XIII Euralex International Congress 25 years studying dictionaries

July 15 - 19 2008

Book of abstracts

- 0. Plenary Lectures
- 1. Computational Lexicography and Lexicology
- 2. The Dictionary-Making Process
- 3. Reports on Lexicographical and Lexicological Projects
- 4. Bilingual Lexicography
- 5. Lexicography for Specialised Languages Terminology and Terminography
- 6. Historical and Scholarly Lexicography and Etymology
- 7. Dictionary Use
- 8. Phraseology and Collocation
- 9. Lexicological Issues of Lexicographical Relevance
- 10. Other Topics

FrameNet Meets Construction Grammar

Fillmore, Charles J.

O. Plenary Lectures

Plenary Lecture

Lexical Patterns: from Hornby to Hunston and beyond

Hanks, Patrick

O. Plenary Lectures

Plenary Lecture

Twenty-five years of dictionary research: Taking stock of conferences and other lexicographic events since LEXeter'83

Hartmann, Reinhard R. K.

♠ 0. Plenary Lectures

Plenary Lecture

Sobre la discontinuidad de las palabras en un diccionario histórico

Pascual, José Antonio

O. Plenary Lectures

Plenary Lecture

Tradició i innovació en la lexicografia catalana

Rafel i Fontanals, Joaquim

♠ 0. Plenary Lectures

Plenary Lecture

Approaches to Computational Lexicography for German Varieties

Abel, Andrea; Anstein, Stefanie

1. Computational Lexicography and Lexicology

Poster

Corpora built for linguistic varieties of a pluricentric language such as German are an indispensable

resource for a detailed and systematic variety comparison and dictionary development. We present desiderata and suggestions as well as methods from computational linguistics to systematically apply variety corpora for the enrichment, i.e. confirmation, extension and generation, of lexical entries in distinctive variant dictionaries for German. Examples are those variant dictionaries developed by Ammon et al. (2004) and Abfalterer (2007), where we focus on the South Tyrolean German language. On the one hand, we conducted a systematic frequency analysis in newspaper variety corpora for approved lists of South Tyrolean special vocabulary in order to possibly refine corresponding dictionary entries with corpus evidence. On the other hand, we filtered the list of words of our South Tyrolean corpus which could not be lemmatised by a tool developed for the variety in Germany. After removing special vocabulary collected for the South Tyrolean variety in other projects-e.g. legal terms, the remaining list was manually checked for possible new variant dictionary entries, thus-as an innovative variety corpus lexicographic approach-also automatically filtering a huge amount of data to extract only relevant data to be investigated in detail. In addition, we semi-automatically extracted lexical cooccurrences of our two newspaper corpora and compared their frequencies-with the assumption that those cooccurrences are worth being more closely investigated that have high frequency in the South Tyrolean corpus and very low frequency in the corpus from Germany. With these three methods we were not only able to refine dictionary entries for South Tyrolean German, but also to add new ones. The findings on variants can be re-used for further corpus annotation resulting in again better resources for computational variant lexicography of the kind described, which is also to be extended to more complex linguistic levels.

AnCora-Verb: Two Large-scale Verbal Lexicons for Catalan and Spanish

Aparicio, Juan; Taulé, Mariona; Martí, M.Antònia

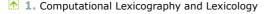


Short Paper

In this paper, *AnCora-Verb* is presented: two large-scale verbal lexicons used for the semantic annotation with arguments, thematic roles and semantic class of AnCora corpora (AnCora-Cat for Catalan and AnCora-Esp for Spanish). Each corpus contains 500,000 words with a multilayer annotation in different linguistic fields-from morphology to pragmatics. AnCora-Verb lexicons focuses on syntactic functions, arguments and thematic roles of each verbal predicate taking into account the verbal semantic class and those alternations in diathesis where the predicate can participate. This paper concentrates on the definition and characterization of verb classes and the criteria followed in the assignment of a verb to a specific class.

Multi-level Reference Hierarchies in a Dictionary of Swahili

Bański, Piotr; Wójtowicz, Beata



Short Paper

This paper can be classified into at least two categories: Computational lexicographty and Reports on lexicographical projects, bordering on yet another, the dictionary-making process. The context is a lexicographic project that creates an electronic, TEI XML-encoded Swahili-Polish learner dictionary-with a goal of 10 000 entries in the first stage. Here, we focus on one of the innovative features that we want to introduce in the dictionary, at a relatively small cost-due to the way the dictionary will be compiled out of a Swahili corpus: explicit visualization of derivational hierarchies-essentially a learner-oriented feature, but also serves as a basis for further lexicographic/lexicological applications. We primarily discuss our motivation for this idea and its XML implementation. Nevertheless, by the Conference date, we should also be able to present an actual visualization of it, going beyond a mere set of colourful hyperlinks, which is the way it is presented in our test dictionary-composed of 300 hundred selected illustrative entries, currently being expanded to 1500, for database testing.

Multidimensional Ontologies: Integration of Frame Semantics and Ontological Semantics

Barzdiņs, Guntis; Grūzītis, Normunds; Nespore, Gunta; Saulīte, Baiba; Auziņa, Ilze; Levāne-Petrova, Kristīne

1. Computational Lexicography and Lexicology

Poster

Today FrameNeta-a state-of-the-art implementation of frame semantics-provides one of the best insights into lexical semantics and their interaction with the syntactic structure of the sentence. The main limitation of the current implementation is the insufficient level of formalization of frame

descriptions, making it unsuitable for automatic text annotation without human supervision.

Meanwhile, FrameNet usability would greatly benefit from more rigorous formalization and the consequential possibility for automatic annotation. Previous attempts at formalization have focused on enforcing strict ontological control of the *semantic types* for the frame fillers-despite their insignificant use-due to high ambiguity-in the actual FrameNet. We propose a different approach relying on representation of FrameNet as a 4D multidimensional ontology that allows capturing of the "precedent" knowledge encoded in the manually annotated texts, like FrameNet's full-text annotation reports. This allows both to re-create FrameNet ontology from semantically annotated texts, as well as to use this representation for semantic annotation of new texts. Further extensions of this approach with 5th dimension for anaphora annotation is discussed as an alternative for the informal semantic type mechanism of FrameNet.

The Structure of the Lexicon in the Task of Automatic Lexical Acquisition

Bel, Núria; Espeja, Sergio; Marimon, Montserrat



Poster

In the task of automatic lexical acquisition, i.e. the induction of lexical information from texts, there have been no attempts to exploit theoretically-based models of the structure of the lexicon. Works like those of Bybee (1988) and Langacker (1987) propose a highly structured lexicon where words are related paradigmatically by phonological similarity and where lexical features are an emergent characteristic of the resulting structure. If so, a machine learning algorithm such as a Decision Tree (DT, Quinlan, 1945) should be able to learn the correlation between particular lexical features and the formal characteristics of words. In our experiment, the machine learner should be able to find a correlation between characters that form the words used for training it and the nominal feature /mass/. The ability of the trained learner to predict correctly whether nouns that it has not been shown in the training phase are mass nouns or not is proof that such a correlation exists and that it can be considered an emergent feature of the paradigmatic relations that relate words in the lexicon. The obtained results prove that a structured lexicon can provide information on lexical features.

SAOL Plus-A New Swedish Electronic Dictionary

Berg, Sture; Holmer, Louise; Hult, Anki

1. Computational Lexicography and Lexicology

Software Demo

In September 2007, a CD version of the Swedish Academy Glossary, SAOL Plus, was released. In SAOL Plus, all inflected forms are shown in full text and virtually every text fragment is searchable. Standard search functions include Search lemma, Search inflected forms and Search article text. Advanced searches can be made with the usual wild cards. It also has an advanced tool for fuzzy search based on pronunciation.

SAOL Plus can be an asset for the public as well as an efficient and functional utensil for linguists. Moreover, thanks to the fuzzy search, it is useful for people with reading and writing disorders, as well as secondary language users.

System requirements: Windows 98/NT/2000/XP/Vista.

Matching Verbo-nominal Constructions in FrameNet with Lexical Functions in MTT Bouveret, Myriam; Fillmore, Charles, J.

1. Computational Lexicography and Lexicology

Full Paper

Matching verbo-nominal constructions in FrameNet with Lexical Functions in MTT Multiword expressions are described in the FrameNet project as complex lexical units linked as wholes to their semantic frames, or as instances of special grammatical constructions. Amongst multi-word expressions, we explore in this paper verbo-nominal constructions in Framenet and more specifically support verb expressions. This category, described in FrameNet litterature, has not yet received a systematic treatment in the project. We have looked at several problems regarding its actual classification and theoretical background. Instead of light verbs, or support verbs, we define the object of our study as Support Verb Constructions (SVC). We explore in this connection a well known model in lexicography, the Explanatory and Combinatorial Lexicology, part of the Meaning Text Theory to

observe the feasibility of using Lexical Functions in FrameNet for the purpose of encoding SVC. The verbs have functions varying from that of light verbs, where they contribute only the ability to treat the noun's frame as a verbal-e.g., tense-bearing, entity-through a variety of aspect, perspective, and register values. These structures are treated in various ways in FrameNet and in the Explanatory and Combinatorial component of the Meaning-Text-Theory.

Our goal in the present research is to understand the nature of these differences, and to consider whether the results obtained in one of them can be aligned with or incorporated into the other. We approach the parameters of comparison of the two systems by means of three themes characterizing verbo-nominal expressions:

- i. lexicalized collocations of the verb and the nominal head of its syntactic dependent,
- ii. the manner in which the kind of situation-the semantic frame-evoked by the noun is given verbal expression,
- iii. and the manner in which the syntactic arguments of the verb are interpreted as matching the semantic roles associated with the noun's frame.

Syntactic Behaviour and Semantic Kinship of Selected Danish Verbs

Braasch, Anna

erwie –

1. Computational Lexicography and Lexicology

Full Paper

The paper discusses relationships between the syntactic behaviour and meaning of selected verbs, with the focus on exploiting observable syntactic similarities for uncovering of semantic kinship. The investigation is inspired by the demand in language technology for large-scale lexicons that combine morphological, syntactic and semantic descriptions of the lemmas. The development of such a lexical resource is rather demanding, therefore, an enhancement of existing resources with additional information types is a worthwhile task. The computational lexicon for *Danish SprogTeknologisk Ordbase* (STO) comprises a comprehensive syntactic layer which is assumed to be suitable for enhancement with semantic information. The theoretical background for the current approach is the consensus on obvious relationships between a syntactic behaviour and a particular sense of lemmas, as a surface complementation structure reflects the underlying semantic argument structure. The idea is to test the feasibility of deriving semantic information systematically from the syntactic structures encoded in syntactic patterns.

In the pilot project, a sub-set of trivalent verbs that share syntactic constructions are extracted from STO; the material consists of 216 verbs subcategorising for a direct object and a prepositional object covered by eight syntactic patterns. The examination takes a syntactically based grouping of these verbs as its starting point and focuses on defining lexical classes in terms of shared prevalent meaning components. These components form the basis of the semantic label assignment to the particular groups. The material provides 20 basic semantic groups, such asforce, urge, judge, consider, remove, cheat, etc. that can be refined into sub-groups along further semantic features or generalized into classes-e.g. communicate-persuade, cause-change-of, according to different degrees of granularity required. The present classifications of the verbs are also examined in relation to Levin's English verb classes (1993).

Our findings suggest that it is feasible-though within recognized limits-to exploit systematically the formalised syntactic descriptions in meaning group prediction.

An Author's Dictionary: The Case of Karel Capek

Čermák, Frantisek

ewe ...

1. Computational Lexicography and Lexicology

Short Paper

After a brief reminder of the long tradition of manually-based author's dictionaries, the possibility of a dictionary based on a full corpus and verified in a number of aspects against a large corpus has remerged. Specifically, the plan of Karel Capek's dictionary and its realisation is being discussed and its final shape shown, having a number of new, hitherto unused features. The Dictionary, being in fact split into four separate ones, is accompanied by the full Capek's corpus on a CD where a lot of additional information can be found.

de Loupy, Claude; Gonçalves, Sandra

1. Computational Lexicography and Lexicology

Full Paper

Morphosyntactic lexica are a very important resource for natural language processing. Many exist; some are freely available for research. But many organisms still produce lexica, even for languages with available resources. In this paper, we present some techniques that can be leveraged to produce lexica more efficiently. Firstly, the format of the lexicon is important. We use a very simple format based on the association of a lemma and a flexion rule, avoiding dozens of entries for a single lemma. Secondly, the linguist must describe some basic elements: the tag list, the tool words and the flexion rules. Thirdly, a specific guesser makes the completion of the lexicon easier. We describe two ways of adding entries to the lexicon using a guesser which associates a lemma and a flexion rule to a word, or a flexion rule to a lemma.

Bottom-up Editing and More: The E-forum of The English-Chinese DictionaryDing, Jun

1. Computational Lexicography and Lexicology

Student Paper

"Computer assistance may enable the lexicographer to prepare and revise dictionaries more quickly"-Barbara Kipfer's prediction made 20 years ago has already become a reality in our age of advanced information technology. Yet how much more quickly can the revision of dictionaries be carried out today? The envelope is now being pushed by the editors of *The English-Chinese Dictionary* (Unabridged) (ECD) through *bottom-up editing*, a new form of online lexicography. Following the launch of the second revised edition of ECD (April 2007), an electronic forum was introduced, linked to the website of Shanghai Yiwen Press, the publisher of the dictionary. For the time being, this e-ECD-forum is attracting more and more of its users to take part in *bottom-up editing*, i.e., pointing out errors and other problems detected in the dictionary directly to its editors through the Net. Three editors including the editor-in-chief participate in the e-forum discussion on a daily basis. Once the problems identified by the users are checked and properly edited by the editors, they will be listed in the e-newsletter linked to the e-forum. This paper will first explain the functioning of the e-ECD-forum and how such direct interaction between users and editors of ECD proves rewarding to both parties, and secondly, illustrate the mistakes and deficiencies published on the e-forum. Lastly it will explore the potential benefits and problems of online collaborative lexicography in the near future.

An Electronic Lexicon for Turkish Idiomatic Compounds Headed by Verbs

Eyigoz, Elif

1. Computational Lexicography and Lexicology

Poster

Turkish is a very creative language in terms of idiomatic compounds headed by verbs. Although traditional dictionaries include such compounds, syntactic and morphological properties of compounds are left unrepresented. Moreover, although essential elements of idiomatic compounds can be represented in subcategorization frames that refer to the argument positions of the verbs, it has been observed that subcategorization frames are impractical and even inadequate for representing the argument structure of idiomatic compounds headed by verbs in Turkish. This paper presents a design for representing properties of Turkish idiomatic compounds in a machine readable dictionary, which has been showcased in a sample dictionary for 322 idiomatic compounds.

Mordebe Admin-A Lexical Management System

Ferreira, José Pedro; Barbosa, Sílvia; Janssen, Maarten

1. Computational Lexicography and Lexicology

Software Demo

The *Portal da Língua Portuguesa* is a website containing information about the Portuguese language oriented towards the general public. The largest part of the information on the *Portal* is lexical information concerning formal characteristics of words, such as orthography, derivations, loanwords and gentiles. The lexical information comes from a lexical database called MorDebe-or more precisely, a network of lexical databases called the Open Source Lexical Information Network (OSLIN). This abstract shows the general set-up of and major functions of MorDebe Admin, which is the lexicon management system for OSLIN. MorDebe Admin provides an easy and secure way of updating and

editing the content of the different databases of OSLIN. Furthermore, much of the data on the Portal are organised as mini-dictionaries and MorDebe Admin provides an integrated collection of tools dedicated to the maintenance of these mini-dictionaries, as well as a built-in neologism tracking system. The software demonstration will illustrate these functions from a user perspective, and how easy ir is to maintain the data behind the Portal.

Lexicon Creator: A Tool for Building Lexicons for Proofing Tools and Search Technologies Fontenelle, Thierry; Cipollone, Nick; Daniels, Mike; Johnson, Ian

and the second s

1. Computational Lexicography and Lexicology

Full Paper

In this paper, we describe Lexicon Creator, a tool designed to help developers produce lexical data for its use in a variety of linguistic applications such as spell-checkers, word-breakers, thesauri, etc. The tool enables developers to work on existing wordlists derived either directly from corpora or from previously created wordlist data. The key feature of the tool is that it enables linguists to rapidly create the morphological rules that are necessary to generate all the inflected forms of a given item. In many languages, a given word may have many forms, each distinguished by different endings attached to the stem of the word. A language like English is rather simple, morphologically-the verb walk only has the following forms: walk, walks, walked, walking, while other languages may have a number of different forms for a word. Yet, it is essential to create lexicons that can recognize and generate all the inflected forms of a given word, especially for applications such as spell-checkers-where overgeneration should be avoided, thesauri, grammar checkers, morphological analyzers/generators, speech recognition, and handwriting recognizers. It would be extremely time-consuming to code each of these forms individually, so it is necessary to develop this data more efficiently. Lexicon Creator allows linguists to classify these variations of the same word into templates, or morphological classes, which allow the automatic generation of all valid forms of a word. Once the templates describing the aforementioned variations have been defined, the data-coding task consists of assigning an input word to the correct template and checking that the forms generated automatically are valid. The article will also focus on the additional types of linguistic information which can be attached to words, depending on the intended application that will use the resulting full-form lexicon.

Generation of Word Profiles on the Basis of a Large and Balanced German Corpus

Geyken, Alexander; Didakowski, Jörg; Siebert, Alexander

1. Computational Lexicography and Lexicology

Full Paper

In this paper we present the DWDS word profile system, a unified approach to the extraction of collocations for German, based entirely on finite state transducers. The system is intended as an additional informational source for the DWDS web-platform (www.dwds.de). The DWDS website-with 2.5 million page impressions per month-is a widely used internet platform that provides a wordinformation system based on a large monolingual German dictionary and the DWDS-Kerncorpus, a balanced corpus of German texts of the 20th century. The DWDS word profile consists of two parts: a language-specific part-which consists of a complete German morphology and an efficient syntax parser for German, and a language-independent part comprised of a database management system for collocations and a corpus query engine, together with a web interface. We have applied the DWDS word profile to a balanced German corpus of the 20th century and subsequently present some technicalities. Another experiment using the DWDS word profile in conjunction with a tabloid newspaper shows that there may be significant differences between corpora, underlining the importance of the corpus choice for language learning as well as for the construction of lexical resources. Future work will focus on language learning; in particular, we will use a simplified tag set and a more systematic description of the word profile differences between corpora. We also plan to create word profiles for the DWDS-extended corpus, a 2 billion token corpus.

El Dicionario de dicionarios do galego medieval

González Seoane, Ernesto; Álvarez de la Granja, María; Boullón Agrelo, Ana Isabel; Rodríguez Parada, Raquel; Rodríguez Suárez, María; Suárez Vázquez, Damián

1. Computational Lexicography and Lexicology

Software Demo

This paper wishes to present some of the most notable tools of the Dicionario de dicionarios do galego

medieval. This recently-published work is an electronic multidictionary that includes fourteen glossaries and vocabularies born out of Galician or Galician-Portuguese texts or textual collections from the Middle Ages.

Shimmering Lexical Sets

Hanks, Patrick; Ježek, Elisabetta



1. Computational Lexicography and Lexicology

Full Paper

For natural language processing and other applications, it has long seemed desirable to group words together according to their essential semantic type-[[Human]], [[Animate]], [[Artefact]], [[Physical Object]], [[Event]], etc.-and to arrange them into a hierarchy. Vast lexical and conceptual ontologies such as WordNet and BSO have been built on this foundation. Examples such as fire a [[Human]] (=dismiss from employment vs. fire a [[Weapon]] (=cause to discharge a projectile) have led to the expectation that semantic types such as [[Weapon]] and [[Human]] can be used systematically for word sense disambiguation. Unfortunately, this expectation is often unwarranted. For example, one attends an [[Event]]-a meeting, a lecture, a funeral, a coronation, etc., but there are many events-e.g. a thunderstorm, a suicide-that people do not attend, while some of the things that people do attende.g. a school, a church, a clinic-are not [[Event]]s, but rather [[Location]]s where specific events take place. The sense of attend is much the same in all these examples, unaffected by differences in the semantic type of the direct object. Nevertheless, the pattern [[Human]] attend [[Event]] is well established and intuitively canonical.

