – Room 7

Analyzing Specialized Verbs in a French-Italian-English Medical Corpus: A Frame-based Methodology

Anna Riccio
University of Naples "L'Orientale"
ariccio@unior.it

Abstract

The aim of this study is to investigate the semantics and syntax of verbs in French, Italian, and English medical discourse by exploring the relationship between verb semantics and argument realization. The verbs under consideration are common lexical units which have acquired the status of a term through their specialization of meaning, such as *affect*, *involve*, etc. Unlike terminological verbs (e.g. *keratinize* or *lyophilize*), they have a lower level of technicalness, and co-occur with arguments (usually terms) in syntagmatic units. The data are extracted from the parallel EMEA corpus including documents published by the European Medicines Agency. The description of the verbs is based on the theoretical model of Frame Semantics (Fillmore 1977a-b, 1982, 1985; Fillmore and Atkins 1992) and on the FrameNet methodology (Ruppenhofer et al. 2010). The resultant analysis of the collected data reveals a sentence-level scenario (i.e., the *Damaging* frame) which groups together verbal forms which share similar syntactic and semantic valence patterns, both within and across languages.

Language of Presentation: English

Fri, 18/7/2014 – Room 7

Neoclassical Formatives in Dictionaries

Pius ten Hacken, Renáta Panocová
Leopold-Franzens-Universität Innsbruck, P.J. Šafárik University Košice
Pius.ten-Hacken@uibk.ac.at, Renata.Panocova@upjs.sk

Abstract

Neoclassical formatives are elements that occur in neoclassical word formations such as *ortho* and *paed(o)* in *orthopaedic*. We can be sure that a separate system of neoclassical word formation is in place when we find words such as *orthopaedic* that use classical elements but cannot have been borrowed from a classical language, because the concept they refer to did not exist yet. However, it must be taken into account that such words can also have been borrowed from a modern language that has such a system. In the lexicographic treatment of neoclassical word formation, a central question is whether neoclassical formatives should be treated in separate entries. We investigated how different English and Russian dictionaries treat them. In order to arrive at an unbiased sample of formatives, we used a Catalan word formation dictionary.

The results of our investigation support the hypothesis that neoclassical word formation constitutes a separate system in English, but not in Russian. This means that in English neoclassical formatives should have entries. In electronic dictionaries they can usefully be connected to the full class of words they appear in. In Russian, neoclassical formations are borrowings from languages such as English or French and their internal structure belongs to the domain of etymology.

Language of Presentation: English

Sat, 19/7/2014 – Room 7

Reports on Lexicographical and Lexicological Projects

Revision and digitisation of the early volumes of *Norsk Ordbok*: Lexicographical challenges

Sturla Berg-Olsen, Åse Wetås
Norsk Ordbok 2014, University of Oslo
sturla.berg-olsen@iln.uio.no, ase.wetas@iln.uio.no

Abstract

2014 will see the work on the 12th and final volume of the academic dictionary *Norsk Ordbok* (NO) finished. Still, the dictionary will remain heterogeneous due to variation in editorial practice throughout the project and incomplete in the sense that its early volumes are not digitally available. The online version of NO currently only covers the alphabet from the letter *i*. This paper describes the present state of the different parts of NO and argues that the early volumes of the dictionary must be revised and digitised to bring them up to the standards of the rest of the work. The revision and digitisation will not only give the dictionary a unitary profile but also make it possible to use it for a number of other purposes and facilitate the continuous process of keeping the dictionary up to date. The paper discusses some of the lexicographical challenges involved in the planned revision project and displays examples of the changes that must be made to the structure of the early material. It also touches upon questions concerning project organisation and funding.