The CPA (Corpus Pattern Analysis) project at Masaryk University, Brno, provides two steps for dealing with this kind of inconvenient linguistic phenomenon:

- 1. Non-canonical lexical items are coerced into "honorary" membership of a lexical set in particular contexts, e.g. school, church, clinic are coerced into membership of the [[Event]] set in the context of attend, but not, for example, in the context of arrange.
- 2. The ontology is not a rigid yes/no structure, but a statistically based structure of shimmering lexical sets.

Thus, each canonical member of a lexical set is recorded with statistical contextual information, like this: [[Event]]: ... meeting. Thus, the semantic ontology is a shimmering hierarchy populated with words which come in and drop out according to context, and whose relative frequency in those contexts is measured. A shimmering ontology of this kind preserves, albeit in a weakened form, the predictive benefits of hierarchical conceptual organization, while maintaining the empirical validity of natural-language description.

The Use of Context Vectors for Word Sense Disambiguation within the ELDIT Dictionary Ignatova, Kateryna; Abel, Andrea



1. Computational Lexicography and Lexicology

Short Paper

The aim of this paper is to tackle the problem of Word Sense Disambiguation (WSD) within the ELDIT system. ELDIT (Elektronisches Lernwörterbuch Deutsch-Italienisch) is an online dictionary of German and Italian, as well as a web-based language-learning system targeted at language learners at elementary and intermediate level. In ELDIT, each word is linked with the corresponding dictionary entry with a list of senses. Nevertheless, selecting the suitable sense of a polysemous word as well as choosing the appropriate homonym in the lookup process is not a trivial task, especially for language learners at elementary level. Therefore, it is desirable to make the dictionary work easier by automatically selecting the right sense of a word in a given context, which is a Word Sense Disambiguation task. While WSD has been studied intensively in fields such as Information Retrieval (IR), Machine Translation (MT), Question Answering (QA), etc., we present a novel setting, in which WSD is performed within an integrated dictionary system. For performing WSD, we first utilize different kinds of knowledge contained in the ELDIT dictionary, namely part of speech information, morphological knowledge, collocation patterns, and various example sentences as the basis for the context vectors technique. Besides, when the ELDIT dictionary does not provide sufficient data for building a context vector for a word, we fall back upon the vast Internet knowledge. By combining all these sources of information, the implemented module is able to automatically choose the most appropriate meaning of a word in a particular context. It achieves an average precision of 96% for disambiguating Italian and 93% for disambiguating German homonyms. The results for polysemous words greatly depend on how distinct the senses are and how many senses a word has. The

evaluation, however, has shown that the approach we apply always outperforms the baseline systemnamely, a simplified Lesk algorithm-and gives quite promising results. In addition to that, we show that the data obtained during our work can be re-used in a number of interesting tasks to serve the further improvement of the ELDIT system.

Meaningless Dictionaries

Janssen, Maarten



1. Computational Lexicography and Lexicology

Full Paper

The creation of word meaning is one of the most time consuming parts of creating a dictionary. Although it is commonly thought that providing definitions for words is the primary function of dictionaries, it is not the most frequent one. Most dictionaries are used for looking up much more basic information, such as to see whether a word exists or to see whether it is spelled correctly. Dictionaries are relatively good at providing complete definitions for individual words but are not necessarily well equipped for more basic tasks. For many of these smaller tasks, users would be better off using smaller databases-or dictionaries-that focus only on the information the user is looking for rather than searching in a general language dictionary. A dictionary that leaves out most of the details traditionally included in the lexical entry not only makes it easier for the user to find the information he is looking for but also allows the lexicographer to put more focus on the relevant data. It does this by focussing on a single type of information; it becomes more feasible to treat it completely, consistently and coherently for the entire lexicon. The Open Source Lexical Information Network-henceforth OSLIN-is an attempt to create such singe-task lexical resources. This paper explains both the advantages and problems of such an approach.

Software Demonstration: The TshwaneLex Electronic Dictionary System

Joffe, David; MacLeod, Malcolm; de Schryver, Gilles-Maurice

1. Computational Lexicography and Lexicology

Software Demo

In this presentation, use of the *TshwaneLex Electronic Dictionary* software module will be demonstrated. This module provides a complete, customisable solution for the publication of a dictionary electronically-as a CD-ROM, or for sale or download on the Internet, or both. The "base package" provides all the functionality of a modern, user-friendly Electronic Dictionary, and can be fully customised for the desired 'look and feel', branding, dictionary content and language(s) of a publisher's product. This provides a highly cost-effective solution for creating a professional Electronic Dictionary product, obviating the need for the kind of expensive custom-developed solutions that have traditionally been required. The system can be used for the immediate publication of dictionaries already in TshwaneLex, or in a comparable structured format-such as XML.

GDEX: Automatically Finding Good Dictionary Examples in a Corpus

Kilgarriff, Adam; Husak, Milos; McAdam, Katy; Rundell, Michael; Rychlý, Pavel

1. Computational Lexicography and Lexicology

Full Paper

Users appreciate examples. If a dictionary entry includes contextualized examples of the different senses a word may have, then the user generally gets what they want in a quick and straightforward way. Thus, there are grounds for including lots of examples and contexts. Producing good examples, however, can be labour-intensive, thus, expensive. We automatically found good candidate sentences in a corpus, with which lexicographers could work. The technology used to add examples to an online version of a leading dictionary: we describe and evaluate the project. We consider a range of other ways in which the finding of good examples can bridge the gap between corpuses, dictionaries, and language learning.

Finding the Words Which are Most X

Kilgarriff, Adam; Rychlý, Pavel

1. Computational Lexicography and Lexicology

Software Demo

Which English words are most distinctive of American English? Which Spanish verbs have a strong

tendency to occur in the gerund? Which English nouns are most often used in the plural? All these questions can be answered in quite a straightforward with a suitable corpus with appropriate markup. The task usually takes a moderate amount of programming. We present a tool which means that it is easy to produce lists of this kind-and many others-which needs no further programming. The work takes place in the framework of a leading corpus query tool.

Corpus as a Means for Study of Lexical Usage Changes

Křen, Michal; Hlaváčová, Jaroslava



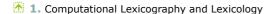
Full Paper

The paper presents a corpus-based method for obtaining ranked wordlists that can characterise lexical usage changes. The method is evaluated on two 100-million representatively balanced corpora of contemporary written Czech that cover two consecutive time periods. Despite similar overall design of the corpora, lexical frequencies have to be first normalised in order to achieve comparability. Furthermore, dispersion information is used to reduce the number of domain-specific items, as their frequencies highly depend on inclusion of particular texts into the corpus. Statistical significance measures are finally used for evaluation of frequency differences between individual items in both corpora.

It is demonstrated that the method ranks the resulting wordlists appropriately and several limitations of the approach are also discussed. Influence of corpora composition cannot be completely obliterated and comparability of the corpora is shown to play a key role. Therefore, although highly-ranked items are often found to be related to changes of language usage, their relevance should be cautiously interpreted. In addition to several general language words, the real examples of lexical variation are found to be limited mostly to temporary topics of public discourse or items reflecting recent technological development, thus sketching an overall picture of lifestyle changes.

Non-heads of Compounds as Valency Bearers: Extraction from Corpora, Classification and Implication for Dictionaries

Lapshinova-Koltunski, Ekaterina



Short Paper

This paper describes an approach to the classification of nominal compounds based on their subcategorisation. German compound noun predicates, such as *Grundproblem*, *Beweislast* and *Schlussfolgerung* subcategorizing for a subordinate clause are semi-automatically extracted from text corpora and classified according to which of their components, the head or the non-head, is the valency bearer. In over 40% of cases the subcategorisation of compounds is not determined by their heads. This kind of information should be included in subcategorisation lexicons as well as dictionaries for human users. We show that our semi-automatic approach can be applied in natural language processing, especially in lexicon and dictionary creation.

The Lexicographic Portal of the IDS: Connecting Heterogeneous Lexicographic Resources by a Consistent Concept of Data Modelling

Müller-Spitzer, Carolin

1. Computational Lexicography and Lexicology

Software Demo

The *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online Vocabulary Information System German) of the *Institut für Deutsche Sprache* (IDS; German Language Institute) in Mannheim is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the IDS. It is an explicit goal of OWID, not to present a random collection of unrelated reference works but to build a network of actually related lexicographic products. Hence, the core of the project is the design of an innovative concept of data modelling and structuring. The goal of this granular data modelling is to allow flexible access of each individual lexicographic resource as well as access across diverse dictionary resources. At the same time, fine-grained interconnectedness of all resources should be made possible. Every lexicographic resource within OWID-*elexiko*, *Neologismenwörterbuch*, *Wortverbindungen online*, *Schulddiskurs im ersten Nachkriegsjahrzehnt*-accomplishes this requirement with regard to data modelling and structuring. The paper explains the underlying consistent concept of

the data modelling for the overall heterogeneous lexicographical resources. Also it is shown, how the modelling potential has been converting into the Internet presence of OWID.

Multilingual Open Domain Key-word Extractor Proto-type

Panunzi, Alessandro; Fabbri, Marco; Moneglia, Massimo

1. Computational Lexicography and Lexicology

Software Demo

Automatic Keyword extraction is now a mature language technology. It enables the annotation of large amount of documents for content-gathering, indexing, searching and for its identification, in general. The reliability of results when processing documents in a multilingual environment, however, is still a challenge, particularly when documents are not limited to one specific semantic domain. The use of multi-term descriptors seems to be a good mean to identify the content. According to our previous evaluations (Panunzi et al. 2006a, 2006b), the availability of multi-term keywords increases the performance with respect to mono-term keywords of 100% relative factor. The LABLITA tool presented in this demo works now in a multilingual environment, as well. The demo calculates on the fly the number of mono-term and multiword keywords of parallel documents in English, Italian, German, French and Spanish, and will allow the audience to judge: a) the enhancement bared by multiword keywords for the identification of content; and b) the comparability of performance obtained by the tool processing different languages.

Refining and Exploiting the Structural Markup of the eWDG

Schmidt, Thomas; Geyken, Alexander; Storrer, Angelika

1. Computational Lexicography and Lexicology

Full Paper

In this paper, we describe a semi-automated approach to refine the dictionary-entry structure of the digital version of the Wörterbuch der deutschen Gegenwartssprache (WDG, en.: Dictionary of Presentday German), a dictionary compiled and published between 1952 and 1977 by the Deutsche Akademie der Wissenschaften that comprises six volumes with over 4,500 pages containing more than 120,000 headwords. We discuss the benefits of such a refinement in the context of the dictionary project Digitales Wörterbuch der deutschen Sprache (DWDS, en.: Digital Dictionary of the German language). In the current phase of the DWDS project, we aim to integrate multiple dictionary and corpus resources in German language into a digital lexical system (DLS). In this context, we plan to expand the current DWDS interface with several special purpose components, which are adaptive in the sense that they offer specialized data views and search mechanisms for different dictionary functions-e.g. text comprehension, text production-and different user groups-e.g. journalists, translators, linguistic researchers, computational linguists. One prerequisite for generating such data views is the selective access to the lexical items in the article structure of the dictionaries which are the object of study. For this purpose, the representation of the eWDG has to be refined. The focus of this paper is on the semiautomated approach used to transform eWDG into a refined version in which the main structural units can be explicitly accessed. We will show how this refinement opens new and flexible ways of visualizing and querying the lexicographic content of the refined version in the context of the DLS project.

An Anglo-Saxon Dictionary and a Morphological Analyzer of Old English

Tichy, Ondrej; Cermak, Jan

🐧 1. Computational Lexicography and Lexicology

Short Paper

The main stages in the project of the digitization of the *Anglo-Saxon Dictionary* by J. Bosworth and T.N. Toller are described and the value of the resulting data is considered. The paper suggests that the dictionary data need to be structurally tagged if we are to further benefit from the project beyond the current dictionary application. It is also noted that the re-tagging process can be partially automatized, but that it will have its complications due to the ambiguity of typographical tagging currently included in the data. An outline of the development of an Old English morphological analyzer, now in its early stages, is offered using the valuable digitized data of the Dictionary and drawing on a model of a functional Czech morphological analyzer. Envisaged problems, such as the building of stem- and affix-lexicons, Old English vowel variation and stem-final variation, are discussed and several solutions are proposed. The paper also proposes and accounts for some divergence from the model of the Czech analyzer reflecting differences between Czech and Old English morphology and slight differences in the

final uses of the Modern Czech and Old English analyzers. Finally, the analyzer's future use, both as a part of the dictionary and as a stand-alone tool for parsing the corpora, for connecting the lexicon entries with text, etc., is suggested and some possibilities of future improvements, e.g. a word-formation or a syntactic analyzer, are indicated.

El programa de ejemplificación en los diccionarios didácticos

Bargalló Escrivá, María



Poster

In taking into account the value users place on exemplification, we intend, in this work, to show how the information contained in the microstructure is interlinked with that offered in the examples provided in Spanish didactic dictionaries. Additionally, in the terminology used by Rey-Debove (2005), we will observe the interrelation between the information programme and the exemplification programme.

In order to centre our discussion upon these questions, we will try to show to what extent redundant grammatical information is used between the descriptive and illustrative parts, or the complementarity between both these parts. We will analyze the various ways in which this relationship is manifested in order to posit generalizations about the exemplification programme in the dictionaries in question.

To conclude, we will show that, from our point of view, little attention is given to the project as a whole, since a study of some of the questions linked to grammatical information reveals that there are no unified criteria about how the relationship between the information programme and the exemplification programme should develop.

Sobre las construcciones pronominales y su tratamiento en algunos diccionarios monolingües de cuatro lenguas románicas

Battaner, Paz; Renau, Irene

2. The Dictionary-Making Process

Full Paper

Some non-native student errors in Spanish show the difficulty of pronominal constructions in Spanish language, which is an aspect that has been analysed under all possible approaches in the grammatical bibliography, but has received little attention from a lexicographical point of view. Our aim in this paper is to propose proper treatment of this issue in the Spanish Learner's Dictionary for Foreign Speakers (DAELE). We review the advances in grammatical analyses of these constructions for Spanish, and later we observe the treatment they have received in several dictionaries of other Romance languages, to decide which parameters we will take into account for verb entries presenting these constructions. Basing ourselves on the grammatical studies, we establish a classification of ten types of pronominal uses, grouped according to if they are: A) uses deriving from the grammar, in which the pronoun represents an argument of the verb; B) constructions in some languages, such as Spanish and Catalan, that admit what appear to be reflexive pronouns which which in fact do not represent arguments of the verb; and C) alternations with or without the pronoun that may or may not display a change in meaning. With this classification in mind, and although the description for Spanish cannot be generalized to all Romance languages, we review some general monolingual and new learner's dictionaries for foreign speakers. For the selection of verbs consulted, we take the examples from French provided by Fontenelle (2004). The treatment of the pronominal uses is analyzed for their appearance in headwords, in a specific sense or subsense, in the examples, and in the observations or remarks. There is some variation across dictionaries, and also within a single dictionary, and that some uses are rarely or irregularly included. Our initial conclusion is that in the DAELE we should adopt some solutions that not do not introduce more variation and should strive to simplify the analysis.

La distribució de la informació contextual en els elements estructurals d'un article de diccionari: col·locacions, restriccions lèxiques i definició

Feliu, Judit; Soler, Joan

2. The Dictionary-Making Process

Short Paper

The main goal of this paper is to discuss the need of improving the *encoding dimension* of general monolingual dictionaries by considering the treatment of non-phraseological lexical combinations-

particularly collocations. This will be attained through an analysis of the notion of collocation from a lexicographic point of view taking into account both empirical and theoretical approaches to the general problem of lexical combinations. Moreover, lexical information extracted from corpus must be analysed and included in general monolingual dictionaries bearing in mind that it can be distributed in different elements of an entry-definition, example, etc.- depending on each dictionary structure. Thus, authors will shed some light on the lexicographic task in order to determine whether the co-ocurrence of two lexical items must be retrieved and, if so, how and where this lexical and semantic information should be organised. In this sense, it will be demonstrated that the definition and the example fields are not enough to retain and reflect the real use of collocations extracted from corpus. Some guidelines will be provided in order to help the lexicographer in the dictionary-making task concerning the distribution of the extracted information in the collocation field and also in the semantic-restriction pattern and its corresponding definition.

The Greek High School Dictionary: Description and issues

Gavrilidou, Maria; Giouli, Voula; Labropoulou, Penny



Short Paper

This paper reports on the compilation of a monolingual Greek pedagogical dictionary targeted at young native language learners, namely secondary education students, aged between 12 and 15. The dictionary, which is in printed form, has been designed to be used in the classroom as a supporting tool for language learning, but also as reference work tailored to meet students' needs for language understanding and production both at school and in everyday activities outside school. To this end, considerations on user-friendliness have been accounted for, and the design and implementation of the dictionary content have built primarily on the needs and requirements of schoolchildren pertaining to the specific age group. The dictionary comprises 15,000 lemmas covering general language vocabulary along with terms belonging to subjects taught at the specific level of education. Information that is central to the pedagogical targets of language learning has been encoded for each lemma, i.e., part of speech, morphology-difficult inflectional forms, domain, register, definitions, usage examples, etc. Finally, useful comments focus on interesting aspects of certain words' semantics, usage, register etc. The central feature of the dictionary is the headword organization which employs systematically word formation criteria: derivatives by suffixation are organized in word-families, while prefixes are included in the dictionary as independent headwords accompanied with lists of derivatives or compounds on the basis of derivational and semantic criteria. The paper presents the framework of the project and its specifications, discusses the main methodological principles that underlie its construction and elaborates on the dictionary description, the main problems faced and the solutions adopted in the process of its compilation.

Desafíos de la definición

Gutiérrez Cuadrado, Juan



Short Paper

This paper reviews the definition of several words in current Spanish dictionaries: *mono* 'monkey', *simio* 'ape', *primate* 'primate', *orangután* 'orang-utan', *gibón* 'gibbon', *chimpancé* 'chimpanzee', *gorila* 'gorilla'. The analysis of the definitions attempts to make it clear that there are some serious inconsistencies in the encyclopaedic information of the dictionaries. It will also be shown that the substitution principle of the definiendum for the definiens is problematic in the sense that its conditions of usage are not well defined. Thus, we find that Spanish dictionaries must incorporate the encyclopaedic information in a specific way. On the other hand, our criticism reflects that the substitution principle should be reformulated. The analyses of the questions being examined here reveals that a theoretical and methodological debate is needed in Spanish lexicography. This paper attempts to demonstrate that some of the objections to the definitions of current Spanish dictionaries are due to lack of criticism in the Peninsula regarding the problems related to the encyclopaedic information of the definitions. It would also be advisable to review the questions related to the substitution principle.

The Funny Mirror of Language: The Process of Reversing the English-Slovenian Dictionary to Build the

Framework for Compiling the New Slovenian-English Dictionary

Krek, Simon; Sorli, Mojca; Kocjančič, Polonca

2. The Dictionary-Making Process

Short Paper

The article describes the process of reversing the English-Slovenian dictionary database in XML format to create the framework for compiling the Slovenian-English dictionary. The aim was to maximize the abundance of information in an extensive dictionary database with a complex and detailed structure. The process involved lemmatization and POS-tagging of both source and target languages, construction of routines to form the preliminary list of possible headwords and their translation equivalents, as well as routines which enabled the grouping of numerous dictionary examples available in the original dictionary under the appropriate translation equivalent. The result is the reversed dictionary database in XML format with the DTD and XSL file to control the layout for viewing the database in Internet browsers or other XML-aware-dictionary-editors. The article presents the process of reversing the dictionary and the features of the final database. It also reflects on the linguistic issues concerning the fact that the database represents only the mirror image of the English-Slovenian contrastive relation and argues that the contrastively undistorted lexical information from a monolingual Slovenian reference corpus has to be taken into consideration when compiling the new Slovenian-English dictionary.