Language of Presentation: English

Tue, 15/7/2014 – Room 2-3

A dictionary guide for Web users

Valeria Caruso, Anna De Meo
University of Naples 'L'Orientale'
vcaruso@unior.it, ademeo@unior.it

Abstract

The continuously growing number of specialised lexicographical resources on the Web calls into question the users' ability to solve their information needs autonomously. Neither terminological databanks, nor dictionary aggregators actually represent valuable solutions to these needs (Lew 2011), and in order to guide users towards useful resources, a database was created, which collects evaluation forms of free internet specialised dictionaries and allows users to carry out customised searches on the basis of their subject field expertise (laymen, semi-experts, experts), the desired language, and the kind of support they need (basically with communicative problems or in acquiring new knowledge).

Using a specific evaluation system, the tool displays the best resources available for the desired parameters, assessing dictionaries on the basis of an evaluation scale and explicit guidelines that prevent contradictory responses, such as dictionaries that are simultaneously suited for laymen and experts.

The paper illustrates the current development of the tool, with special reference to its evaluation system, as well as its possible future improvements.

Language of Presentation: English

Tue, 15/7/2014 – Room 2-3

What a multilingual loanword dictionary can be used for: searching the *Dizionario di italianismi in francese, inglese, Tedesco (DIFIT)*

Matthias Heinz, Anne-Kathrin Gärtig
Universität Salzburg
matthias.heinz@sbg.ac.at, anne-kathrin.gaertig@sbg.ac.at

Abstract

The poster outlines the structure and practical use of the *Dizionario di italianismi in francese, inglese, tedesco* (DIFIT), a dictionary registering Italian loanwords in three languages published in print and currently being prepared for digital reedition. This specialized lexicographical resource focuses on language contact resulting in synchronic and diachronic lexical transfer from the Italian language to French, English and German. Italian as a culturally extremely rich and diverse source for borrowings has extensively influenced other languages in different historic phases all along the past centuries. The DIFIT reflects this rich European heritage of borrowings in a maximal variety of lexical fields.

Major methodological aspects in and prior to editing the dictionary included the delimitation and exploration of the source material and deciding on depth of historical coverage as well as defining the information programme of the microstructure. At the same time, limitations of the general lexicographical documentation on both Italian and the three recipient languages presented particular challenges. The poster displays results of some first quantitative research on linguistic aspects of the borrowed elements in DIFIT and gives an outlook on the ongoing project of digitizing the lexicographical data by means of an online database documenting the linguistic impact of Italian as a global donor language (*Osservatorio degli italianismi nel mondo/OIM*, hosted by the Accademia della Crusca).

Language of Presentation: English, Italian, German

Thu, 17/7/2014 – Corridor seminar rooms

The Basic Estonian Dictionary: the first monolingual L2 learner's dictionary of Estonian

Jelena Kallas, Maria Tuulik, Margit Langemets
Institute of the Estonian Language
jelena.kallas@eki.ee, maria.tuulik@eki.ee, margit.langemets@eki.ee

Abstract

This paper is a report on a lexicographical project completed by the Institute of the Estonian Language. The Basic Estonian Dictionary was published in print in March 2014, and the online version will be available by September 2014. The BED is aimed at learners of Estonian as a foreign language or as a second language at the elementary and intermediate levels (A2 to B1) according to the Common European Framework of Reference for Languages.

The dictionary contains about 5,000 headwords, including single items and multi-word lexical items. The BED provides lexicographical information on pronunciation, morphological information, definitions, word formation, government and collocation patterns, multi-word phrases, semantically related words and usage notes. In the online version, sound recordings (mp3 audio files) are provided. Morphological information was generated automatically. The most frequent government and collocation patterns were analysed and selected using the Sketch Engine corpus query system. The BED contains approx. 400 illustrations, study pages and picture pages (e.g. related to animals). In the appendix, geographical names and grammar tables are included.

In addition, the Dictionary of the Estonian Sign Language (containing approx. 6,700 video recordings), based on the BED database, was published online in March 2014.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Creating a German-Basque electronic dictionary

David Lindemann
UPV-EHU University of the Basque Country
david.lindemann@ehu.es

Abstract

With this poster, we introduce the new electronic dictionary project EuDeLex, that is currently being worked on at UPV-EHU University of the Basque Country (<http://www.ehu.es/eudelex/>). The primarily targetted user groups are Basque-L1 German learners and German to Basque translators.

The language pair German-Basque, which can be called less-resourced or medium-density, does not count with any lexicographical antecedents that could be updated or adapted. Nevertheless, the editing process of the new dictionary is supported by existing monolingual lexicographical databases and a newly set up German-Basque parallel corpus. We explain our proposals regarding a

macrostructure based on corpus frequency data for German and Basque and a microstructure for the German-Basque part of EuDeLex, as well as a proposal for an online user interface and the publishing process of a first version.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Die fremdsprachige Produktionssituation im Fokus eines onomasiologisch-konzeptuell orientierten, zweisprachig-bilateralen Wörterbuches für das Sprachenpaar Deutsch - Spanisch: Theoretische und methodologische Grundlagen von DICONALE

Meike Meliss
Universidad de Santiago de Compostela
meike.meliss@usc.es

Abstract

Der Beitrag beschäftigt sich mit den verschiedenen Such-, Auffindungs- und Auswahlprozessen, die für die fremdsprachige Produktion notwendig sind und von DICONALE-*online*, einem onomasiologisch-konzeptuell ausgerichteten, zweisprachig-bilateral konzipierten Verbwörterbuch der spanischen und deutschen Gegenwartssprache, besonders berücksichtigt werden. Der Ausgangspunkt von DICONALE ist ein unbefriedigendes Informationsangebot in den bestehenden ein- und zweisprachigen Lernerwörterbüchern für den L2-*output* und bestätigt das Projektteam in der Notwendigkeit, ein neuartiges benutzer- und situationsdefiniertes *online*-Nachschlagewerk zu erstellen. Zwei Bezugsrahmen bilden die Grundlage für einen komplexen, konzeptuell und fragegeleiteten Zugriffspfad, der dem Benutzer bei der Suche und Auswahl von Ausdrucksmöglichkeiten und der adäquaten Anwendung behilflich sein soll. Das Novum dieses Wörterbuchprojekts besteht hauptsächlich darin, eine onomasiologisch-konzeptuelle Perspektive für den fremdsprachigen Produktionsprozess nutzbar zu machen und mit einem semasiologischen Zugriff zu verbinden, durch den es möglich ist, die inter- und intralingualen Unterschiede zwischen den Lexemen eines lexikalisch-semantischen (Sub)Paradigmas hervorzuheben.

Ziel des Beitrages ist es daher, den Ausgangspunkt, sowie die theoretischen und methodologischen Grundlagen von DICONALE-*online* unter der speziellen Perspektive der Benutzer- und Situationsorientiertheit zur Diskussion zu stellen, die einzelnen Zugriffspfade für den Such- und Auffindungsprozess.

Language of Presentation: German

Wed, 16/7/2014 – Room 2-3

Graph-Based Representation of Borrowing Chains in a Web Portal for Loanword Dictionaries

Peter Meyer
Institut für Deutsche Sprache, Mannheim
meyer@ids-mannheim.de

Abstract

This paper reports on an ongoing lexicographical project that investigates Polish loanwords from German that were further borrowed into the East Slavic languages Russian, Ukrainian, and Belorussian. The results will be published as three separate dictionaries in the Lehnwortportal Deutsch, a freely available web portal for loanword dictionaries having German as their common source language. On the database level, the portal models lexicographical data as a cross-resource directed acyclic graph of relations between individual words, including German 'metalemmata' as normalized representations of diasystemic variants of German etyma. Amongst other things, this technology makes it possible to use the web portal as an 'inverted loanword dictionary' to find loanwords in different languages borrowed from the same German etymon. The different possible pathways of German loanwords that went through Polish into the East Slavic languages can be represented directly as paths in the graph. A dedicated in-house dictionary editing software system assists lexicographers in producing and keeping track of these paths even in complex cases where, e.g., only a derivative of a German loanword in Polish has been borrowed into Russian. The paper concludes with some remarks on the particularities of the dictionary/portal access structure needed for presenting and searching borrowing chains.