Making a Thesaurus for Learners of English

Lea, Diana

2. The Dictionary-Making Process

Short Paper

This paper explains the principles and methodology behind the selection and presentation of synonyms in the Oxford Learner's Thesaurus-a dictionary of synonyms (April 2008). The needs of learners when consulting a thesaurus are different from those of native speakers: so different, in fact, that they need a completely different kind of thesaurus to consult. Native speakers have a large bank of language stored in their brains; the thesaurus, for them, is simply a means of accessing this information. It reminds them of words that they already know but cannot bring to mind. For language learners, the traditional thesaurus contains far too many words, and not nearly enough information about any of them. They need a thesaurus that will not only enable them to access information, but will also teach them things they did not know before. The first task was to decide which words to include. A conceptual framework was established, dividing the language into areas of thought and experience. Words under each heading were sorted into groups of near-synonyms. A system of frequency counting was used to order the synonyms and eliminate the less frequent. The resulting entry list was checked against a core vocabulary for learners. Then the entries had to be written. Here, the list of synonymsforming pretty much a complete entry in a traditional thesaurus-was just the beginning. Each synonym was defined and exemplified. Careful thought was given to register, usage and collocation. Notes contrast the meaning and usage of pairs or groups of words that are particularly hard to tease apart. The aim of the learner's thesaurus is to expand the learner's word bank. It both adds words to the bank, words that the learner did not even know before, and helps learners choose more effectively between words that they have met before, where their knowledge of the exact meaning and usage of the words was previously incomplete.

Structure de la définition lexicographique dans un dictionnaire d'apprentissage explicatif et combinatoire Milićević, Jasmina

2. The Dictionary-Making Process

Short Paper

The paper focuses on the construction of lexicographic definitions for an electronic dictionary targeting intermediate-to-advanced learners of French as a second language. It proposes a learner-friendly adaptation of definition formats developed for a theoretical lexicon of a particular type-*Explanatory-Combinatorial Dictionary* [= ECD] of Contemporary French. It is argued and, hopefully, demonstrated that it is possible to construct lexicographic definitions that are both theoretically sound and palatable for language learners. The work on the adaptation of existing definition formats has shed light on these definitions themselves-which in some cases needed to be modified-thus demonstrating, once again, the interdependence of applied and fundamental research. After presenting the basic structure of an ECD-style definition, the paper details the types of modifications needed to make it learner-friendly. These range from 'superficial' modifications, aimed at a better readability of the definition-

typography, indentation, color, etc., to substantial ones-simplification of the vocabulary and syntax of the defining language, omission of the definition components deemed non-essential in a learner's dictionary, changing the word-sense division within a polysemic word in case the level of detail seems too high for our purposes, etc. The proposed approach is illustrated with the example of evaluation verbs, such as *approuver* '(to) approve', *désapprouver* '(to) disapprove', *blamer* '(to) blame', *critiquer* '(to) criticize', etc., for which a definition template and definitions themselves are given.

Frames and Semagrams. Meaning Description in the General Dutch Dictionary Moerdijk, Fons

2. The Dictionary-Making Process

Full Paper

This paper discusses the semagram, an innovation in the way of describing meaning in lexicography, as used in the *Algemeen Nederlands Woordenboek* (General Dutch Dictionary). A semagram is the representation of knowledge associated with a word in a frame of *slots* and *fillers*. *Slots* are conceptual structure elements which characterise the properties and relations of the semantic class of a word-e.g. colour, smell, taste, composition, components, preparation for the class of beverages. The abstract meaning frame for such a semantic class is called *type template*. After a motivation for the use of frames in lexicography we reveal how semantic classes are determined and how type templates are composed. We illustrate this with the type template of the animal names and show how the semagram of *cow* is based upon it. We conclude by summing up the main advantages of the use of semagrams.

A Systematic Approach to the Selection of Neologisms for Inclusion in a Large Monolingual Dictionary O'Donovan, Ruth; O'Neill, Mary

2. The Dictionary-Making Process

Full Paper

For each new edition of *The Chambers Dictionary*, around 1,000 new words are selected by Chambers' lexicographers for inclusion. In preparing the latest edition, we seized the opportunity to use new corpus and database technology to improve neologism detection and selection. Our resources included the large, recently built Chambers Harrap International Corpus (CHIC), our automated word-tracking system, the databases developed for our new words monitoring programmes and a new tool for ranking words by corpus frequency. We report on the results of our work in this area: a systematic approach to neologism detection and investigation that complements the expertise of lexicographers.

Lexicographic Treatment of Italian Phrasal Verbs: A Corpus-based Approach Onesti, Cristina

2. The Dictionary-Making Process

Poster

Italian phrasal verbs-or *verbi sintagmatici*-have seen a growth of interest in the Italian linguistic panorama. However, a more systematic analysis is needed to clarify the theoretical status of these verb-particle constructions and to improve their lexicographic treatment, which is still inconsistent, as shown by an overall comparison between the major Italian monolingual dictionaries. Difficulties are related with both an unclear semantic classification of them and the lack of frequency and productivity data about their formation. Following the classification of Masini 2005 in terms of intensification, direction, metaphoric and actional meaning, the present paper carries out a case study aiming at frequency data in phrasal verbs withvia, in particular about the presence of an Aktionsart contribution of the particle. The meaning of accomplishment seems to be clearly traceable in the newsgroup messages analyzed (corpus NUNC-It), although very few dictionaries record it. Further corpus data should offer schemes of regularity in the usage of other verb-particle constructions making their lexicographic treatment more effective.

Il Dizionario Garzanti nel quadro della lessicografia italiana contemporanea

Patota, Giuseppe

2. The Dictionary-Making Process

Full Paper

Over the last fifteen years, Italian lexicography has achieved significant results in the field of historical as well as general dictionaries, thanks to the publication of outstanding works which provide the users

with a wide range of information, both in paper and digital form, namely: phonemic transcriptions, etymological indications, the dating of the entries with the original occurrence, use frequency, phraseology, synonyms and antonyms, polyrematic units, grammatical notes. As they are specifically geared to provide surveys of written and spoken Italian based on real evidence, these works have come to be crucial methodological tools for knowledge not merely linguistic. The *Dizionario Italiano Garzanti* is to be placed within the framework of this 'new' lexicography. The dictionary, which from the beginning has featured clarity, accessibility, and comprehensiveness, has profoundly changed its scope over time, and its 2008 version can be regarded as the achievement of a long-standing commitment to language evolution. My paper will address the most significant transformations of the Dizionario by accounting for its developments in the context of the most recent history of Italian lexicography.

Lemmatisierungspraxis und -problematik im Autorenwörterbuch am Beispiel des *Goethe-Wörterbuchs* Schares, Thomas; Schlaps, Christiane



Short Paper

The macrostructure of an author's dictionary as determined by (the rules of) lemmatization so far has attracted little attention in practical and theoretical lexicography, with most publications on the topic covering small-sized dictionaries that represent only specific segments of a writer's works and vocabulary. The Goethe Dictionary (GWb), in contrast, endeavors to treat the complete vocabulary used by Johann Wolfgang Goethe in his prolific literary, scientific, philosophical, historiographic etc. writings as well as in his letters and, to some extent, his conversations. The word list contains over 90.000 headwords, thus making the project the largest dictionary on an author's idiolect worldwide. The form and placement of lemmas in the GWb is determined by a set of rules that over the decades of work on this dictionary have been collected in a style manual but which are far from comprehensive or even static. Examples from the published parts of the GWb will demonstrate a number of decisions that typically arise in author's dictionaries and the way they are solved in the GWb, including questions of an author's idiolectal aberration from the orthographic norm, the frequency of hapax legomena, and morphological idiosyncrasies that in the GWb led to specialized sublemmas. In our paper, we would like to emphasize the need for research, both practical and theoretical, into the special problems of lemma presentation in author-centred lexicography and its methodological foundation as we believe that to show the full range of an individual's lexicon will not only contribute to the field of metalexicography in particular but will also help to gain valuable insights into the lexicological, semantic, grammatical, and, incidentally, historical dimensions of language in general.

Von der Markierung zur Beschreibung: Besonderheiten des (Wort-) Gebrauchs in elexiko Schnörch, Ulrich



Full Paper

Elexiko is a lexicological-lexicographic, corpus-guided German Internet reference work (cf. www.elexiko.de). Compared to printed dictionaries, in elexiko, restrictions on space disappear. Specific comments on the use of a word do not need to be given in traditional abbreviated forms, like the so-called field labels or usage. In this paper, I will show its advantages for the description of the particular pragmatic characteristics of a word: I will argue that traditional labelling such as formal, informal, institutional, etc. cannot account for the comprehensive pragmatic dimension of a word and that these are not transparent, particularly for non-native speakers of German. The main focus of the paper will be on an alternative approach to this dictionary information-as suggested by elexiko. I will demonstrate how narrative, descriptive and user friendly notes can be formulated for the explanation of the discursive contextual embedding or tendencies of evaluative use. I will outline how lexicographers can derive such information from language data in an underlying corpus which was designed and compiled for specific lexicographic purposes. Both, the theoretical-conceptual ideas and their lexicographic realisation in elexiko will be explained and illustrated with the help of relevant dictionary entries.

Requirements for the Design of Electronic Dictionaries and a Proposal for Their Formalisation Spohr, Dennis

2. The Dictionary-Making Process

Full Paper

We discuss recent analyses of the requirements for the design of electronic dictionaries, building primarily on the accounts by de Schryver (2003), Chiari (2006), Heid (2006) and Tarp (2008). These requirements suggest a richer formalization of dictionary models than is usually the case in traditional database and plain XML-based approaches, and we therefore argue in favour of a formalisation of these requirements in the framework of a strongly typed formalism. The discussion focusses on users' needs, needs of specific applications of Natural Language Processing, and multifunctionality-in the sense suggested by Gouws (2006) and Heid/Gouws (2006). We further point out the benefits of a richer formalization of dictionary models that goes beyond the traditional view on lexical resources, and strengthens our claim by providing evidence from related work on lexicon modelling in OWL DL (Burchardt et al., 2008).

Alphabetic Proportions in Estonian Monolingual and Bilingual Dictionaries

Veldi, Enn

2. The Dictionary-Making Process

Short Paper

The paper discusses alphabetic proportions in Estonian general monolingual and bilingual dictionaries with Estonian on the left-hand side. As no data about the Estonian alphabetic proportions were available, the alphabetic proportions were calculated on the basis of the corpus-based Frequency Dictionary of Standard Estonian (Kaalep and Muischnek 2002). The findings were then used to configure the Estonian ruler in TshwaneLex dictionary compilation software-alphabetic proportions for English, Afrikaans and several Afrikan languages as well as an excellent background to this problem can be found in De Schryver 2005. Subsequently, the established proportions were used as a yardstick for comparing three monolingual and six bilingual dictionaries. The six bilingual dictionaries included four general Estonian-English dictionaries-one of them not completed as of yet but revealing potential problems-and two school dictionaries-Estonian-German and Estonian-Russian. The findings show that while alphabetic proportions have generally been followed quite successfully, some Estonian dictionaries show a tendency to be skewed-some dictionaries become more thorough towards the end of the alphabet while others show the opposite trend. The problem is more challenging for those dictionary projects that require decades for completion and where the dictionary is published in fascicles-in the Estonian lexicographic practice one has in mind the Explanatory Dictionary of Standard Estonian, which started publication in 1988. There can also naturally be instances where single alphabetic stretches may reveal perceived overtreatment or undertreatment. The paper also argues whether the alphabetic proportions of certain letters can vary to some extent depending on the selection of words listed under them-e.g. inclusion of large numbers of foreign and learned words beginning in a, b, d, g, f in Estonian dictionaries can increase the proportions of these letters as they are less typical of native words. The paper ends with the firm conviction that in recent years it has become much easier to control the progress of a lexicographic project.

Dictionnaire de Néologismes du Portugais Brésilien (décennie de 90): conception et processus d'élaboration Alves, Ieda Maria

3. Reports on Lexicographical and Lexicological Projects

Poster

The Brazilian Portuguese Neologisms Dictionary (1990s), expected to be completed in early 2008, was created to present neologisms from contemporary Brazilian press from January 1993 to December 2000. The dictionary corpus has been taken from the Contemporary Brazilian Portuguese Neologism Database, a tool aimed at collecting and investigating neologisms from contemporary Brazilian pressnewspapers Folha de S. Paulo and O Globo and magazines IstoÉ and Veja-since January 1993. Samples were randomly taken from these sources: the newspaper O Globo, from the first Sunday of the month; the magazine IstoÉ, from the second week of the month; the newspaper Folha de S. Paulo, from the third Sunday of the month; and the magazine Veja, from the last week of the month. In the investigated period, 13 500 neologistic-lexical units were collected. The collected data shows that with 30%, the prefixed formations are the most frequent in the group of neologistic words. The other formations correspond to subordination composition (19%), borrowings (17%), syntagmatic formations (13%), suffixed formations (8%), coordination composition (5%), semantic neology (4%), blending (2%) and other formations (2%). These results are reflected in the dictionary's macrostructure, which will present about 3 000 entries according to the frequency proportion, e.g. the prefixed formations will correspond to 30% of the entries. Each article demonstrates necessarily: lexical units, grammatical references, definitions, context, context references, linguistic notes-types of

formation, attested years and eventually label, variant, acronym or abbreviated form, encyclopedic notes and attestation in dictionaries after 2000. For example: *jogador-chave* sm *Jogador que se destaca em uma equipe. O relatório que Matsunaga entregou ao técnico Akira Nishino aponta o meia-atacante Juninho como o do Brasil. (FSP, 21-07-96) Composição por subordinação*. Atestado (4) em 1994, 1996, 1998.

A Digital Dictionary of Catalan Derivational Affixes

3. Reports on Lexicographical and Lexicological Projects

Bernal, Elisenda; DeCesaris, Janet

Poster

This paper presents the digital dictionary of Catalan derivational affixes, awarded with the Laurence Urdang Award 2005. The aim of this dictionary is to provide a tool for lexicographers that will help them systematize the representation of the language's morphology in dictionaries, and provide an indepth description of Catalan derivational affixes in the form of a dictionary that should be of interest both to linguists and language professionals and those seeking a model for similar projects dealing with other languages.

Recopilación y estructuración del vocabulario de especialidad en el *Nuevo Diccionario Histórico del Español* (RAE)

Carriazo Ruiz, José Ramón; Gómez Martínez, Marta

3. Reports on Lexicographical and Lexicological Projects

Short Paper

The New Historical Dictionary of the Spanish Language is the latest work the Spanish Royal Academy has decided to undertake. In this lexical compilation, scientific and technical vocabulary will be taken into account and included, regardless of what other previous dictionaries, such as general language dictionaries, have done so far. For this purpose, a team of lexicographers in Cilengua (La Rioja) is studying the method to select, extract and tag the vocabulary used in specialized areas of knowledge. This presentation explains the different steps that will be followed in order to include the specialized terms and their history in the dictionary, such as the establishment of a representative corpus, the selection of terms or the introduction of subject matter labels.

Portal de léxico hispánico: una herramienta para el estudio del léxico

Clavería, Gloria; Prat, Marta; Torruella, Joan; Buenafuentes, Cristina; Freixas, Margarita; Julià, Carolina; Massanell, Mar; Muñoz, Laura; Varela, Sonia

3. Reports on Lexicographical and Lexicological Projects

Short Paper

The Portal de léxico hispánico (Hispanic vocabulary portal) is the result of the research project "Banco de datos diacrónico e hispánico: morfología léxica, sintaxis, etimología y documentación", carried out in the last few years thanks to funding by the Ministerio de Educación y Ciencia (Ref Nº: HUM2005-082149-C02-01) and the Generalitat de Catalunya (Ref Nº: SGR2005-00568). This website brings together scientific information about the vocabulary of the Ibero-Romance languages and their dialectical variations. The objective of this site is two fold: on the one hand, it aims to provide a tool for all Internet users to obtain both diachronic and synchronic information about Hispanic vocabulary, and on the other hand, it aims to provide a useful tool for the work on the Nuevo Diccionario Histórico of the Real Academia Española. The Portal de léxico hispánico fundamentally brings together information of bibliographic, lexicologic and documentary nature relating them to a wide range of data bases. It contains data from the digital version of the Diccionario crítico etimológico castellano e hispánico by J. Coromines and J. A. Pascual (Madrid: Gredos, 1980-1991) together with data from other sources. The consultation interface of the portal allows bibliographic, diachronic, diastratic, diatopic, etymologic, graphic, phonetic-phonological, morphosyntactic and semantic information to be searched for specific words, as well as their documentation in ancient and modern texts. In the communication, detailed information will be available about the origin, aims, characteristics and present state of the Portal de léxico hispánico. There will also be a demonstration of the consultation interface available on the Internet.

Derouin, Marie-Jeanne; Le Meur, André

3. Reports on Lexicographical and Lexicological Projects

Full Paper

Many things have changed in the field of dictionary production during these last ten years. With the introduction of digital support and networking, the lifespan of dictionaries has been considerably extended. The dictionary manuscript has become a unique data source that can be re-used and manipulated time and again by numerous in-house and external experts. The traditional relationship between author, publisher and user has now been expanded to include other partners, such as data-providers-either publishers or institutions or industry-partners, software developers, and language-tool providers. All these dictionary experts need a basic common language to optimize their work flow and to be able to co-operate in developing new products while avoiding time-consuming and expensive data manipulations. Dictionary users also need to receive more reliable information about new lexicographic products. In this paper we will first of all present the ISO standardization for Lexicography which takes these new market needs into account, and then go on to describe the two new standards: *Presentation/Representation of entries in dictionaries* which was published in March 2007 and *Lexicographical production and marketing: Concepts and vocabulary* which was launched in the summer of 2007. In conclusion, we will outline the benefits of standardization for the dictionary publishing industry.

Construir un diccionario de derivación del español en el siglo XXI. La arquitectura de la información al servicio de la lexicografía

Díaz García, María Teresa; Mas Álvarez, Inmaculada

3. Reports on Lexicographical and Lexicological Projects

Short Paper

In the age of communication, new technological advances are made everyday in the field of information and linguistics. *Deriv@* is a system that makes the most of linguistic data. All data managed by this application are connected with derivational morphology, since derived words constitute the starting point from which relationships with other linguistic elements are established. *Deriv@* is a linguistic database with two models of representation: one for all Spanish words created by means of derivation; the other for the corresponding Latin words. We offer a grammatically-analysed corpus, both synchronically and diachronically, to make customised queries according to the user's interests, and also to create a dictionary of derived words. *Deriv@* allows and facilitates non-restricted access to all the information available in the two databases. This model is valid for any form derived from Spanish or from other Romance languages. It offers users the possibility of using it real time, thus letting them interact with the system and contribute to its improvement.