Language of Presentation: English

Tue, 15/7/2014 – Room 2-3

The Problem of Scale. Handling (More) Data with the *Wiki Lexicographica*

Bruno Bon, Krzysztof Nowak

Institut de recherche et d'histoire des textes, CNRS; Institute of the Polish Language, Polish Academy of Sciences
bruno.bon@irht.cnrs.fr, krzysztof@ijp-pan.krakow.pl

Abstract

The main purpose of the *WikiLexicographica* was to provide a unified access to several various dictionaries of Medieval Latin. The prototype which was constructed offered clear separation of the basic and advanced views, rich visualisation capabilities, advanced search and browsing features. The test interface was, subsequently, populated with significantly larger amount of lexicographic data from three various lexicographic works. Although some of the features of the original design (namespaces, categorising etc.) scale very well, others seem to pose a challenge to the application of the Semantic MediaWiki as a dictionary tool. In the software presentation, two major problems will be discussed. As a means of dealing with visualisation issues, data reductions techniques will be presented. Problem of inconsistencies between microstructures of the dictionaries, in turn, needs to be solved by applying common ontologies.

Language of Presentation: English

Wed, 16/7/2014 – Room 8

At the Beginning of a Compilation of a New Monolingual Dictionary of Czech (A Report on a New Lexicographic Project)

Pavla Kochová, Zdeňka Opavská, Martina Holcová Habrová

Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i.
kochova@ujc.cas.cz, opavska@ujc.cas.cz, holcova@ujc.cas.cz

Abstract

The aim of the article is to present the new lexicographic project that is being implemented at the Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i. Since 2012, its Department of Contemporary Lexicology and Lexicography has worked on the creation of a new medium-sized monolingual dictionary of Czech with the working title *Akademický slovník současné češtiny* (The Academic Dictionary of Contemporary Czech). With its size and method of treatment, the dictionary ranks among academic dictionaries, i.e. dictionaries with an elaborated, standardised and structured explanation of the meaning of lexical units, with an adequately rich exemplification documenting the typical use of lexical units, with a sufficiently elaborated description of the basic semantic relations, mainly synonymy and antonymy, with the appropriate description of the grammatical properties of lexical units and with usage labels (a description of the stylistic, temporal, spatial, frequency and pragmatic markedness) of lexical units.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Frame Semantics and Learner's Dictionaries: Frame Example Sections as a New Dictionary Feature

Carolin Ostermann

Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
carolin.ostermann@fau.de

Abstract

Frame semantics has so far been neglected or even been rejected in the context of EFL-lexicography, although lexicographic description within a frame semantics approach would have advantages for learners, e.g. the coherent presentation of several relevant lexical items at a time, as well as their conceptual connection, both of which would also further vocabulary acquisition. This proposal will detail how a frame semantics approach for the example section in English monolingual learner's dictionaries can contribute to the notion of cognitive lexicography, i.e. a lexicography that puts an emphasis on how users process language, which would in turn facilitate a user's understanding of an entry. For this purpose, so-called *frame example sections* were developed on agentive nouns (e.g. *bridegroom*, *plaintiff*); these are small coherent text passages that define and exemplify the noun in relation to its whole frame. The frame example sections mention related frame elements, collocating verbs and describe the typical scenario underlying a semantic frame, in order to promote decoding, i.e. understanding the meaning of lexical items, as well as encoding, i.e. learning words and finding related language material. The paper will be rounded off by presenting the results of a small user-study that was conducted on the frame example sections.

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

Translating action verbs using a dictionary of images: the IMAGACT ontology

Alessandro Panunzi[°], Irene De Felice*, Lorenzo Gregori[°], Stefano Jacoviello[†], Monica Monachini*,
 Massimo Moneglia[°], Valeria Quochi*, Irene Russo*
[°]University of Florence, *ILC CNR (Pisa), [†]University of Siena
 alessandro.panunzi@unifi.it, irene.defelice@ilc.cnr.it, lorenzo.gregori@unifi.it, stefano.jacoviello@gmail.com,
 monica.monachini@ilc.cnr.it, moneglia@unifi.it, valeria.quochi@ilc.cnr.it, irene.russo@ilc.cnr.it