El Diccionari de l'Institut d'Estudis Catalans (2007): el tractament de la pronominalitat verbal Fradera, Imma; Fullana, Olga; Montalat, Pere; Santamaria, Carolina

3. Reports on Lexicographical and Lexicological Projects

Short Paper

The aim of this paper is to explain the work that has been carried out in the Oficines Lexicogràfiques to regularize the treatment of verbal pronominality in the *Diccionari de la llengua catalana* of the Institut d'Estudis Catalans (DIEC) and to show the result that the second edition of this dictionary, published in April of 2007, reflects. It is well known that the DIEC is not a dictionary started from scratch, but it is based on the *Diccionari general de la llengua catalana* of Pompeu Fabra (DGLC). The first and second editions of the DIEC represent an updating of the DGLC, not only regarding nomenclature and definitions but also regarding the revision and the systematization of some lexicographic treatments, as in the case of the treatment of verbal pronominality. Our exposition is divided into five sections:

- 1. First, we will briefly present the treatment of verbal pronominality in the DGLC and in the first edition of the DIEC (DIEC1).
- 2. Then, we are going to analyze which type of linguistic constructions are considered pronominal in these two dictionaries.
- 3. After that, we are going to state the theoretical frame that allowed us to establish the bases on which the lexicographic criteria are built.
- 4. Fourth, we are going to explain the lexicographic criterion that governs the treatment of verbal pronominality in the second edition of the DIEC (DIEC2) and some results of the application of this criterion will be shown.
- 5. Finally, we will conclude our exposition by presenting the reach of this application.

Diskurswörterbuch - Zur Konzeption eines neuen Wörterbuchtyps

Kämper, Heidrun

.....

3. Reports on Lexicographical and Lexicological Projects

Short Paper

After a brief discussion on the term discourse, discourse will be related to the tasks of a discourse dictionary. The paper goes on developing the subject of discourse lexicography, which is a lexicographic presentation of discourse vocabulary, of the net of its semantic relations, and of the societal and historical circumstances of the usage people have made of it. This background will be useful for the presentation of two types of discourse dictionaries. On the one hand, they are based on the same primary conception. On the other hand, they are adapted to the respective discourse constellations. The first example is the result of a project on the early post-war period and presents the already-existing discourse dictionary of this project. The content of this dictionary is the vocabulary of three different groups, which participate in one discourse and specifically represent its main item. Since this dictionary also exists in electronic version, this concept will be proved by examples taken out of this version. The second example refers to a project running on the 1967/68 protest period. The vocabulary of this discourse makes up a set of several single discourse items, while these items constitute the leading subject of the discourse of 1967/68: democracy. Thus, the task of the lexicographic description of a complex discourse like this is not at least: to assign the discourse vocabulary to the single discourses and to describe the different usages relating to these single discourses. The paper ends with a draft of a lexicographic program based on the type discourse dictionary.

Turning Roget's Thesaurus into a Czech Thesaurus

Klégr, Ales

3. Reports on Lexicographical and Lexicological Projects

Short Paper

Turning *Roget's Thesaurus* into a Czech Thesaurus in a report on how a thesaurus of the Czech language was compiled on the basis of *Roget's Thesaurus*, the following issues are covered:

- Reasons for undertaking the thesaurus project-to redress the unbalance between the semasiological and onomasiological description of Czech by compiling a counterpart to the two large alphabetical dictionaries of Czech:
- 2. Strategy and philosophy, and the choice of the source text-combination of translation and original compilation; decision to use an available and well-proven model, a shorter version of *Roget's Thesaurus*, to resolve the issue of a classificatory system and format;
- 3. Phase one: a project grant-awarded by Charles University for a three-year project, Computerized Thesaurus of the Czech Language, resulting in a preliminary translated version of the Czech thesaurus and the publication of a sample volume as an output;
- 4. Phase two: expanded version for publication-moving from translation to original compilation for greater autonomy of the Czech thesaurus and expanding the average of 80 items per entry to 300 using Czech sources; specific rules required for entry structure, the type and order of subentries, etc, to ensure the uniform format of the entries:
- 5. Compiling the index-to achieve the standard index-length equal to that of the dictionary text, a procedure combining manual and mechanical shortening was devised to abridge the dictionary text;
- 6. Conclusion. Compilation of a thesaurus via translation from another language is a possible procedure. Supplementing translation with original compilation based on target-language resources is nevertheless recommended if a truly national thesaurus is to result.

MedLex+: An Integrated Corpus-Lexicon Medical Workbench for Swedish

Kokkinakis, Dimitrios; Toporowska Gronostaj, Maria

3. Reports on Lexicographical and Lexicological Projects

Full Paper

This paper reports on the work carried out developing MedLex+, a medical corpus-lexicon workbench for Swedish. This project, which is still under active development, has been going on for some years now within the Department of Swedish language at Göteborg University. At the moment, the workbench incorporates:

- i. an annotated collection of medical texts-including 20 million tokens and 45,000 documents,
- ii. a number of language processing software programs, including tools for collocation extraction, compound segmentation and thesaurus-based semantic annotation, and
- iii. a lexical database of medical terms-containing 5,000 medical entries. MedLex+ is a multifunctional lexical resource due to a structural design and content which can be easily queried. The medical workbench is intended to support lexicographers compiling lexicons and also lexicon users more or less initiated in the medical domain. MedLex+ can also assist researchers working on either lexical semantics or natural language processing (NLP) applications with focus on medical language. The linguistically and semantically annotated medical texts in combination with a set of smart queries turn the corpora into a rich repository of semasiological and onomasiological knowledge about medical terms and their linguistic, lexical and pragmatic properties. These properties are recorded in the lexical database with a cognitive profile. The MedLex+ workbench seems to offer a constructive help in many different lexical tasks.

Semiotic Conceptualization of Human Body: Lexicographical or Database System Description? Kreydlin, Grigory E.

3. Reports on Lexicographical and Lexicological Projects

Short Paper

The paper focuses on some possible means of lexicographical description of the human body both in natural languages and in nonverbal semiotic systems. The Russian language and the Russian body language present significant material for constructing semiotic representation of the human body and its parts. Two basic modes for such a representation-explanatory dictionaries and database systems-are discussed in detail. It is argued that database systems provide, on the one hand, more explicit and rigorous format for the comparative analysis of gestures, postures, mimics and other nonverbal signs, and natural language expressions, on the other hand, than explanatory language and gesture dictionaries.

Repertorio analitico dei dizionari bilingui francese-italiano

Lillo, Jacqueline

3. Reports on Lexicographical and Lexicological Projects

Short Paper

This research has aimed at analytically listing bilingual French-Italian, Italian-French dictionaries available in public and private libraries. The team of about thirty researchers has visited almost 400 libraries in France and Italy and also in Netherlands, Spain and Great-Britain. 800 different editions have been found from the first in 1583 to 2000-conventional date. An analytical description has been provided for each of them. It gives general information on their author, title, printing city, publishers, volume measures, typology, etc., and more specific information on the metalexicographical languages, the paratext-introduction, illustrations, etc., the nomenclature and the microstructure itself-phonetics, etymology, descriptive glosses, labels, examples, etc. All the information, registered in a data base, allow us to present a pretty realistic view on bilingual French-Italian and Italian-French lexicography from the very beginning. Various figures are included in this article to show: the number of dictionaries per century, the most productive authors-over 15 editions, the production per author-it is very interesting to see that almost half of the authors have published only one dictionary, the places of publication per century and all together. This bibliography-Quattro secoli di lessicografia franco-italiana 1583-2000. Repertorio analitico di dizionari bilingui-is published by Peter Lang.

Ein elektronisches Lexikon im OLIF-Format für die Erzählanalyse

Luder, Marc; Clematide, Simon; Distl, Bernhard

3. Reports on Lexicographical and Lexicological Projects

Poster

We present the JAKOB lexicon, a semantically rich German lexical resource, and its migration to the OLIF format (Open Lexicon Interchange Format). This lexicon is part of a web-based text and narrative analysis application. The JAKOB narrative analysis is a qualitative research tool to systematically analyze patient's narratives. It conceptualizes narratives as dramaturgically-constructed linguistic productions and interprets them with regard to the un-conscious conflicts of the narrator contained there in. In this process, narratives are extracted from transcripts, then a linguistic analysis is performed, and after that the vocabulary is encoded according to predetermined psychological conceptual categories incorporated in the JAKOB lexicon. The need for the proper treatment of multi-

word units in the JAKOB project made OLIF a reasonable target format. OLIF is word-sense oriented and allows a broad linguistic description-syntactical, morphological, and semantic-for each lexical entry. The OLIF data categories and attributes are well defined in the case of German but it turned out that the data-category labels in OLIF aren't specified very clearly sometimes. In addition to that, there are few resources that prove their practical use. In a corporate project, the lexicon was halfautomatically reassessed and finally migrated. OLIF is an open XML-based standard for structuring lexical data and provides a rich choice of linguistic categories and predefined values. Multi-word entries represent an essential improvement for the JAKOB application. The narrative texts represent spoken language; therefore the utterances aren't well formed, in most cases, and not eligible for a standard syntactic analysis. We use a construction-grammar approach to gather the sense of multi-word expressions in the text and to match them to lexicon entries with their corresponding conceptual categories. We use multi-word entries as containers for constructions-form-meaning units-like idioms and collocations. Further investigations will show to which extent more general constructions can be lexicalized. Our project goal is to improve precision in coding the JAKOB narratives. We decided to create an OLIF database, using the XML schema as the basis for the database structure. Thus, import and export of OLIF data is straight-forward. The implementation is object-oriented and solely based on open source software using PHP / MySQL.

GEST 2.0: A Gestionary of Emblems for Cross-Cultural Communication & Media Accessibility

Mesa Lao, Bartolomé; Bartoll Teixidor, Eduard

3. Reports on Lexicographical and Lexicological Projects

Short Paper

Nowadays, media accessibility is gaining relevance and it is becoming socially more and more important in almost all areas of our daily lives. Among the different types of media products, the most common ones are those combining the audio and the visual channel. In such products, semantic content can be transmitted by using the audio channel as well as the visual, or both at the same time. From this perspective, an important semantic aspect that can be conveyed by visual means is body language. Body language is a broad term for forms of communication using body movements or gestures instead of-or in addition to-sounds, verbal language, or other forms of communication. Due to the relatively high degree of information contained within human gestures, it seems to be necessary to open new fields of research based on this paralanguage. The idea behind this study of body language is to present the structure of a multicultural 'gestionary', a multimedia dictionary in progress of culturally coded gestures for audiovisual translators. The study of semantic aspects of culture-based gestures should prove useful for audio describers when dealing with meaning, context of use and verbal formulation of such gestures. For instance, compared to the field of second-language learning, this topic has not kept pace with the level of interest in the area of Audiovisual Translation. The creation of a new multimedia dictionary of gestures reflects our interest in putting together in a single project three complementary fields:

- 1. The creation of new tools for audiovisual translators.
- 2. The possibilities of web 2.0 technologies to develop socially generated projects.
- 3. The need to find new ways to go further in media accessibility.

Introducing BAWE: A New Lexicographical Resource

Nesi, Hilary

3. Reports on Lexicographical and Lexicological Projects

Poster

This paper reports on the compilation of the British Academic Written English (BAWE) corpus, a collection of almost 3000 proficient student assignments produced at three representative universities in the UK. BAWE was designed to fill a gap in current corpus resources by complementing other writing collections which represent expertly written academic text-such as the TOEFL 2000 Spoken and Written Academic Language Corpus, or non-expert and non-discipline specific student writing-such as the Louvain Corpus of Native English Essays, and the Cambridge Syndicate Examination corpus. Prior to the development of BAWE the few small corpora of writing produced by university students within their disciplines had either been compiled for individual scholarly purposes, or were in the form of inadequately documented and unannotated 'essay banks' for student use. The BAWE corpus, in contrast, is a large, formally compiled collection of assignments at four levels of study, from first year undergraduate to masters level, accompanied by detailed contextual information. Thirteen broad macrogenres have been identified in the corpus, including the *essay*, of course, and writing generically

similar to the published research article, but also including other types of writing, neglected in the literature, which reflect the purpose of university level study. The full corpus will be freely available to researchers from January 2008, and it is foreseen that it will provide a-currently unique-resource for designers of dictionaries for advanced learners, particularly those learners studying at university level in the medium of English.

Development of the Integrated Concordancer for the Corpus of the 17th to 19th Century Culinary Manuscripts

Paek, Doo-hyun; Nam, Kil-im; Lee, Mi-hyang; Ahn, Eui-jeong; Song, Hyeon-ju



Poster

The aim of this project is to develop the Integrated Concordancer for food-related terms used in Korean culinary manuscripts from the 17th to the 19th century. The Integrated Concordancer may be utilized by Korean linguists who wish to make use of culinary manuscripts as research materials for the history of the Korean language. Additionally, it might be useful for culinary scholars of traditional foods, and also for the general public. The tasks of the current project are twofold. The task is, firstly, to construct a corpus by collecting hand-written culinary manuscripts written between the 17th century and the 19th, and develop a web-based search engine. Secondly, to extract headwords of everyday words from the corpus of the 17th to 19th century manuscripts and compile a source book for traditional culinary terms by making and utilizing concordance data by frequency, part of speech, and semantic pattern. The current project is a two-year project-starting in August 2007, ending in July 2009-which will eventually become available to the public.

Diccionario de los glifos maya con descripción visual estructural

Pichardo-Lagunas, Obdulia; Sidorov, Grigori;

⚠ 3. Reports on Lexicographical and Lexicological Projects

Short Paper

The deciphering of Mayan script is an intricate but interesting problem. During years, the community of Mayan researchers was not open to the usage of computer tools. Still, the progress of the computer science and the current state of Mayan research proves the necessity of this type of software. We present the project related to the development of Mayan script database, which is the first necessary step in development of computer representation of Mayan script data. The database contains several tables and allows for various queries. The main idea of the project is the development of the system that would allow managing Mayan script data for specialists as well as for persons without any previous knowledge of Maya. This includes structural visual description of glyph images, expert system facilities, and, in future, calculation of glyphs similarity and development of digital corpus for analysis of similarity of the contexts on the fly. Another possible direction of further investigations is confirmation of deciphering results using large corpus data.

Presentación del Diccionario Coruña de la lengua española actual

Porto Dapena, José-Álvaro; Conde Noguerol, Eugenia; Córdoba Rodríguez, Félix; Muriano Rodríguez, Montserrat

🐧 3. Reports on Lexicographical and Lexicological Projects

Short Paper

The *Diccionario Coruña de la lengua española actual* has been a work in progress since the year 2000. It is a monolingual dictionary of current standard Spanish language that covers both European and American Spanish. One of the main features of our dictionary is the possibility of two access methods-alphabetically/semasiologically and onomasiologically. Our work begins with an alternative scheme of the structure of a lexical semantic field. This field guides the whole process involving the compilation of the semasiological section. It is useful, for instance, to separate real meanings-i.e. invariant or paradigmatic meanings, but not senses-although different senses will be present in the dictionary under its relating meaning. We understand that there are different meanings in a word when it belongs to different lexical paradigms. The verb *componer* in *Albéniz compuso Iberia* belongs to the field of the verb *crear* 'to create', but in *El relojero compuso el reloj* the verb means *arreglar* 'to repair'. The paradigmatic section will become a structural dictionary of the Spanish language. We are not trying to create a thesaurus, but rather to describe the structure of Spanish vocabulary applying the linguistic criteria of structural semantics. This structure will be a set of trees-one for each field-showing the

semantic relations: synonyms, hyponyms, hypernyms, meronyms, etc., as well as relations like causativity. Every meaning in the alphabetical section is linked with these trees.

Pedagogical Criteria for Effective Foreign Language Learning: A New Dictionary Model

Pujol, Dídac; Masnou, Joan; Corrius, Montse

3. Reports on Lexicographical and Lexicological Projects

Poster

This paper presents the pedagogical criteria used in the making of the Easy English Dictionary with a Catalan-English Vocabulary (EED), a new dictionary model for lower intermediate learners of English as a foreign language. The dictionary described renders an account of the philosophy and the results of a specific lexicographical project centred on English as the L2 and Catalan as the L1. The pedagogical criteria on which the EED is based are: structural criteria, linguistic criteria, cultural criteria and illustration criteria. The paper examines the treatment that each of these four aspects has received in different types of dictionary and, after pointing out their weaknesses and limitations, proposes a new dictionary model that seeks to promote a more effective learning of foreign languages. The most innovative aspect of the EDD concerns its structure: the EED is a bilingualized dictionary-i.e. it combines the advantages of both monolingual and bilingual dictionaries, but unlike in classical, immediate bilingualized dictionaries, in the new-deferred-dictionary model the L1 translation does not minimize the L2 definition. The EED also takes advantage of the L1 language and culture, something which the vast majority of dictionaries for L2 learning do not do: the new model uses L2 words similar to L1 ones as well as cultural referents familiar to the L2 learner. Finally, the new dictionary model presented in this paper considers illustrations as an important means of contextualization and linguistic production.

On The Lexis of Cloth and Clothing Project

Rutten, Stuart Nels

3. Reports on Lexicographical and Lexicological Projects

Short Paper

A proposal for discussing the goals, limits, benefits and problems of creating a multi-lingual dictionary, using the web-based Lexis of Cloth and Clothing Project as a basis for consideration. Using PowerPoint slides and examples from ongoing work, the presentation will demonstrate both the methods in use for the dictionary and will raise questions regarding lexical practice when developing dictionaries for describing the lexis of multilingual communities.

ISLEX-An Icelandic-Scandinavian Multilingual Online Dictionary

Sigurðardóttir, Aldis; Hannesdóttir, Anna; Jónsdóttir, Halldóra; Jansson, Håkan; Trap-Jensen, Lars; Úlfarsdóttir, Þórdís

3. Reports on Lexicographical and Lexicological Projects

Full Paper

This paper presents ISLEX, an inter-Nordic project based in Reykjavík, Iceland, with partners in Gothenburg, Bergen and Copenhagen. The aim of the project is to develop an online dictionary site with Icelandic as the source language and the three Scandinavian languages-Swedish, Norwegian (with two official standards) and Danish-as the target languages. The dictionary is planned to contain 50,000 lemmas, with a development period of six years. In 2011, or possibly sooner, the site will be publicly available on the Internet, free of charge. In this article, the main features of the project are presented with particular emphasis on database design, editorial principles and priorities.

e-LIS: Electronic Bilingual Dictionary Italian Sign Language-Italian

Vettori, Chiara; Felice, Mauro

3. Reports on Lexicographical and Lexicological Projects

Short Paper

This paper presents the design of e-LIS (Electronic Bilingual Dictionary Italian Sign Language (LIS) - Italian), an ongoing research project at the European Academy of Bolzano started in 2004. It is the first attempt to build a sign-language reference dictionary that contains definitions and examples in the sign language itself and which offers a search engine that guides the user in the process of reconstructing and retrieving the sign he is looking for. Thus, not only can users go from Italian to LIS, but also from LIS to Italian.

Verbs of Science and the Learner's Dictionary

Williams, Geoffrey

rwi 🕳 🕳

3. Reports on Lexicographical and Lexicological Projects

Short Paper

This paper looks at how the verbs of science are displayed in the OALD and then compares them to a specialised corpus. The individual entries will be studied to see whether the scientific aspect is signalled, how the definitions are structured, and if implicit information is carried in the examples. The examples are analysed using Halliday's Systemic and Functional Grammar (SFG). The results in each section are compared with usage cases found in a subcorpus from the British National Corpus and in a specialised corpus.

La définition dans les dictionnaires bilingues: problèmes de polysémie et d'équivalence interlangues Bouchaddakh, Samia

4. Bilingual Lexicography

Short Paper

In this paper, we intend to tackle the issue of the definition in the so-called "encoding bilingual dictionaries" or "active bilingual dictionaries". We focus more specifically on French-Arabic dictionaries. Our main objective is to demonstrate the interest of definition, precisely the one based on the principles of the Explanatory and Combinatorial Lexicology, for both the lexicographer and the user of the active bilingual dictionary. This kind of definition allows us to identify the internal structure of lexical meaning, to select the best equivalent and to make explicit the relations of polysemy and equivalence between the two languages.