Abstract

Action verbs have many meanings, covering actions in different ontological types. Moreover, each language categorizes action in its own way. One verb can refer to many different actions and one action can be identified by more than one verb. The range of variations within and across languages is largely unknown, causing trouble in all translation tasks. IMAGACT is a corpus-based ontology of action concepts, derived from English and Italian spontaneous speech corpora, which makes use of the universal language of images to identify the different action types extended by verbs referring to action in English, Italian, Chinese and Spanish. This paper presents the IMAGACT search interface and the various linguistic information the user can derive from it. IMAGACT makes explicit the variation of meaning of action verbs within one language and allows comparisons of verb variations within and across languages. Because the action concepts are represented with videos, extension into new languages beyond those presently implemented in IMAGACT is done using competence-based judgments by mother-tongue informants, without intense lexicographic work involving underdetermined semantic descriptions.

Language of Presentation: English

Wed 16, Thu 17/7/2014 – Room 8

Degrees of Synonymity as the Basis of a Network for German Communication Verbs in the Online Reference Work *Kommunikationsverben* in OWID

Kristel Proost, Carolin Müller-Spitzer
 Institut für Deutsche Sprache, Mannheim
 proost@ids-mannheim.de, mueller-spitzer@ids-mannheim.de

Abstract

This contribution presents the procedure used in the *Handbuch deutscher Kommunikationsverben* and in its online version *Kommunikationsverben* in the lexicographical internet portal OWID to divide sets of semantically similar communication verbs into ever smaller sets of ever closer synonyms. *Kommunikationsverben* describes the meaning of communication verbs on two levels: a lexical level, represented in the dictionary entries and by sets of lexical features, and a conceptual level, represented by different types of situations referred to by specific types of verbs. The procedure starts at the conceptual level of meaning where verbs used to refer to the same specific situation type are grouped together. At the lexical level of meaning, the sets of verbs obtained from the first step are successively divided into smaller sets on the basis of the criteria of (i) identity of lexical meaning, (ii) identity of lexical features, and (iii) identity of contexts of usage. The stepwise procedure applied is shown to result in the creation of a semantic network for communication verbs.

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

Job-hunting in Italy: Building a glossary of “English-inspired” job titles

Virginia Pulcini, Angela Andreani
Università degli Studi di Torino
virginia.pulcini@unito.it, angela.andreani@unito.it

Abstract

This paper reports on a study of “English-inspired” job titles retrieved from a specialized corpus of job advertisements posted on Italian web pages. This corpus was created using the WebBootCat tool in the Sketch Engine, following the methodology described by Baroni and Bernardini (2004) and Baroni et al. (2006). The aim is to build a glossary of English job titles to be published online as a tool for prospective job applicants. Checking their status in English and Italian dictionaries, we will establish whether the titles collected are current English terms, false Anglicisms, or English-inspired creations. The preliminary findings consist of a list of 30 job titles which are analyzed in terms of form and meaning, and grouped into categories depending on whether an Italian equivalent is available or not. The corpus of job postings is used to analyze the lexical profile of job titles, their meaning and/or possible covert manipulative intent. In fact, data shows that some English job titles may be preferred to Italian equivalents to attribute greater status to the actual job designation and description. Moreover, some job titles are characterized by complex pre-modification which may confuse the ultimate users, i.e. job hunters themselves.

Language of Presentation: English

Wed, 16/7/2014 – Room 1

A small dictionary of life under communist totalitarian rule (Czechoslovakia 1948-1989)

Věra Schmiedtová
The Institute of the Czech national corpus, Charles University in Prague
vera.schmiedtova@ff.cuni.cz

Abstract

The intention is to preserve the vocabulary of the period for the younger generation; also to remind the older generation of vocabulary that they used to encounter, but are gradually forgetting. The dictionary is specific in that it is made up of two types of vocabulary – the language of communist propaganda and the spoken language emerging from how people reacted to the pressure of propaganda, often including popular humour. The first type of vocabulary has been collated through Corpus of Totalitarianism, for the second type a corpus-based source does not exist (it is the language as spoken, which it was possible to collate through quotes from fiction, journalistic writings and from the author’s own observations. This language has been checked in contemporary written corpora, on some occasions it is to be found in the Corpus of Totalitarianism or on the internet).

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

A Frequency Dictionary of Dutch

Carole Tiberius, Tanneke Schoonheim, Adam Kilgarriff
Institute of Dutch Lexicology, Lexical Computing Ltd
{carole.tiberius,tanneke.schoonheim}@inl.nl, adam@lexmasterclass.com

Abstract

In this paper, we present a corpus-based frequency dictionary of Dutch containing the 5000 most frequent words of Dutch. The dictionary has been published at the beginning of 2014 as part of the Routledge Frequency Dictionaries series, a well-established series with titles available for 11 languages at the time of writing. Novel in the Dutch frequency dictionary is that genre has been foregrounded. The dictionary does not contain one single frequency list, but multiple lists are presented, of which four are genre specific covering fiction, newspaper, spoken and web. Throughout the dictionary there are also thematically organised lists featuring the top words from a variety of key topics such as animals, food and other areas of daily and cultural life. Words specific to Dutch in Belgium are also included. The dictionary is based on a 290-million-word corpus which includes both written and spoken material from a wide range of sources.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

The Corpus of the Croatian Church Slavonic Texts and the Current State of Affairs Concerning the Dictionary of the Croatian Redaction of Church Slavonic Compiling

Vida Vukoja
Old Church Slavonic Institute, Zagreb
vidal@stin.hr

Abstract

Croatian Church Slavonic is a literary, bookish idiom used in Croatia from the XI/XII until the XVII c. based on Old Church Slavonic, an idiom created by Sts. Cyril and Methodius, and shaped by the Croatian vernacular.

The Croatian Church Slavonic corpus consists of the texts excerpted from 11 breviaries, 4 missals, 3 psalters, 3 rituals, 15 miscellanies. It also incorporates all the 26 fragments dated from the period up to and including the XIII c. and several auxiliary sources. The corpus was created over a period of thirty years (from 1959 to the early 1990s) at the Old Church Slavonic Institute in Zagreb. It is a historical, referential, representative paper card-file of excerpts. It is also a parallel corpus, as it contains Latin and Greek parallel texts, those that were identified as closest to the actual source texts for the translational Croatian Church Slavonic texts.

The *Dictionary of the Croatian Redaction of Church Slavonic* has been compiled on the basis of the Croatian Church Slavonic corpus. The fascicles of the *Dictionary* have been published since 1991. So far, 1 (1991)–19 (2012), with the dictionary articles A–ŽRBTV (according to the Old Cyrillic alphabet) have been printed.

Language of Presentation: English

Wed, 16/7/2014 – Room 2-3

Others

Considerations about Gender Symmetry in the Dictionary of Bavarian Dialects in Austria

Isabella Flucher, Eveline Wandl-Vogt, Thierry Declerck
Austrian Academy of Sciences, ICLTT, Austria; DFKI GmbH, Language Technology Lab, Germany
bella_flu@yahoo.de; eveline.wandl-vogt@oeaw.ac.at; declerck@dfki.de

Abstract

This poster summarizes the first results of a study that has been pursued as part of an internship at the Austrian Academy of Sciences. The aim was to investigate cases of gender symmetry or asymmetry in a dictionary. As case study we focused on a traditional dialectal dictionary. An annotation schema has been developed and first natural language processing steps have been established. In this poster, forms of gender symmetry as well as asymmetry are presented, on the basis of analysis of the vocabulary employed in example sentences or excerpts used in the dictionary. The analysis of gender asymmetry was also based on the consideration of a selection of derogatory names. The work described in this poster provides a critical insight into lexicographical work, design and implementation from a feminist perspective and opens new perspectives for the development of gender-symmetric lexicographic works.