La place de la morphologie constructionnelle dans les dictionnaires bilingues: étude de cas Cartoni, Bruno

4. Bilingual Lexicography

Short Paper

In this study, we questioned the role and the place of constructional morphology in bilingual dictionaries. As it is the case with monolingual dictionaries, it is becoming increasingly common to find morphological elements in bilingual dictionaries. Presenting such elements in a bilingual context, however, raises questions as to representations of meaning, translation selection and choice of examples. We show a comparison of the treatment of productive prefixes of Italian in 5 bilingual dictionaries-French was the target language in three of them, English in the other two. This comparison shows the differences in the treatment of prefixation and its coverage. First, we notice that not all the prefixes are represented in these dictionaries, and that some lexical elements are labelled as prefixes even if this status is contested. In terms of treatment, we examine in particular aspects of polysemy and of multiple translations. Regarding the polysemy of some prefixes, many dictionaries simply avoid marking the difference, while others add specific sense indicators. The most frequently used method for presenting the translation of prefixes, however, is through examples. In analysing these examples, we notice the inadequacy of certain translation equivalents, especially from a productive point of view. There is a noticeable absence of any information on the productivity of these constructional elements. From the perspective of understanding neologisms and sometimes even producing them, this lack of precision is regrettable.

Friend or Enemy? A Case Study of Lexical Comparison between Italian, German and Japanese Bilingual Dictionaries

Costantino, Mauro

4. Bilingual Lexicography

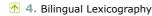
Student Paper

This work aims to demonstrate with a case study of bilingual Italian-German, Italian-Japanese, and German-Japanese dictionaries, exactly to what extent lexical drifting can manifest itself. Through a few examples, the study will raise issues of gender and cultural problems in bilingual dictionaries, and the peculiar case of a distant language like Japanese in comparison to German and Italian. The importance of considering both cultural and background knowledge when both building and consulting bilingual dictionaries will be stressed, in order for one not to obtain an outcome contrary to expectation. Finally,

suggestions about how to overcome these problems will be made, with consideration given to the difficulty of dealing with strongly culturally-bound terms and their meanings.

User-friendly Dictionaries for Zulu: An Exercise in Complexicography

de Schryver, Gilles-Maurice; Wilkes, Arnett

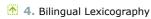


Full Paper

In this paper the main features of Bantu lexicography are analysed through several case studies of Zulu dictionary features. Examples from both existing dictionaries as well as a forthcoming reference work are used in the analysis, which develops from verbs and nouns, gradually including more word classes, and ending with a detailed study of possessive pronouns. The latter serves as one example of the complex mappings that occur in the creation of bilingual dictionaries where the two languages involved have very different grammatical structures. In this case, one concept-that of a possessor and its possession-has only a few members in English, but hundreds in Zulu. It is shown how one can deal with such a mass of data in a structured, systematic and linguistically-sound way, all the while aiming to produce a user-friendly end product. All the members of this single concept are collectively referred to as a paradigm, and it is indicated that some members are homonymous with members of other paradigms-a fact which exponentially complicates the dictionary treatment. Several suggestions are made for the lexicographic treatment of conjunctively written Bantu languages, and all the claims, as well as all the data, are based on facts derived from a large general-language Zulu corpus.

Systemhaftigkeit in zweisprachiger Lexikographie: Zur Darstellung deutscher und russischer Possessivpronomen

Dobrovol'skij, Dmitrij; Sarandin, Artem



Short Paper

In this paper, we discuss systematic approaches to the lexicon representation in bilingual dictionaries. For empirical data we take how possessive pronouns are treated in the *New Comprehensive German-Russian Dictionary(NCGRD)*. Every type of pronouns builds a closed class of words, and its members have to be described in the same terms and in the same format. In the ideal case, each deviation from lexicographic uniformity must be understood as a signal that a given word displays unique linguistic features as compared to other members of the same class.

An additional difficulty is that the German and the Russian pronoun systems are arranged according to non-identical principles. Therefore, every German possessive pronoun can be translated into Russian not only by its "regular" equivalent (мой, твой, его etc.), but also by the pronoun свой, under certain syntactic and discursive conditions. These conditions have to be explicitly stated in the entry. There are also contexts where a given German possessive pronoun has to be omitted in the Russian translation, cf. hast du (dir) deine Hände schon gewaschen? - ты уже вымыл руки? hast du schon mit deiner Миtter gesprochen? - ты уже поговорил с матерью? This phenomenon clearly depends on the semantic class of the noun modified by the pronoun in question, i.e. it is rule-governed; these rules need to be declared in the lexicographic description.

The pronoun *mein* in certain contexts must be translated not by мой my but by наш 'our'; cf. *meine Fakultät - наш факультет*. The reason is that the Russian word мой 'my', being different from the German word *mein*, denotes exclusive possession, i.e. мой means 'mine and only mine'.

All these cases will be illustrated by entries from NCGRD. They demonstrate far-reaching uniformity, so that every deviation from the uniform format-as compared to both various possessive pronouns in German and their Russian equivalents- is a meaningful constituent of their lexicographic representation.

La equivalencia en los diccionarios bilingües: un enfoque semántico

Fernández Fernández, Juan

🚹 4. Bilingual Lexicography

Student Paper

In this paper, we present a proposal to analyze bilingual dictionaries' lexical equivalents. Lexical correspondence between two languages can be analyzed adopting the point of view of different

linguistic domains-e.g. pragmatics or syntax. Our approach is based upon semantics. The aim of this paper is to find out ways of discovering conceptual differences between bilingual dictionaries' equivalents which are supposed to have the same meaning. For this reason, we put into practice a semantic analysis based on semantic decomposition of monolingual dictionaries' definitions of the equivalents given by bilingual dictionaries. We make use of the lexicographical definition, the semantic metalanguage-especially, the metalanguage which consists in the up-to-date Wierzbicka's semantic primes, and corpus linguistics. These conceptual analysis' results are shown by means of conceptual trees which relate the different concepts which are part of the given equivalents' definitions. Thanks to the Wierzbicka's semantic primes, which have proven to be universals, we can follow an objective reasoning behind the differences and similarities between proposed lexical correspondences by these dictionaries. As a result, we can gain valuable conceptual knowledge that will sensitize us to languages' lexical variety and richness. This is something that cannot be easily shown in commercial bilingual dictionaries by reasons of time and pressures of the publishing market. Thus, our proposal is a small contribution to reflect on these dictionaries' functions regarding their users-language learners or language professionals. A further step is to provide an alternative to their common conceptual structure, in order to compare the lexicon of two languages in a more reasonable way, or in accordance to language's anisomorphism.

Le casse-tête des dictionnaires bilingues pour traducteurs: le cas des dictionnaires arabes bilingues Franjié, Lynne

4. Bilingual Lexicography

Full Paper

Translators have long called for translators' bilingual dictionaries that would include ready-to-use equivalents. In order to make them, lexicographers must ponder on the bilingual dictionary as a translator's working tool, hence not only from a lexicographic point of view, but also from a translation one. Studying the translations included in dictionaries-such as Arabic bilingual ones-shows that they are problematic as they are *out-of-context* and *transitional*. The issue becomes even more complex when the entries at stake are culture-related, for it is common knowledge that 'shared culture'-for example, that related to social and religious realities-often varies from one culture to another, as it is the case between Arab and French cultures for instance. Semantic voids are numerous in these cases and the lexicographer finds himself compelled to make difficult choices. It is these choices that this paper means to examine by analysing cultural entries in largely-used Arabic-French and Arabic-English bilingual dictionaries. Studying these cultural entries entails determining the types of translations included in the dictionaries. One could then conclude that Arabic bilingual dictionaries are in fact meant for translators although they are deficient in some ways. One solution would be to include authentic functional translations, namely by using parallel corpora from which translations can be extracted, thus enriching dictionary entries.

What to Say about *mañana*, *totems* and *dragons* in a Bilingual Dictionary? The Case of Surrogate Equivalence Gouws, Rufus; Prinsloo, Danie

🚹 4. Bilingual Lexicography

Short Paper

There are frequent instances in any given language pair where a suitable translation equivalent is not available to be treated as source and target language in a bilingual dictionary. This is known as zero equivalence and can be regarded as the most complex type of equivalence to be dealt with in a bilingual dictionary. This paper will focus on the various ways in which lexicographers of different dictionaries deal with the lack of equivalence and the subsequent use of surrogate equivalents. There are a number of strategies that the lexicographer can use when dealing with instances of zero equivalence, e.g. the use of glosses, paraphrases, illustrations and even text boxes with lexicographic comments. This paper suggests different types of surrogate equivalents based on user needs, and it will be done in accordance with the relevant dictionary functions, i.e. the cognitive function and the communicative functions of text reception, text production and translation. A linguistic gap can be identified when the speakers of both languages are familiar with a certain concept but when one language does not have a word to refer to it, whereas the other language does have such a word. A referential gap can be postulated when a lexical item from language A has no translation equivalent in language B. This would be because the speakers of language B do not know the referent of the lexical item from language A. Acknowledging different degrees of complexity in the relation of surrogate equivalence leads to a tiered view of the concept. The first level in the hierarchy provides for linguistic gaps where a mere gloss or brief paraphrase of meaning will suffice. More complicated are the gaps where the surrogate equivalent also has to provide grammatical guidance. The top tier in the hierarchy provides for referential gaps where taboo, culture-specific or sensitive values have to be expressed.

Du support d'information à l'outil lexicographique: la lexicographisation du guide touristique Leroyer, Patrick



4. Bilingual Lexicography

Short Paper

The development of lexicographic products for tourists is one of the most productive lexicographic activities in the world, with the publication of paper and online bilingual travel dictionaries, phrase books, and tourist guides often containing a dictionary component. Additionally, software companies propose multilingual, downloadable dictionary solutions that can be printed on demand or consulted via a PDA or a WAP phone. There are two explanations to this lexicographic infatuation: the huge expansion of tourism world-wide and the extensive communicational and knowledge-oriented informational needs of tourists. However, metalexicography has shown very little promise to this field of lexicography, and has solely dealt with the communicative needs of tourists. In this contribution, I will outline a new lexicographic method that can be used to satisfy the aforementioned needs of tourists-also, namely lexicographisation-which is the lexicographic transformation of tourist guides performed to ensure fast and easy access to user and situationally-adapted information.

Lexical Entries and the Component of Pronunciation in Tshivenda Bilingual Dictionaries Mafela, Munzhedzi James



4. Bilingual Lexicography

Short Paper

Lexical Entries and the Component of Pronunciation in Tshivenda Bilingual Dictionaries Pronunciation is defined by Allen (1990) as the way in which a word is pronounced, especially, with reference to a standard. It involves a set of symbols, each of which always represents the same sound. Languages pronounce orthographic symbols differently. In some languages, orthographic symbols written identically are pronounced differently. Tshivenda is characterized by orthographic symbols which are written identically, but can be pronounced differently. These are orthographic symbols such as tsh, ts, tsw and pf. The same orthographic symbol can be pronounced as an aspirated sound or ejected sound. For example, the orthographic symbol ts can be pronounced as [ts'] in tsika (to press down) or [tsh] in tsimbi (metal). Poulos (1990) says that the actual pronunciation is determined by the words in which the orthographic symbols are used. Definitions of headwords in a dictionary consist of many components, for example word category, morphology, pronunciation, etymology, meaning and illustrative examples. The pronunciation element becomes a necessity for bilingual dictionaries because the addressees of these dictionaries may be learners of a foreign language. "Pronunciation is, after all, the integral part of the lexical item", as Sobkowiak put it in 2003. Giving the meanings of words is often thought to be the main purpose of a dictionary. It should be also noticed, however, that "the dictionary also contains other areas of information useful to the user" (Underhill 1980). Knowledge about pronunciation helps in checking any spelling the user is not sure of. Almost all Tshivenda dictionaries are bilingual and are therefore learner's dictionaries. The compilers of the dictionaries did not include the component of pronunciation in the definition of lexical items. Therefore, learners of Tshivenda find it difficult to pronounce orthographic symbols which denote more than one phonetic sound. This presentation seeks to highlight the lack of pronunciation component in Tshivenda bilingual dictionaries and its effects on learners of the language. Three Tshivenda bilingual dictionaries will be used to illustrate some points in this regard.

L'accès aux Séquences Figées dans les dictionnaires électroniques bilingues Français - Italien Murano, Michela



4. Bilingual Lexicography

Short Paper

This paper presents the results of our research on a group of electronic French-Italian, Italian-French dictionaries on CD-Rom-DIF, Boch, Garzanti Clic, and Garzanti interattivo. We examine whether the characteristics of the electronic support can influence the access to the fixed sequences. This work deals particularly with the importance of diversified typography and new types of complex search-e.g. full-text search-which are now available for dictionary users.

Méthode sociolinguistique d'étiquetage du niveau de langue dans les dictionnaires bilingues (sur l'exemple d'un dictionnaire français-ukrainien)

Shevchenko, Natalya

4. Bilingual Lexicography

Short Paper

This article describes a new sociolinguistic method in the labelling of unconventional units in bilingual dictionaries. This study was undertaken as part of the preparation of a French-Ukrainian dictionary of unconventional language.

On the Presentation of Onomastic Idioms in Bilingual English-Polish Dictionaries of Idioms Szerszunowicz, Joanna

4. Bilingual Lexicography

Poster

The paper discusses the lexicographic description of onomastic idioms in contemporary English-Polish dictionaries of idioms, with a special focus on the cultural character of the onymic component. Onymic idioms are distinguished as a group of particular interest for lexicographers, since onyms tend to be culture-bound elements of international, national or local character. Thirteen English-Polish dictionaries of idioms have been analyzed so that the presentation of onymic idiomatic expressions in such lexicographic works could be discussed. The macro- and micro-structures of such dictionaries are analyzed in order to identify the problem areas in the bilingual description of onomastic idioms. From the cultural-linguistic point of view, two methods of presenting onomastic idioms are observed in the dictionaries, i.e. the inclusion of cultural information regarding the onym or the exclusion of such information. In the case of the inclusion of cultural information, the lack of consistency is common in one dictionary, i.e. some of the onyms are commented on, while others are not described at all. Since onyms tend to be culturally-specific components of idioms, cultural information is essential to ensure a proper understanding of the idiom. The problem of insufficient lexicographic description of such fossilized phrases is presented in order to draw attention to the need for the creation of an onomastic idiom dictionary, enabling both users and advanced learners (of English) to have an insight both into the language and the culture. Bearing in mind that idioms undergo various modifications when used in particular contexts, such an approach to describing onomastic units of idiomatic character renders it possible for the user to acquire a proper command of idioms containing onomastic components.

QRcep: A Term Variation and Context Explorer Incorporated in a Translation Aid System on the Web Abekawa, Takeshi; Kageura, Kyo

5. Lexicography for Specialised Languages - Terminology and Terminography

Poster

In this paper we describe the method of exploring term variations and the contexts in which terms occur using the Web, to help English-to-Japanese translators working online. Many English-Japanese terminological dictionaries are available in electronic form, but most of them do not provide rich examples of terminological use including variations. This is a problem for translators, who may not have sufficient knowledge on the use of terms in a specific subject they are translating. In order to augment this information gap, we have developed a system that explores actual use of terms using the information on the Web.

The system proceeds as follows:

- 1. when an electronic text in source language (English) is given, the system automatically looks up entries in terminological dictionaries including their variations, using the variation expansion rules;
- 2. map the English entry to the Japanese translations;
- 3. expand variations of Japanese terms on the basis of Japanese variation rules;
- 4. search the Web and provide actual use of the term including variations within the actual context. For variation expansion, we are using Fastr Platform and defining corresponding rules for English and Japanese variations. The system is incorporated into the system that helps online volunteer translators and augments the terminology look-up functions.

ECODE: A Pattern Based Approach for Definitional Knowledge Extraction

Alarcón Martínez, Rodrigo; Sierra Martínez, Gerardo; Bach Martorell, Carme

5. Lexicography for Specialised Languages - Terminology and Terminography

Poster

In this paper we present a pattern-based approach to the automatic extraction of definitional knowledge from specialised Spanish texts. Our methodology is based on the search of definitional verbal patterns to extract definitional contexts related to different kinds of definitions: analytic, extensional, functional and synonymic. This system could be a helpful tool in the process of elaborating specialised dictionaries, glossaries and ontologies.

Environmental Terminology in General Dictionaries

Alonso Campo, Araceli

5. Lexicography for Specialised Languages - Terminology and Terminography

Poster

This paper discusses how some specific Environment-related terms commonly used in general discourse have been represented in monolingual and learner's dictionaries in Spanish. Our discussion falls within wider research on the characterization of Environment-related lexical units and the relationship between specific domain terminology and lexicographic representation. We briefly compare the information provided in general language dictionaries of Spanish with that found in other lexicographical traditions-for instance, in the English tradition-and find a lack of precision in lexicographic practice in relation to Environment-related terms. We outline some guidelines for improved representation of these units.

Gestor de terminologia multilingüe d'accés lliure

Bover Salvadó, Jordi; Grané Franch, Marta

5. Lexicography for Specialised Languages - Terminology and Terminography

Software Demo

Attending the demand of several sectors asking for a terminology management tool suitable for specific or personal use, TERMCAT has developed a free-access terminology manager, available at our website (www.termcat.cat). The tool, addressed to anyone interested in carrying out terminographic work, enables the management of any multilingual project that involves the compilation of terms in different languages and the systematization of concepts from different fields of knowledge. We would like to underline that every user would be able to customize the terminology management tool according to the project features and their personal needs, in order to speed up the process of data creation and modification. The most relevant contributions of the TERMCAT free-access terminology management tool are the following:

- Organizing the information in conceptual files.
- Including or deleting denominations, definitions, notes, contexts, observations and their attributes-grammatical category, range, linguistic hierarchy, source-in n languages.
- Consulting and modifying the properties of a dictionary-name, description, languages, ordering.
- Creating, maintaining and consulting the concept structure of a dictionary.
- Organizing the files in thematic or alphabetical order, and according to language.
- Allowing search based on the combination of a wide range of criteria: denominations, definitions and notes; and also according to language, field structure, hierarchy, grammar category or source.
- Consulting the alphabetical or thematic index of a dictionary.
- Consulting the files by edition mode or consultation mode.
- Importing and exporting a selection of files in several formats.

TESAURVAI: Extraction, Annotation and Term Organization Tool

Cardeñosa, Jesús; Gallardo, Carolina; Maldonado-Martínez, Ángeles; Vergara, Jorge

5. Lexicography for Specialised Languages - Terminology and Terminography

Software Demo

TESAURVAI is a tool for extracting, annotating and organizing terms from a collection of digital documents. The main contribution of TESAURVAI is the unification of a term extractor and a thesauri builder in the same tool. The term extractor identifies terms, words and phrases in the input digital texts that are transferred to the thesaurus builder. TESAURVAI follows the international standards for the construction and management of thesauri, and it provides the following facilities: on the one hand, it is a tool to create thesaurus from scratch, allowing for the extraction, creation, edition and annotation of terms, as well as providing a user-friendly interface for establishing relations between terms and performing basic or advanced searches of terms. On the other, it is a tool to manage several thesauri and to import and export existent thesauri from text or XML files. Finally, TESAURVAI can build alphabetical, hierarchical and permuted indexes to be printed or exported as reports. TESAURVAI has been developed in Java and requires and external database to store the user's thesauri. The tool is compatible with any database manager provided with a Java Database Connectivity (JDBC) file, such as MySql or Postgres. This tool has been developed within the framework of the PATRILEX (HUM2005-07260/FILO) project, sponsored by the Spanish Minister of Education. Currently, TESAURVAI is in a provisional version. A new version of the tool, which will be accessible on the Internet, will be available in July 2008.