Language of Presentation: English

Thu, 17/7/2014 – Corridor seminar rooms

Advancing Search in the Algemeen Nederlands Woordenboek

Carole Tiberius, Jan Niestadt, Lut Colman, Boudewijn van den Berg
Institute of Dutch Lexicology, belberg
{carole.tiberius,jan.niestadt,lut.colman}@inl.nl, boudewijn@belberg.eu

Abstract

In this paper, we will take a closer look at the advanced search option in online dictionaries from the perspective of the user. In the context of dictionaries, the advanced search allows users to retrieve sets of words which match a particular description, for example 'archaic compounds consisting of three syllables'. However, the possible sets of words that could be retrieved are endless, and the challenge is how to present all these options to the user in a way that he can grasp and understand. Studies on dictionary use and log file analyses suggest that existing solutions offered by online dictionaries are not very successful as users do not really seem to use this search option. We will discuss different types of advanced search that can be distinguished and we will present a new, more user-friendly approach to advanced search in dictionaries using the *Algemeen Nederlands Woordenboek* as our test case.

Language of Presentation: English

Wed 16, Thu 17/7/2014 – Room 8

List of Authors

Adachi Ryoko	32
Alonso Campo Araceli	45
Andreani Angela	80
Apel Ulrich	32, 39
Arts Tressy	15
Baisa Vít	39
Baiwir Esther	65
Baldwin Timothy	29
Bańko Mirosław	69
Barbierik Kamil	15, 16
Bargalló María	40
Battista Simonetta	18
Benko Vladimír	35
Berg-Olsen Sturla	75
Berti Barbara	39
Betbeder Marie-Laure	49
Biber Hanno	69
Bon Bruno	78
Bosch Sonja	53
Bothma Theo J.D.	55
Buendía Castro Miriam	45
Calderone Basilio	35
Caruso Valeria	75
Castell Andreu	40
Castellarin Michele	29
Català Natàlia	40
Čermák František	16
Chiari Isabella	35
Chiocchetti Elena	45
Coffey Stephen	59
Colman Lut	85
Convery Cathal	55
Cook Paul	29
Costa Pérez Ana	69
Cresti Emanuela	36
Czajkowska-Kisil Małgorzata	31
Dallaji Ines	31
De Felice Irene	79
De Mauro Tullio	35
De Meo Anna	75
DeCesaris Janet	59
Declerck Thierry	30, 85
Děngeová Zuzana	15, 16
Didakowski Jörg	30
Dobrovol'skij Dmitrij	59
Dotto Diego	65
Dziemianko Anna	23
Faaß Gertrud	53
Faber Pamela	45
Fábián Zsuzsanna	40
Fabijanić Ivo	16
Fabre Cécile	46
Fanderl Wolfgang	32
Flinz Carolina	23
Flucher Isabella	85
Fontenelle Thierry	9
Gagliardi Gloria	17
Galafacci Gisele	42
Gantar Polona	30
Gärtig Anne-Kathrin	76
Gavriilidou Zoe	23
Germain Bérenger	49
Giacoma Luisa	41
Giacomini Laura	60

Giovannini Michela	17
Gornostay Tatiana	46
Gouws Rufus H.	48
Gregori Lorenzo	36
Hajič Jan, jr.	32
Hathout Nabil	35
Hebal-Jeziarska Milena	69
Heid Ulrich	9, 55
Heinonen Tarja Riitta	60
Heinz Matthias	76
Holcová Habrová Martina	78
Hult Ann-Kristin	24
Hvelplund Holger	18
Jacoviello Stefano	79
Jakubiček Miloš	39
James Mafela Munzhedzi	42
Jarý Vladimír	15, 16
Johannsson Ellert Thor	18
Josselin-Leray Amélie	46
Kageura Kyo	32
Kallas Jelena	76
Kettnerová Václava	70
Kilgarriř Adam	39, 80
Kochová Pavla	78
Køhler Simonsen Henrik	24
Koplenig Alexander	26
Koponen Eino	53
Kosem Iztok	30
Kovář Vojtěch	39
Krek Simon	30
Krisch Jennifer	47
Łacheta Joanna	31
Langemets Margit	76
Lau Jey Han	29
Lea Diana	18
León-Araúz Pilar	47
Lew Robert	25
Lindemann David	41, 76
Linde-Usiekniewicz Jadwiga	31
Liška Tomáš	15, 16
Lišková Michaela	15, 16
Logar Nataša	30
Lopatková Markéta	70
Lorentzen Henrik	19
Manterola Iker	41
Marello Carla	10
Margalitadze Tinatin	71
Mayer Martina	48
Meier Stefanie	24
Meliss Meike	77
Metsmägi Iris	19
Meyer Peter	26, 77
Miyata Ray	32
Monachini Monica	79
Moneglia Massimo	79
Moon Rosamund	11
Mörth Karlheinz	31
Mulhall Chris	54, 61
Müller-Spitzer Carolin	26, 79
Murayama Ryo	32
Navigli Roberto	31
Nazar Rogelio	41
Nied Curcio Martina	25
Niestadt Jan	85
Nimb Sanni	19
Nowak Krzysztof	55, 78
Ó Mianáin Pádraig	55

ó'Diollúin Seamus	54
Oja Vilja	19
Opavská Zdeňka	78
Ostermann Carolin	78
Panocová Renáta	72
Panunzi Alessandro	36, 79
Papadopoulou Lena	36, 71
Picton Josette Rebeyrolle Aurélie	46
Pinnavaia Laura	39
Plaisantin Alecu Blandine	49
Planas Emmanuel	46
Prinsloo Danie J.	48, 55
Procházka Stephan	31
Proost Kristel	79
Ptasznik Bartosz	25
Pulcini Virginia	80
Quochi Valeria	79
Radtke Nadja	30
Ralli Natascia	45
Reimerink Arianne	47
Renders Pascale	65
Riccio Anna	72
Rodríguez Sánchez Ignacio	61
Rundell Michael	29
Russo Irene	79
Rutkowski Paweł	31
Rychlý Pavel	39
Sajous Franck	35
San Vicente Iñaki	41
Saralegi Xabier	41
Schmiedtová Věra	80
Schnörch Ulrich	36
Schoonheim Tanneke	80
Silvestre João Paulo	65
Šindlerová Jana	32
Šorli Mojca	36
Squillante Luigi	61
Stantcheva Diana	48
Storjohann Petra	36
Takeuchi Koichi	32
Taljard Elsabé	48
ten Hacken Pius	72
Tenório Zucchi Angela Maria	26
Thomas Izabella	49
Tiberius Carole	80, 85
Tosques Fabio	29
Trap-Jensen Lars	19
Tuulik Maria	76
van den Berg Boudewijn	85
Vasiljevs Andrejs	46
Veldi Enn	56
Veselovská Kateřina	32
Villalva Alina	65
Virus Miroslav	15, 16
Vogel Iris	32
Vukoja Vida	81
Wandl-Vogt Eveline	30, 85
Wetås Åse	75
Widmer Alessandra	24
Williams Geoffrey	59
Wolfer Sascha	26
Zavaglia Adriana	42

Under the patronage of

AUTONOME
PROVINZ
BOZEN
SÜDTIROL



PROVINCIA
AUTONOMA
DI BOLZANO
ALTO ADIGE

PROVINCIA AUTONOMA DE BULSAN
SÜDTIROL



Città di Bolzano
Stadt Bozen



ACCADEMIA DELLA CRUSCA

With the support of

Sprachstelle
im Südtiroler KULTURinstitut

Sketch  Engine

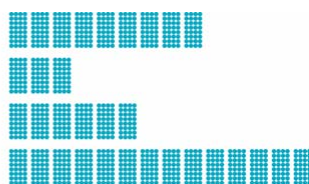


KDICTIONARIES

Hornby

A. S. Hornby Educational Trust

ISTITUT
LADIN
MICURÀ
DE RÜ



SDLO 

SÜDTIROL



OXFORD
UNIVERSITY PRESS