Risotto, spaghetti, vino: Ingredients for a Good Gastronomic Dictionary Corino, Elisa

5. Lexicography for Specialised Languages - Terminology and Terminography

Gastronomy is commonly recognized as a basic "ingredient" of culture and tradition. As a central axis of various cultural components, food is a common denominator connecting both the Fine Arts and Science, as well as History, Anthropology, Sociology, etc. Additionally, gastronomy has become one of the world's most important professions and continues its ascent. Italy has recently witnessed this growth as being the origin of the Slow Food movement and the hosting of the first University of Gastronomic Sciences, where specific lecture courses expressly focus on culinary jargon and on the linguistic, typological and historical analysis of menus, recipe books and recipes alike. The increasing need for thorough glossaries and dictionaries devoted to detailed studies of the subject is apparent. This paper is meant to deal with the vocabulary connected to food in its broad sense, and will attempt to provide a cross section of the lexicographical state of the art and propose a possible original source to be held up as a model for gastronomy dictionaries: Newsgroup corpora on cooking. The Langenscheidt Praxiswörtebuch Gastronomie Italienisch (2005) will be investigated as an example of an exhaustive dictionary: its word list compared with the 500 most frequent occurrences of nouns, adjectives and verbs in the NUNC-cooking (Newsgroup UseNet Corpora), amongst both its Italian and German versions. Finally, a case study on adjectives describing wine is presented to suggest new entries for a wine glossary.

Léxico específico de la piel. Presentación de un proyecto terminográfico García Antuña, María

5. Lexicography for Specialised Languages - Terminology and Terminography

Poster

Poster

This project is framed within the project I+D of the MEC Linguistic strategies applied to social communication: study of communicative necessities and design of materials in the social environment of medicine, administration and business, the head of which is Doctor Miguel Casas Gómez. For the realization of this project we have had the collaboration of the business world, according to an agreement with the Association of Andalucian Leather Manufacturers (Asociación de Empresas Andaluzas de la Piel, EMPIEL) and the Technological Center of Leather. A specific agreement and a contract of service supply is in the process of being signed, all with the support and unfailing advice of the Oficina de Transferencia de Resultados de Investigación (OTRI) of the Vicerectorate of Research, Technological Development and Innovation at the University of Cádiz. The main objective of this project is the management of a base of terminological facts of leather work that permits the development of effective translation tools and the regulation of the specific language in these fields. The introduction of this lexicon is important from the formative point of view, since a lexicographic work does not exist today in which the specific terminology of leatherwork can be investigated. This lack of specific terminology is an obstacle for the communication of knowledge among the professionals involved. Furthermore, this work will complete other advancement objectives, since it is effective and useful to describe the characteristics of a product, to create a positive attitude towards the sector in the customers. This will compensate the introduction into the market of low quality products from countries like China or India, and will contribute to convincing the customers of the superiority of the offer versus that of the competitor. Furthermore, this terminological project can contribute significantly to the regulation of commercial and technical language within the sector.

Slovene Terminology Web Portal

Gorjanc, Vojko; Krek, Simon; Vintar, Spela

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

Work in the field of terminology is extensively supported worldwide as it enhances the transfer of science and technology. In Slovenia, there is a series of terminology-related activities running, and a significant number of terminology dictionaries and terminological data exist, but they are methodologically heterogeneous and often unavailable for public use. Traditionally, terminology work in Slovenia is closely connected with other activities in the filed of lexicology and lexicography, especially regarding the methodological approach to the compilation of dictionaries of specialised languages. Therefore, terminology work is mostly regarded only as a process involving the compilation of dictionaries. The paper presents the Slovene Terminology Web Portal project. The main objective of the project is to develop the Slovenian terminology portal to offer basic information on the principles of terminological work and to present a terminological database in a unified format. In the core of the presentation, there is a process of conversion of different types of existing terminology data from different sources into XML format with a simple DTD/schema and from there to unified TBX database. Simultaneously, the feasibility of linking textual resources and the extraction of term candidates with the terminological database is also shortly presented.

Prototypes and Discreteness in Terminology

Hacken, Pius ten

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

Characterizing the nature of terms in their opposition to general language words is one of the tasks of a theory of terminology. It determines the selection of entries for a terminological dictionary. This task is by no means straightforward, because terms seem to have different properties depending on the field that is studied. This is illustrated by a brief discussion of examples: terms in mathematical linguistics, traffic law, piano manufacturing, and non-terms in the reporting of general experiences. Two properties can be derived from these discussions as candidates for the delimitation of terms from general words. Firstly, the degree of specialization. This property distinguishes specialized expressions in mathematical linguistics and in piano manufacturing from non-specialized expressions in traffic law and reporting general experiences. Secondly, the lack of a prototype. In mathematical linguistics and in traffic law, the definition of terms concentrates on the boundaries of the concept. In piano manufacturing and in reporting general experiences, concepts have a prototype and fuzzy boundaries. Defining the word term as a disjunction of the two properties implies that it is a less coherent concept than general language word, because it is only the complement of the latter. When the two properties are considered in isolation, it can be shown that the degree of specialization is a gradual property whereas the lack of a prototype is an absolute property. Whether or not we choose to use the name term for it, the latter property identifies a concept that is ontologically different from general vocabulary. I will reserve the name term for concepts that do not involve prototypes and call the professional expressions in piano manufacturing specialized vocabulary. By focusing on the boundary instead of the prototype, a terminological definition creates an abstract object for which there is no equivalent in general language words. Whereas general language words only exist in the competence of the speakers, the abstract object associated with a term can exist independently of the knowledge of individual speakers. There are interesting parallels between the nature of these abstract objects and the nature of a piece of music. The creation of such an object on the basis of general language words can proceed by the selection of properties or the choice of a specific boundary on a scale.

New Voices in Bilingual Russian Terminography with Special Reference to LSP Dictionaries Karpova, Olga; Averboukh, Konstantin

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

The article is devoted to the general review of modern bilingual LSP dictionaries in Russia. Main trends in current Russian bilingual terminography are distinguished through criticism of new types of LSP dictionaries of different subject areas with special reference to linguistic and encyclopedic reference books. Evolution of lexicographic description of different special domains in English-Russian and Russian-English terminography is being traced from humanitarian subject fields-economics and

finance, business, law, mass-media and public relations, social work, immigration policy and the likeand natural sciences-biology, botany, physics, zoology-to technical disciplines-aviation, electronics, civil and nuclear engineering, etc.-and other subject areas-agriculture, architecture, philosophy, statistics, etc. Special attention is given to the analysis of Russian-English polytechnic dictionaries published in a new millennium showing the latest changes in Russian terminological vocabulary connected with borrowings of new terms and whole terminological systems-computers and new information technologies, logistics. Current developments and perspectives in progress in Russian bilingual terminology will be mentioned in the presentation.

LSP Dictionaries and Their Genuine Purpose: A Frame-based Example from MARCOCOSTA León Araúz, Pilar; Faber, Pamela; Pérez Hernández, Chantal

5. Lexicography for Specialised Languages - Terminology and Terminography
Full Paper

A dictionary is written and designed for a specific addresse (*user group*). Primary considerations in this respect are users' profiles and the special needs of the user group (Bergenholz and Nielsen 2006). User needs are inevitably linked to the knowledge level of potential readers, who have a situational context and engage in activities, which can be facilitated by lexicographic data. Such information significantly affects both the micro and macrostructural design of the lexical resource, and is directly related to Wiegand's conception of genuine purpose (Wiegand 1998:52). These theoretical parameters dealing with users' profiles, users' needs and use situation should necessarily be reflected in the way information is packed in lexicographical entries, i.e. in the way definitions are organized and structured. This article examines how LSP dictionaries deal with this issue. The example chosen is the term *aquifer*. After a brief overview of how this term appears in current dictionaries, we show how it is represented in MarcoCosta, a frame-based lexical resource that facilitates the acquisition of specialized knowledge.

Marqueurs définitionnels et marqueurs relationnels dans les définitions du DAAFAPS Ligas, Pierluigi

🖲 5. Lexicography for Specialised Languages - Terminology and Terminography Short Paper

This paper analyzes relational and definitional markers and their function in meronymic, derivational and approximate definitions of nouns as they appear in the *Dictionnaire alphabétique et analogique du français des activités physiques et sportives*, currently under preparation. It is argued that definitional markers are semantically weak lexical substitutes with a metonymic or meronymic character, placed at the beginning of the definition and belonging to the same grammatical category as the defined lexical item. It is also argued that relational markers are words or groups of words whose function in discourse is to establish logical, spatial or temporal relations between two or more elements and which thus contribute to organize the definitional sentence and to illustrate the concept denoted by the lexical item. As mentioned supra, we have decided to exclude hyperonymic definitions-since they do not start with definitional or relational markers-and to concentrate on three types of definitions: meronymic-based on the relation between a whole and its parts, derivational-based on the relation between root and affixes, and approximate-that make use of markers such as *sorte de*, *espèce de*. We will analyze a corpus of such definitions and try to establish how these markers contribute to the fulfillment of the definition's role, by following mainly R. Martin's, E. Wüster's, J. Rey-Debove's, A. Auger's, A. Condamines's, E. Martin's definitional theories.

A Constructional Approach to Terminological Phrasemes

Montero Martínez, Silvia

5. Lexicography for Specialised Languages - Terminology and Terminography
Full Paper

Specialized discourse shows regularities in the lexical and syntactic patterning of terminological units. This fact, evidenced by corpus-based analysis, has spurred a number of studies on polilexical terminological units. In spite of the available linguistic data, however, the systematic management of these units in specialized lexicography is still lacking. Apart from a few exceptions, terminological products, especially dictionaries, are inconsistent with their treatment of these units. Such arbitrary approaches are worthless within the context of the newer terminological knowledge bases. In this paper, we describe how the Lexical Grammar Model can offer an in-depth, principled description of

such units. Meaning and grammar are seen as interdependent and complementary layers. So, the basic unit of grammar is a form-meaning pairing or construction that can be described as a conventionalized combination of form and meaning. In this vein, the lexical profile of a specialized concept is composed of constructions, which reflect its collocational patterns both at a lexical and a syntactic level. Thus, we use the umbrella term terminological phraseme (Meyer and Mackintosh 1994) to include entrenched, conventional combinations of linguistic units in the form of complex nominals and predicate-argument structures. These units are conceived as constructions codifying conceptual, experiential and syntactic information concerning the lexical concepts of a cognitive frame. Consequently, the frame is the element which constrains the potential relations holding between the lexical concepts, and the construals that the frame allows are only a subset of the construals allowed by the argument-taking heads. The basic qualia structure and the domain-specific relations account for such combinations and for the inheritance phenomenon. In sum, we present a theoretical and methodological approach that accounts for the lexical profiles of concepts in a consistent way, including the description of conceptual relations as well as the terms' combinatorial potential.

Bilingual Terminology Acquisition from Unrelated Corpora

Nazar, Rogelio

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

This paper presents a simple yet effective technique for the extraction of term equivalents in different languages. In general, techniques for bilingual lexicon extraction have been related to the elaboration of parallel corpora and have yielded accurate results. However, parallel corpora of different domains and languages are not easy to compile. Because of this, some authors have explored techniques to extract a bilingual lexicon from nonparallel but comparable corpora, which are pairs of texts that are not exactly translations of each other but that roughly "talk about the same things". This paper describes an algorithm that performs bilingual terminology extraction without the need of large amounts of data; dealing with infrequent units; needing not the corpora to be comparable nor other resources like an initial bilingual lexicon to use as *seed words*. In spite of its simplicity, the results of this algorithm are comparable to those of the state of the art techniques, however it supersedes them considering that it offers a domain and language independent method specially suitable for the extraction of specialized terminology, which is the most dynamic part of the lexicon and the most difficult to acquire.

El sistema métrico decimal en la lexicografía española del s. XIX

Pascual Fernández, Luisa

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

The metric system is one of the clearest examples of the universal acceptance of scientific and technological vocabulary in nineteenth-century language. Its introduction into Spanish language coincided in time with its introduction in the other European languages. This vocabulary, however, has not always been rigorously included in dictionaries, as shown by dictionaries of the nineteenth-century. For this reason, I have decided to study the inclusion of the vocabulary related with the metric system in Spanish nineteenth-century dictionaries, this century being particularly interesting as far as the history of science and lexicography is concerned. The analysis is structured into two main parts. The first part is committed to the study of the already-mentioned nomenclature of the eleventh edition (1869), the twelfth edition (1884), and the thirteenth edition (1899) of the Diccionario de la Real Academia Española de la Lengua. The second part is devoted to the analysis of how metric vocabulary is incorporated in non-academic dictionaries-including the Nuevo Tesoro Lexicográfico de la Lengua Española. These parts are complemented with the comparison of that Spanish vocabulary with its French, English and Italian counterpart. In this sense, our focus lies in the European perspective. The conclusion of the research provides wide information about the first instance of metric vocabulary in Spanish within the European context and, consequently, we hope to shed some light on the way the analyzed dictionaries influenced each other. We hope to also be able to conclude the position of the Spanish Academy regarding this kind of vocabulary.

An English-Polish Glossary of Lexicographical Terms: A Description of the Compilation Process

Podhajecka, Mirosława; Bielińska, Monika

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

In the present paper we describe the consecutive phases in the compilation of an English-Polish glossary of lexicographical terms, which is part of a larger dictionary project-still in the making. In doing so, we address some of the issues that made the compilation procedure methodologically difficult. On theoretical grounds, the main dilemma was whether lexicographical-i.e. mainly descriptive-or terminological/terminographical-i.e. mainly prescriptive-principles should be followed, inasmuch as they result in different coverage, organisation and description of data. The most pertinent practical problem that we faced was, on the one hand, the variability of terms in English lexicographical discourse and, on the other one, the incompatibility of English and Polish terminological frameworks. It was therefore envisaged that, for the glossary to be used successfully in text reception, allowing alternative terms and determining various levels of equivalence between interlingual terms would be a necessity. The issues discussed here have been illustrated with selected English-Polish contrastive material.

Wissensdarstellung und Benutzerfreundlichkeit in einem zweisprachigen terminologischen Rechtswörterbuch: Der Fall Hochschulrecht

Ralli, Natascia; Wissik, Tanja

5. Lexicography for Specialised Languages - Terminology and Terminography

Poster

This paper presents the Italian-German Terminological Dictionary for University law in Italy and Austria. In particular, we will describe the microstructure of the dictionary and the typology of the given information with regard to the needs of the target group. The dictionary was produced and printed in 2007 by the Institute for Specialised Communication and Multilingualism of the European Academy of Bolzano on behalf of the Department for the Right to Education, University and Scientific Research of the Autonomous Province of Bolzano/Bozen-South Tyrol. The aim of this work is to compare the Italian and Austrian terminology of university law as well as to record their most recent changes and developments.

Palabras y términos "lingüística y contextualmente determinados"

Sanz Espinar, Gemma

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

Our first concern is the specificity of the terminology of the human and social sciences, which are said to be *language-dependent* or *culture-dependent*. However, we will consider the creation and the use of terms-as of all words-*language* and *context-dependent*. This assessment is contrary to traditional terminology theory, which considers terms to be univocal relations between concepts and their designation-contrary to words, and also considers terms from pure sciences and technics as language-independent and culture-independent. We will analyze what *language* and *context-dependency* mean for terms and no terms, for terms of more technical or positive sciences, as well as for terms of the human or social sciences. From a pragmatic point of view, the creation and use of any word or term is supposed to be influenced by the context in which this word or term was created/used, so that they are linked, to some extent, to the creator-author-dependency-and his language-language-dependency, the culture-culture-dependency, the place-geography-dependency, the historical period-history-dependency, and the communicative aim-dependency on the communicative aim, which includes the type of circumstance or the person the speaker talks to. This process means that for translation and for terminographic purposes we will find some specificities in these cases, but we can formulate strategies to cope with them.

Terminology Practice in a Non-standardized Environment: A Case Study

Taljard, Elsabe

5. Lexicography for Specialised Languages - Terminology and Terminography

Short Paper

Terminology as independent discipline, as well as its practical applications is not yet well established for the South African Bantu languages. The aim of this paper is to illustrate some strategies that are currently employed to ensure sound terminology practice in a non-standardized environment, and at the same time contribute to terminology and language standardization of Northern Sotho, a language

of lesser diffusion spoken by approximately 4 million people in the Republic of South Africa. Within the South African context, standardization of terminology needs to contribute to the elevation of the status of a previously disadvantaged language to that of fully-fledged official language. In the case of Northern Sotho, apart from its direct impact on terminological development, any form of terminological activity therefore must contribute to terminological standardization, and within the broader sociolinguistic context, to language standardization, since Northern Sotho has not yet been fully standardized. This paper presents the results of a case study based on the compilation of a quadrilingual explanatory LSP dictionary for chemistry in order to illustrate that sound terminology practice is indeed possible in an environment where the terminological infrastructure is not ideal, and that it can contribute not only to terminology development and standardization as such, but also over a wider spectrum to standardization of an as yet only partially standardized language.

La reforma pombalina de la enseñanza: de la *Prosodia* de Bento Pereira al *Parvum Lexicon* de Pedro da Fonseca

Borges, Ana Margarida

(X)

🔥 6. Historical and Scholarly Lexicography and Etymology

Short Paper

The end of Jesuits' control over education in Portugal in 1759 and the consequent remodelling of Portuguese education led to new and promising procedures at social and state levels such as dismissal and nomination of teachers, syllabus planning, elaboration and fiscalization of didactic material. In fact, it is in this context of education reforms and of modification of the structure and customs of the Portuguese society that Marquês de Pombal forbids Bento Pereira's *Prosódias*, a group of dictionaries that supported the teaching of Latin and Portuguese. Therefore, the need to compose urgently a new dictionary that might answer the needs of Pombal's aims in relation to education and that could at the same time fulfil the capacities of school usage, emerges.

This new dictionary would be Pedro José da Fonseca's *Parvum Lexicon Latinum*, which would be concluded and published three years later, in 1762, under royal order. The simple idea of the usefulness of a little dictionary that made the learning of Latin and Portuguese easier represented the beginning of the modernization of the bilingual Latin-Portuguese lexicography that would, later on, allow for the improvement of the techniques used in the making of dictionaries.

The aim of this investigation is to establish a link between Bento Pereira's *Prosodia* and Pedro da Fonseca's *Parvum Lexicon*, by pointing out the main innovations in lexicography present in the nomenclature and in the structure of the articles.

Un diálogo implícito: la relación entre Joan Corominas y José Luis Pensado a través de su producción lexicográfica

Cotelo García, Rosalía



6. Historical and Scholarly Lexicography and Etymology

Student Paper

Our paper is part of a broader research into the profound change that transformed the Diccionario Crítico Etimológico de la Lengua Castellana (1954) by Joan Corominas, into the Diccionario Crítico Etimológico Castellano e Hispánico (1980) by Joan Corominas and José Antonio Pascual, the latter being a considerably more comprehensive and extensive edition. Our proposal stresses the importance of the implicit dialogue that Joan Corominas and Jose Luis Pensado kept through their lexicographic works. This dialogue would substantially improve the Diccionario Crítico Etimológico de la Lengua Castellana (1954). Thus, not only did Pensado include comments on this latter dictionary, but numerous corrections as well, in the Prologue of Catálogo de Voces y Frases Gallegas (1973), which he edited. Corominas assessed them, accepting most of the corrections and he introduced them in his new dictionary, the Diccionario Crítico Etimológico Castellano e Hispánico(1980). This huge lexicographic work arises our interest since most of its macro- and microestructural enlargement is based on a massive inclusion of galician entries-thanks to the editorial work of Pensado, actually. In consequence, this presentation seeks, firstly, to reflect the importance and consequences of this fruitful dialogue and, secondly, to vindicate the figure of Jose Luis Pensado in the Diccionario Crítico Etimológico Castellano e Hispánico, as well as Corominas' appreciation and recognition of his philological authority and erudition. Finally, we expect to highlight the undeniable productivity of scientific dialogue in the field of lexicography, since, as in any specialized area, it plays an essential role in the advance of modern research.

Velázquez de la Cadena y la lexicografía bilingüe inglés / español

Garriga, Cecilio; Gállego, Raquel

6. Historical and Scholarly Lexicography and Etymology

Short Paper

Mariano Fernández de la Cadena (Mexico City 1778 - New York 1860), professor at Columbia University, was the author of a bilingual English-Spanish dictionary of great prestige, which is subsequently reissued and revised even today. However, the dictionary is a relatively unknown work, primarily because bilingual dictionaries have not been the focal point of much attention by researchers, and secondly, because the dictionary's sphere of influence has been centered in the United States. If we survey the primary literature, we can see that the work is referred to inaccurately. At first glance it is clear that this is an innovative dictionary that is clearly rooted in the Spanish lexicographical tradition that emerged in the mid-19th century at a significant point in the revival of Spanish lexicography. Likewise, as a bilingual dictionary that aims to meet the needs of American students, it manages to escape the asphyxiating dominion exercised by the Royal Academy in the field of Spanish lexicography. In this study, the characteristics of *A Pronouncing Dictionary of the Spanish and English Languages* are examined in detail in terms of both their macrostructure and microstructure, and special attention is paid to how the sciences technical fields are treated lexically, as they constitute one of the realms most sensitive to revival during this period. All of this is duly contextualised within contemporary lexicographical trends.

Description of Loan Words in French School Dictionaries: Treatment of Words of Foreign Origin in Dictionnaire Hachette junior (2006) and Le Robert junior illustré (2005)

Gasiglia, Nathalie

6. Historical and Scholarly Lexicography and Etymology

Poster

French children learn to use dictionaries at the very beginning of their schooling. Between the ages of eight and twelve, they have access to general-purpose dictionaries which may deal with certain loan words. Our study analyses borrowings which are dealt with in a selection of this type of dictionaries: two French general dictionaries for cycle 3 which have substantial etymological content-Dictionnaire Hachette junior (2006) and Robert junior illustré (2005). The four leading general children's dictionaries for eight- to twelve-year-olds note between 116 and 619 borrowings from a selection of 4 to 52 languages. Like the dictionaries for cycle 2-students between the ages of five and height, they may provide information about the phonographic features of the borrowings indicated as such, but as cycle 3 children are supposed to read alone and be at an age when the thirst for new knowledge is very strong, it is logical that dictionaries designed for them should offer more substantial entries in terms of the nature and relative systematization of the information they provide. According to each dictionary's individual structure this information might be presented in a single entry zone-as in Larousse junior (2003)-or three zones-in Robert junior illustré and Dictionnaire Hachette junior. Like the number of zones used, associated information types also vary: the information given is most often metalinquistic-phonographic, lexical, morphological, etymological, etc.-and sometimes cultural. In this analysis of the treatment of anglicisms in French dictionaries for eight- to twelve-year-olds, I propose to build a typology of etymology associated information and to examine how and where this information is given in Dictionnaire Hachette junior and Robert junior illustré, the two dictionaries which have a consistent etymological approach: 619 loan words identified in the first one and 495 in the second one.

Le polirematiche nel TLIO: pratiche lessicografiche, dati e criteri di classificazione Giuliani, Mariafrancesca

🔥 6. Historical and Scholarly Lexicography and Etymology

Full Paper

This paper describes the data, the methodological problems and the directions, as well as the classification criteria involved in the lexicographic treatment of multiword expressions in the TLIO (*Tesoro della Lingua Italiana delle Origini*, cf. *www.ovi.cnr.it*). It is focused the importance of choosing to record and include multiword expressions into the microstructure of the entries, in order to show the semantic and syntactical interconnection binding free, recurrent and fixed combinations in the net of uses involving each data-base item. Particularly I describe and discuss the three level classification-

collocations, idioms, phrases-used to arrange the data-base cooccurences showing features of frequency or idiosyncratic semantic-syntactical structure. Some attention is paid to the definition of the idiomatic field drawn in the editing of a corpus based historical dictionary, often grounded on the decoding activities connected with the lexicographic description; finally I stress the contribution that linguistics and lexicography could get out of the collection and study of a high number of particular form-meaning pairs selected from historical documentation, especially if compared with similar modern lexical corpora.

GASTEREA: Digital Diachronic Thesaurus of Latin Food Words and their Heritage in European Languages Grigorieva, Alexandra; Hautala, Svetlana; Romanova, Natalia

6. Historical and Scholarly Lexicography and Etymology

Short Paper

Our international lexicographical project is set to assemble all classical Latin culinary words of surviving texts to create the first *Digital Diachronic Thesaurus of Latin Food Words and their Heritage in European Languages* in history, an interactive searchable database structure of culinary contexts for exploring the history of the culinary words of antiquitiy and their reception from the Middle Ages and Renaissance up to the present. Each classical Latin word describing food is to be supported by its etymology when possible-with relevant quotations in Ancient Greek for Ancient Greek loanwords-and, if that particular word has progeny in other old and modern, especially Romance European languages, it will be provided with links to derivatives and appropriate contexts. We will also strive to cover the majority of Medieval Latin food words in the same way during the second stage of the project.

The project is in its initial stage now but we would like to show our colleagues the preliminary digital structure of the Thesaurus that allows the display of the historical chains of shifting lexical forms-including dialectal when possible-and their meanings. Every food context in the chain would be provided with a translation and a short commentary in English and in Italian describing its historical, anthropological, cultural and culinary peculiarities. This frontier, interdisciplinary project covers the whole history of European languages and literatures. It should be able to bring to light the varied typology of European culinary vocabulary-something nobody has done before-and, at the same time, help to preserve the rich culinary heritage and diversity of European countries. We hope it becomes an invaluable tool for many Classical, Medieval and Renaissance scholars and other researchers engaged in Language, History, Food Studies and so on.

The role of Foclóir Gaeilge-Béarla Néill Uí Dhónaill in Irish language lexicography in the twentieth century Mac Amhlaigh, Liam

6. Historical and Scholarly Lexicography and Etymology

Poster

This paper sets out to chronicle the compilation and usage of the Foclóir Gaeilge-Béarla-or Irish-English Dictionary-by Niall Ó Dónaill, Tomás de Bhaldraithe and the lexicography team in An Gúm working on behalf of the Department of Education in the Republic of Ireland. As the primary modern dictionary of its time, its effect on the teaching and usage of the Irish language in the last quarter of the twentieth century is profound. This is especially the case in light of the fact that no update or amendment to it has ever been seen fit to be produced. Unlike the forthcoming English-Irish dictionary in motion under the auspices of Foras na Gaeilge-the government body responsible under Irish law for the promotion of the Irish language and Irish language organizations-and Lexicography MasterClass, there is no likelihood of any new Irish-English dictionary being produced in the near future. The evolution of the dictionary began as a development from the publication of the English-Irish Dictionary of Tomás de Bhaldraithe in 1959 when an equivalent resource for language users was desired from the opposite perspective-that of the Irish language user looking for the appropriate and most up-to-date English idiom for the words sought. The paper analyses the strengths and weaknesses of the dictionary together with the reasons that necessitated the production of the dictionary as it was. The paper represents a flavour of the ongoing research in the area of Irish language lexicography of the twentieth century, utilising, among other sources, the papers of Tomás de Bhaldraithe situated in University College Dublin's Cártlann na gCanúintí-Irish language dialect archive, the papers of Muiris Ó Droighneáin, one of Ireland's foremost grammatical consultants and the papers and archive of An Gúm, the Irish language publishing wing of the Department of Education.

Macro- and Microstructure Experiments in Minor Bilingual Dictionaries of XIX and XX century Marello, Carla; Tomatis, Marco

6. Historical and Scholarly Lexicography and Etymology

Short Paper

Two bilingual English and French and English and German dictionaries and two multilingual dictionaries dealing with English, French, German and Italian with a peculiar macro- and microstructure will be considered in order to highlight their efforts to spare space and to help foreign learners of such languages. The first dictionary-Williams Smith, A French Dictionary, on a plan entirely new (1814)tried to help English learners to reproduce the pronunciation of French words, the second-A.F Inglott Bey, A dictionary of English Homonyms pronouncing and explanatory translated into Italian and French (1899)-arranged homonyms in three languages and explained them, the third-Neues Universal-Wörtbuch der deutschen, englischen, französischen und italianischen Sprache (1856)-insisted on comparison among languages and the fourth-Max Bellows' Dictionary of German and English English and German (1912)-tried to have both sections English-German and German-English on the same page and to exploit different types to distinguish parts of speech plus masculine, feminine and neuter gender. The paper will explore suggestions for more innovative format in electronic bilingual dictionaries of the XXI century, since electronic dictionaries on Cd-Rom developed the search window, but did not venture to reinvent the electronic microstructure profile. During the XIX and XX century printers and lexicographers reflected upon improvements in printing layout above all when they aimed to meet the claims of middle-class buyers, asking for effective, pocket-size, not too expensive lexicographic tools.

Le programme TLF-Étym: apports récents de l'étymologie comparée-reconstruction Petrequin, Gilles; Monda Andronache, Marta

6. Historical and Scholarly Lexicography and Etymology

Short Paper

Our topic is the French hereditary vocabulary considered in a new theoretical approach and focuses on the theme "Historical and learned Lexicography and Etymology". From the point of view of the classical etymology any lexeme must find its origin in a graphic form. Therefore, there is nowadays a strong consensus among the specialists of the Romanic studies to revise this "classical" and philological lexicographic practice that puts the graphic form at the centre of the theory, which generates basic contradictions. Recognizing the oral form of a hereditary lexeme and renouncing the "graphic centrist" conception in the treatment of the hereditary vocabulary appears as an obvious necessity in the daily practice of the lexicographer. Moving away the "classical" method of the Romanic etymology, we propose to apply the system of the historical and comparative grammar to the field of the French etymology to reconstruct, by comparing different oral forms from Romanic languages, the oral form of the proto-language. Our submission presents three examples of etymological notes/headwords on the hereditary vocabulary developed and published by the program TLF-Étym of the linguistic laboratory ATILF (Analyse et traitement informatisé de la langue française; CNRS/Nancy-Université, France). These examples will allow us to demonstrate to what extent the practice of the etymology of the French hereditary word pool depends on the progress of the Roman etymology, with which it should go hand in hand from now on.

De la 1re à la 2e édition du Dictionnaire de l'Académie française: marques diastratiques et diaphasiques Pouteaux, Marie-Alix; Dagenais, Louise



6. Historical and Scholarly Lexicography and Etymology

Student Paper

In the history of French dictionaries, the second edition (1718) of the French Academy's Dictionnaire (hereafter ACA2) has generally been perceived as a bare alphabetical re-arrangement of the first edition, published in 1694 (ACA1), in which lexical entries were morphosemantically grouped under their primary root word. However, ACA2's preface and title (Nouveau dictionnaire) suggests that it underwent a more important revision than what has been believed. This research brings to light the significant progress which ACA2 represents in comparison with ACA1. In the first part, the various aspects of the dictionary microstructure of the letter / headwords are compared with each other. The second part is devoted to the analysis of the sociolinguistic marking on the basis of the diastratic and diaphasic usage marks, i.e. bas, populaire, peuple and familier. The results that arise from this study are, firstly, that 57% of the lexical units from the I corpus that are common to both editions are reworked in ACA2 and, secondly, the study shows that 47 to 83% of the lexical units tagged bas,

populaire, people and/or familier were not included in ACA1. We then proceed to demonstrate to what extent the French Academy 1718 Nouveau Dictionnaire constitutes a new edition and not just an alphabetic reprint of the first edition.

L'informatisation du FEW: attentes et modelisation

Renders, Pascale; Nissille, Christelle

....

6. Historical and Scholarly Lexicography and Etymology

Short Paper

The computerisation project of the *Französisches Etymologisches Wörterbuch* aims at taking this reference dictionary on Romance linguistics out of its current state of under-utilization resulting from the complexity of its structures. In view of this, we have submitted on September 2007 a questionnaire designed to better understand the wishes and common practices of FEW users. In the present paper, we examine the first results of the survey and its impact on the future electronic modelisation of the FEW by focusing on cross-searching fields. Implicit information, such as dates, regionalisms or suffixes cannot be found automatically, except by using external tools, that must therefore be taken into account in the computerisation. The solutions considered here do not mean that "classic reading" will become obsolete. Nevertheless, we hope that we will be able to give users new ways to get to the dictionary thus allowing a new and more efficient use of the FEW...

El tratamiento de los números en el diccionario

Rodríguez Ortiz, Francesc; Garriga Escribano, Cecilio

6. Historical and Scholarly Lexicography and Etymology

Poster

The definition of numbers involves a complexity that is often overlooked in dictionaries. Their double nature-both grammatical and semantic-means that on one hand they constitute part of a formal language-ex. *arithmetic*, and on the other, their morphological behaviour means they are no different from any other types of words. This complexity is accentuated by the way they can be considered either nouns or adjectives. On the other hand, words like *one*, *two*, *three*, etc. present the distinction that they can be written in two different ways: *one* / 1, *two* / 2, etc., i.e. either using a linguistic or an arithmetical sign. Additionally, their different forms, dependent on whether they are cardinals, ordinals, fractions, multiplicative, distributive, collective, etc., involve differentiated lexical forms. Numbers also frequently possess figurative meanings or appear in an abundance of set phrases. Dictionaries have dealt with this problem in a variety of different ways. If we look at the design of Spanish dictionaries since the 18th century, we find a certain amount of vacillation that persists until the matter was more firmly established in the 20th century. Nor were there any major differences between dictionaries in different languages. This study presents the state of the question on the basis of an examination of popular Spanish general dictionaries, and proposes certain principles that could improve the coherence of dictionaries in the way they deal with this class of problem.

Aspectos gramaticales en la macro y microestructura de un diccionario bilingüe novohispano Romero Rangel, Laura; Mora-Bustos, Armando

6. Historical and Scholarly Lexicography and Etymology

Short Paper

The purpose of this paper is to expose and explain certain grammatical aspects in one of the most important bilingual dictionaries in La Nueva España on sixteenth century: the *Vocabulario castellano y mexicano y mexicano y castellano* (1571), elaborated by prior fray Alonso de Molina. Currently in bilingual lexicography, there are different criteria to codify syntactic or grammatical information in lexicographical theory. Yet in the middle of the 16th century, these criteria did not exist. It was the linguistic sensibility of Molina that became the only way to express the grammatical functions of the lexical and phraseological units. Our intention with this presentation is to demonstrate the way that Molina, although not a lexicographer, was able to codify the following information:

- 1. marcas gramaticales -grammatical markers, for example: mejor nombre comparativo, mejor adverbio comparativo or he adverbio para demostrar, in Spanish entry, and agora tiempo presente, axca, axcan. Aduer, in nahuatl part;
- contornos sintácticos-syntactic plan/outline, such as: Abituar a alguno en agluna cosa, Cutir una vasija con otra, or Chamuscarse algo; and

3. ematización de compuestos y frases, for example: Abrego viento, Higas dar, Hambrear auer hambre, Hambre hauer o tener hambre de cualquier cosa, Harona bestia, Haldas poner en cinta or Achacoso ser.

This is only a small sample of how to lexicographically handle special types of syntactic information.

Le DÉCT (*Dictionnaire Électronique de Chrétien de Troyes*): un modèle pour la lexicographie d'aujourd'hui? Souvay Gilles, Pierre Kunstmann

6. Historical and Scholarly Lexicography and Etymology

Software Demo

The DECT is an example of today's lexicographic practice. Its realization is completely computerized from the input to the on-lining. It calls on modern concepts of data encoding (XML) and diffusion-free access on the Web. The DECT is not just a dictionary searchable from the entries. It is in fact a real lexicographic tool made up of an annotated textual base-lemma and part of speech-with the manuscript's image, and the lexicon resulting from the texts analysis. It can be consulted in a traditional way-display of a page, of a verse, of an article...-or through specialized search forms, for instance, it is possible to look for co-occurring words in the texts-lemma aimer before an adverb, or to make a multi-criteria query in the lexicon-search for a word in a verb's definition. Moreover, it is always possible for the user to go from the lexicon to the texts and vice versa. The on-line base can be accessed at http://www.atilf.fr/dect. (French and English). The DECT's computerized component is built on a platform developed at the ATILF for historical linguistics projects. The same tools allow the consultation of other lexicographic projects, about ten instancings. The DECT contributed, for a large part, to the platform development and constitutes, for it, the most successful instancing.

${\it Vulgar}$ and ${\it Popular}$ in Johnson, Webster and the OED Wild, Kate

6. Historical and Scholarly Lexicography and Etymology

Student Paper

The use of restrictive labels is one of the most subjective features of modern lexicography, and several studies have shown that dictionaries do not always agree in their application of, for example, *colloquial* and *informal*. Labels are also a problematic feature of pre-20th century dictionaries, which did not provide lists or explanations of the labels they used. The purpose of this paper is to analyse the development of two labels-*vulgar* and *popular*-in Johnson's (1755) *A Dictionary of the English Language*, Webster's (1828) *An American Dictionary of the English Language*, and the first edition of the *Oxford English Dictionary* (1884-1933)-in order to consider how their meanings and connotations have changed, and what their use can tell us about the relative prescriptivism of the three dictionaries.

Papel de los diccionarios de colocaciones en la enseñanza de español como L2

Alonso Ramos, Margarita

7. Dictionary Use

Full Paper

It is generally acknowledged within the Spanish as second language (SSL) community that collocations need to be taught and that collocation dictionaries are useful. Nevertheless, no one has carried out yet any experimental study to investigate what kind of collocation information must be included into a dictionary and how to encode it for a user to take full advantage of it. We describe the results obtained from a small experiment in the use of collocation dictionaries in the teaching of SSL. More precisely, the goal tof this experiment is to verify whether the inclusion of semantic and syntactic information on collocations into the dictionary as well as examples of usage could correlate with a better performance on the part of learners.

This is namely the premise underlying the *Diccionario de colocaciones del español* (DiCE). DiCE is based on the Explanatory and Combinatorial Lexicology (Mel'cuk et al. 1995), where collocations are assigned semantic labels and syntactic tags -lexical functions. In order to weigh up how useful this information is, we had to compare the DiCE with another dictionary which did not include this information: the only dictionary that has been published in Spanish which deals with collocations is the *Diccionario combinatorio práctico* (DCP, Bosque 2006).

The experiment was conducted on 25 learners of Spanish and 5 native speakers. Its goal was to evaluate whether the users of the dictionaries had better results with the dictionary that included

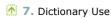
semantic and syntactic information of each collocation. Since we needed to know their previous knowledge, we decided to organize the test according to three different criteria:

- 1. without any collocation dictionary;
- 2. with the DCP, and
- 3. with the DiCE.

On the one hand, the results of the experiment are positive but, on the other, worrying. Positive because they confirm our premise: in general, students perform better when the dictionary includes semantic and syntactic information on collocations, and worrying because they show that in some cases, the performance of the students decreases when they use the dictionaries mentioned above. Further, more extensive studies are needed to investigate this phenomenon.

Frequency in Learners' Dictionaries

Bogaards, Paul



Short Paper

The learners' dictionaries that exist for English all contain a restricted number of items. The vocabulary that is described in these dictionaries is selected on the basis of frequency of appearance in English. A far more limited number of items are marked as the most important ones, as these that all students should know at some time, because they constitute the lexical core of the language. The marking of high frequency is done in different ways in the five learners' dictionaries. The data provided are not always very useful and are sometimes inconsistent from dictionary to dictionary. An analysis is made of some samples taken from the five learners' dictionaries of English and the relevance of different types of frequency information is discussed.

United in Diversity: Dutch Historical Dictionaries Online

Depuydt, Katrien; De Does, Jesse

7. Dictionary Use

Software Demo

The Integrated Language Database of Dutch (ILD) is a project of the Institute for Dutch Lexicology in Leiden, which integrates corpora, computational lexica and dictionaries describing the Dutch language from ca. 500 until the present. In 2007, the dictionary component was released, already containing two major historical dictionaries of Dutch, the Woordenboek der Nederlandsche Taal (WNT, Dictionary of the Dutch Language, 1500-1976) and the Vroegmiddelnederlands Woordenboek (VMNW, Dictionary of Early Middle Dutch, 1200-1300). When, by 2009, the Middelnederlandsch Woordenboek (MNW, Dictionary of Middle Dutch, ~1250 - 1550) and the Oudnederlands Woordenboek ('ONW', Dictionary of Old Dutch, a current project at INL, to be finished in 2008, ca. 500-1200) will have been added, researchers of Dutch will have access to dictionaries covering the complete history of the Dutch language. The choice of a single application, integrating the dictionaries so that a user might query one or more dictionaries simultaneously, was a logical step because of the complementary nature of the dictionaries. The challenge was not only providing the user with optimal access to the dictionary information, but also doing so without compromising the uniqueness of each individual dictionary. We sketch the principles underlying the application.

Noun and Verb Codes in Pedagogical Dictionaries of English: User-friendliness RevisitedDziemianko, Anna

7. Dictionary Use

Short Paper

The aim of the present paper is to assess the user-friendliness of noun and verb coding systems in pedagogical dictionaries of English, measured by the frequency with which relevant information properly used in a productive task is located in codes. The influence of the following independent variables on the user-friendliness of codes is studied: the degree of syntactic congruity between Polish lexical items and English headwords, the form of codes, the grammatical category of headwords and the level of dictionary users' proficiency in English. To investigate the influence of the form of codes on their user-friendliness, codes in noun and verb entries were divided into mainstream-referring to formal categories, transparent and prevalent in pedagogical dictionaries, and alternative-which, used very sparingly in today's dictionaries, include reference to sentence functions-verbs-or many quite

opaque symbols-nouns. Conclusions are drawn on the basis of an experiment in which almost 900 Polish subjects, advanced and intermediate in English, were involved in a translation task in which they had to use English noun and verb entries compiled for the purpose of the study. The results show that differences in grammar between Polish and English did not affect the consultation of either noun or verb codes. Strangely enough, alternative, and seemingly more demanding codes were strongly favored by the intermediate subjects, and-in the case of verbs-also the advanced ones. The part of speech played a very significant role at the higher level of proficiency, but was not important for the reference to codes by the less advanced. Finally, the higher level of proficiency in English made the subjects appreciate codes more fully, which may be seen as an argument for maintaining the over 70-year tradition of encoding syntactic information in pedagogical dictionaries of English.

Teaching the Systematic Dictionary Use as a Strategy for Accuracy and Confidence Building Kambaki-Vougioukli, Penelope

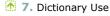


Short Paper

This is a longitudinal study, which started in 2004 and ended in 2005. There participated sixteen highschool pupils-same number of boys and girls, aged 13-15, of similar socioeconomic background, whose MT is Turkish but living in Thrace, Greece and attending Greek State Schools rather than minority Public Schools. The fact is that we expected to have more subjects but, unfortunately, we had to exclude a lot of pupils due to a number of reasons such as differences in the socioeconomic level of the families, gender availability-having more male than female pupils, negative attitude towards the research, etc. What we are investigating is whether and to what extent the systematic use of both monolingual English dictionaries and bilingual Greek-English and English-Greek dictionaries could possibly result in a better reading comprehension and, in the long run, in an improvement and enrichment of their English vocabulary and, to a lesser extent, in Greek. Our aim is to reinforce their general linguistic competence and performance but also their strategic competence by encouraging them to use dictionaries when working at home, too. Furthermore, we are measuring their confidence levels before and after using dictionaries, at certain intervals over the whole period of the experiment. All the participants were given individualised instruction on dictionary use in pair and group work at certain intervals over the whole period of the experiment, too. It is important to notice that we are not really evaluating "certain" dictionaries, it is rather unrealistic as their resources are rather poor; nevertheless, we are trying to exploit what we really have at our disposal, that specific time. The results justified our expectations as most the students that collaborated seem to be very comfortable with dictionary use and confident with the information they expect to find there.

Improving Dictionaries

Kernerman, Ari



Short Paper

Although printed dictionaries have reached a high level of sophistication, there is still much to be improved in order to enhance their usefulness. Prefaces, especially in learners' dictionaries, are not written for users or actual learners, but rather for their teachers, for other lexicographers or for reviewers. For example, Prefaces in learners' dictionaries explain such things as the use of word corpora, the character of the dictionary, the philosophy behind the dictionary, how the dictionary was written, what is different in each particular edition, etc. Interesting, but not helpful information for users. Though intended to be used universally, these dictionaries are culturally biased. Their British culture is irrelevant to the billion learners of English who live in non-English-speaking countries, and need locally or neutrally-oriented dictionaries to help them to communicate with people in other non-English-speaking countries. And the one-size-fits-all principle of monolingual learners' dictionaries does not replace the need to provide mother tongue translation. Many publishers keep adding information to the new editions, much of which is not helpful, reduces the dictionary's efficiency, and does not increase the user's knowledge. On top of that, the absence of a system of lexicography standards makes it difficult for users to refer to more than one dictionary. Giving preference to corporadetermined frequency over the didactic value of presenting basic meanings first is a step backward, not forward. Besides, too much space is unnecessarily devoted to familiar words, at the expense of less familiar words. These, and other deficiencies of our modern dictionaries-including bilingual, native speakers' and specialized dictionaries-are discussed, with suggestions for rectifying them.

Teaching Dictionary-using Skills for Online Dictionaries-An Attempt at a Theoretical Framework for South Africa

Klein, Juliane

7. Dictionary Use

Student Paper

The aim of this paper is to illustrate a theoretical approach to teach dictionary-using skills in South Africa. As the focus is on online dictionaries, only dictionary-using skills will be discussed. Teaching dictionary-using skills in a linguistically heterogeneous society, which has not yet developed a fully functional dictionary culture for all languages, is a difficult task. Not only must the different languages-e.g. conjunctively written languages and disjunctively written languages-be taken into account, but also the different user groups ranging from pupils/university students to ordinary people who want to use a dictionary have to be considered. Although the dictionary users are not a homogeneous group, the aim of teaching dictionary-using skills is the same for all groups: achieving a confident and successful use of dictionaries in the short term and creating a fully developed dictionary culture that includes all the languages which are official in South Africa in the long term. The teaching of dictionary-using skills could be divided into four stages:

- 1. teaching about dictionaries,
- 2. teaching basic skills to access dictionaries
- 3. teaching look-up strategies,
- 4. teaching strategies to decode the information found in the definition given by the dictionary.

Dictionary-using skills should be taught as early as possible in schools, and this teaching should be continued throughout the whole education process, i.e. it should not be taught as a single module, but rather as language methodology. In tertiary education institutions, dictionary-using skills could be integrated into academic literacy modules or taught in separate short language modules. Teaching dictionary-using skills to everybody else will be more difficult, as those who have finished their formal education cannot be reached as easily as pupils or university students. This group will mainly be taught through the dictionaries themselves. Teaching dictionary-using skills to people through dictionaries implies that the dictionaries must be self-explanatory, which implies that the user interface and all instructions should be available in all the languages that the dictionary covers and not only in English. In addition to that, the dictionary should ideally be accompanied by a user manual in all languages the dictionary covers.

Can Dictionary Skills Be Taught? The Effectiveness of Lexicographic Training for Primary-school-level Polish Learners of English

Lew, Robert; Galas, Katarzyna

7. Dictionary Use

Full Paper

In the present paper we examine the question of whether dictionary reference skills can be taught effectively in the classroom. To this end, we test the reference skills of a group of Polish primary-school students attending English classes twice: prior to and following a 12-session specially-designed training program. Despite the subjects' high confidence in their reference skills reported in the accompanying questionnaire, they performed rather poorly on the pre-test. Following a training program, the performance improves substantially and significantly more than in a matched control group. We conclude that a dictionary skills training program may be effective in teaching language learners at this level to use dictionaries more effectively, though different skills benefit to different degrees.

Bringing Bilingual Dictionaries in from the Cold: Challenging Negative Perceptions and Practices in English Language Teaching

Mandalios, Jane

7. Dictionary Use

Full Paper

This paper deals with English language teaching (ELT) and learning. It considers the presenter's research into the use of bilingual dictionaries in those cases where English is the second/foreign language. The research was carried out amongst non-native speaker students and teachers, and also amongst teachers who were native speakers of English in an English-medium university in the United Arab Emirates. It showed that, in contrast to the teaching of other foreign languages, bilingual

dictionaries are generally negatively viewed by ELT theorists and teachers. Yet, after a careful scrutiny of both the lexicography and literature used in ELT you realize that such a view is based on unsubstantiated opinions or questionable research. The study also shows that bilingual dictionaries are almost unanimously considered helpful by learners, yet their preferences are usually ignored or discouraged by teachers, many of whom do not speak the first language of their students, and who feel pressurized to follow the English-only approach that has dominated ELT for the last 40 years (Phillipson 1992; Auberbach 1993). The students in the study exhibited poor dictionary skills, and little understanding of how efficient the use of bilingual dictionaries could be to enrich their receptive and productive vocabulary skills. A small action research component of the study indicated that these skills can be greatly improved by structured bilingual dictionary instruction. The presenter proposes that the findings of the study constitute evidence of a serious imbalance of power within ELT which can be defined as pedagogic imperialism, and calls for a critical reappraisal of both the role of bilingual dictionaries and the use of the native language when teaching English. Closer ties need to be established between the fields of lexicography and ELT, particularly in contexts where the theory and practice of teaching is dominated by native speakers who do not speak the first language of the learners.

Looking Up "Hard Words" for a Production Test: A Comparative Study of the NOAD, MEDAL, AHD, and MW Collegiate Dictionaries

McCreary, Don R.

7. Dictionary Use

Short Paper

We test this hypothesis: The *New Oxford American Dictionary* (NOAD), MW, AHD, and MEDAL equally meet the needs of American college students when they look up a *hard word*. On a production task, writing the word in an appropriate sentence, NOAD users scored much higher than the other three groups on every *hard word*, with only one exception per user. The *Macmillan English Dictionary for Advanced Learners* (MEDAL) users scored higher than the users of the *Merriam Webster's Collegiate Dictionary*, 11th Edition (MW) or users of the *American Heritage Dictionary*, 2nd Edition (AHD), another collegiate desk dictionary. NOAD has several advantages over the other collegiate dictionaries, including microstructure and vocabulary coverage. Unfortunately, overall coverage of *hard words* is problematic in MEDAL, since it is intended for non-natives. MW users were hampered by their tendency to choose the first sense in the entry, which is the oldest historical sense in MW. This also applies to AHD. This suggests that American college students might consider buying NOAD for its usability and its vocabulary coverage.

Giving Them What They Want: Search Strategies for Electronic DictionariesMechura, Michal Boleslav



Short Paper

This paper deals with how humans search electronic dictionaries. It raises the point that users often make dictionary searches with misspellings, with inflected words copied and pasted from elsewhere, with complete sentences or fragments thereof, and with other kinds of low-quality input, and suggests methods for dealing with such phenomena in a preemptive manner. The issues addressed include searching with inflections, dealing with multi-word items, misspelling detection and text normalization. Additionally, the value of log files is emphasized as a source of information on user behaviour.

Adverb Use in EFL Student Writing: From Learner Dictionary to Text Production Philip, Gill

7. Dictionary Use

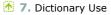
Short Paper

Adverbs, especially those occurring in *adverb+adjective* collocations, play a central role in the language that advanced learners are expected to produce in their argumentative writing. Submodifying adverbs of degree such as *closely*, *deeply*, *strongly* and *widely*, however, have been identified as being problematic for learners of English: Italian learners over-use *very* and *really* to the virtual exclusion of any other adverb (Philip 2007). This situation is due in part to the EFL curriculum, but monolingual and bilingual learner's dictionaries appear to do little to address the issue. This presentation examines the way in which lexical adverbs of degree are treated in the five major English dictionaries for advanced

learners (CALD, COBUILD, LDOCE, MED and OALD). It also evaluates the way these same forms are treated in four bilingual dictionaries specifically aimed at Italian learners of English (Longman, Oxford Study, Rizzoli-Laroussse, and Oxford-Paravia). The analysis reveals that these dictionaries do little or nothing to help students expand their working knowledge of adverbs of degree. In general, the presentation of lexical adverbs is regarded to be subservient to the adjectives from which they are derived. The information boxes which most modern learner's dictionaries include seem to focus on elementary matters of grammar and word choice rather than on the collocation of these polysemous, metaphorically-motivated language items. The presentation concludes by suggesting some ways in which monolingual and bilingual learners' dictionaries might modify their treatment of lexical adverbs in order to enable students to identify and use alternatives to very, really and a lot.

The Electronic Dictionary in the Language Classroom: The Views of Language Learners and Teachers

Ronald, James; Ozawa, Shinya



Poster

The pocket electronic dictionary (PED) has the potential to be a powerful language learning tool. At the same time, it may be seen as an obstacle to communication, a waste of classroom time, and a source of conflict between foreign-language learners and the teachers. This presentation will report an indepth survey of three sets of people influenced by the widespread presence and use of the PED in the classroom: foreign-language students, teachers who share the native language of the students, and teachers who are native speakers of the target language. The survey, which takes into account the beliefs, attitudes, and expectations of Japanese learners of English and of their teachers regarding the PED, revealed important differences in their opinions about how and when the dictionary should be used, in the effect of dictionary use on foreign language vocabulary development, and regarding users' needs for training or guidance in the use of electronic dictionaries. The presentation will also recommend means by which understanding of these differing perspectives may help both language learners and teachers make the most of the potential of the electronic dictionary.

REDES. Diccionario combinatorio del español contemporáneo

Almarza Acedo, Nieves; Lozano Ramírez de Arellano, Yolanda

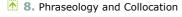
8. Phraseology and Collocation

Poster

REDES. Diccionario combinatorio del español contemporáneo is an attempt to reflect on the lexical restrictions and to analyse the structure of the language. Through different examples we want to go deeply especially into the utility of this dictionary for any speaker of Spanish and mainly for the students of the language. This dictionary demonstrates that it is necessary to know how the words are combined to express ourselves with accuracy.

Propuesta de anotación semántica para una base de datos paremiológica

Alonso Pérez-Ávila, Elena



Student Paper

An electronic tool, such as an on line multilingual paremiological database, that would enable researchers or translators to search paremiological units of many languages and manipulate the information stored about them more easily, would greatly benefit the field of paremiology.

This paper deals with how the information provided by SpanishWordNet and MultiWordnet may be used to tag semantically paremiological units in Spanish and Italian within the database. In other words, tags related to the WordNet ontology are attached to each term in the proverb in order to provide more information about the domain that the proverb belongs to. We propose this annotation as a methodology to classify paremiological units that can be shared by different linguistic communities since it is based on an already widely used lexical resource developed in many languages: WordNet.

Unfortunately, it is not possible to tag the proverb as a whole unit due to its particular features: the meaning of the whole proverb cannot be easily derived from the meanings of its separate components. At the moment we are trying to supply the database with as much information as possible on the semantics of the components of the proverbs and the relations, such as hyponymy, that those parts present with the rest of the lexicon.

From Dictionary to Phrasebook?

Granger, Sylviane; Paquot, Magali



Full Paper

Language is characterized by a large number of conventionalized phrases which, unlike idioms, are largely regular, both semantically and syntactically. Biber et al. (1999) call these phrases *lexical bundles* and highlight the key role they play both in spoken and written discourse. In spite of their high frequency, these types of phrase have not yet received the place they deserve in dictionaries. In this article, we describe how they are integrated into monolingual learners' dictionaries of English and English-French bilingual dictionaries. The description shows that the presentation of these phrases is largely based on intuition and fails to reflect authentic usage as attested by corpus investigation. We make a plea for a more rigorous-corpus-based-integration of these phrases and illustrate our approach with a fully corpus-based section devoted to English for Academic Purposes (EAP), functions that has been integrated as a middle section in the new edition of the *Macmillan English Dictionary for Advanced Learners*.

Collocational False Friends: Description and Treatment in Bilingual Dictionaries

Heid, Ulrich; Prinsloo, Danie

8. Phraseology and Collocation

Short Paper

Our starting point is that of translation equivalents: true friends in their use as individual lexical items often become false friends in collocations. It is the duty of the lexicographer to guide the user, especially in learners' dictionaries aimed at productive-encoding-use, in forming correct collocations and in warning the user of false friend cases. Our arguments are based on evidence from large newspaper corpora as well as on internet research. We will present several lexicographic presentation devices from printed dictionaries that allow lexicographers to warn users about false friend collocations. The study will be limited to false friend relations in general bilingual dictionaries, mainly for German, Dutch and Afrikaans. The compilation of dictionaries for false friends lies beyond the scope of this paper. We adopt a lexicographic notion of collocation, here, as used for example by the *Oxford Collocation Dictionary for Students of English* (2002). We use Hausmann's (2004) terms-base and collocate-to denote the elements of collocations. Klégr (2006) transfers the notion of false friends from single words to collocations and classifies the relevant cases according to categories known from translation theory. We propose the following simple arrangement of false friend collocations, inspired by the concept's basic principles:

- 1. word combination-lexical (co-)selection: if true friend single word equivalents exist in a language pair, we consider collocations as false friends where the cooccurrence of the two single word true friends is impossible in a given language;
- morphosyntactic preferences: if true friend single word equivalents exist in a language pair, we consider collocations false friends where the languages differ with respect to morphosyntactic preferences, individual readings being equivalent;
- 3. differences with respect to usage domains.

Analysis of Collocations in Russian: Corpus vs Dictionary

Khokhlova, Maria

8. Phraseology and Collocation

Student Paper

The paper discusses the results of an experiment in collocation extraction in a corpus of Russian texts. The data obtained is compared to the data given for set expressions in modern Russian dictionaries in order to analyze from the standpoint of traditional lexicography what kind of phrases can be received by such an approach. The paper also explores the role of statistical measures for extracting collocations in Russian..

The Lemmatisation of Lexically Variable Idioms: The Case of Italian-English Dictionaries Mulhall, Chris

8. Phraseology and Collocation

Student Paper

The choice of a suitable point of entry for an idiomatic expression is one of the most complex tasks a lexicographer faces throughout the compilation of a dictionary. This is further exacerbated by the possibility of lexical variation in certain expressions. This paper analyses twenty idioms with variable verbs (ten English / ten Italian) and twenty idioms with variable nouns (ten English / ten Italian) across six bilingual Italian-English dictionaries, *Il Ragazzini* (ZIR) (2006), *Hoepli Grande Dizionario di Inglese* (HGDI) (2003), *Collins Sansoni Italian Dictionary* (CSID) (2003), *Oxford-Paravia Italian Dictionary* (OXID) (2001), *Il Sansoni Inglese* (ISI) (2006) and *Hazon Garzanti Inglese* (HGI) (2006). The analysis highlights a number of problems in the treatment of lexically variable idioms. Firstly, bilingual Italian-English dictionaries do not have a definitive approach to dealing with the problem of lexical variation. Secondly, the consistency and comprehensiveness in the coverage of lexical alternatives varies significantly both within and across the Italian-English and English-Italian sections of dictionaries. The totality of such differences suggests that a more systematic approach is required in order to achieve a greater consistency in the recording of the variable constituents of idioms.

Proyecto para la redacción de un diccionario de locuciones del español

Penadés Martínez, Inmaculada



Poster

The idea of creating a dictionary of Spanish idioms originates in the verification that currently there is no dictionary that solely includes these kinds of phraseology units, in contrast to other publications that compile other types of complex units, such as popular sayings. There are other reasons for the convenience of the project, more concretely, the deficient lexicographic treatment given to the assigning of grammar marks up to this day. This deficiency becomes apparent also in the assigning of syntagmatic combinatory, as well as the diastratic and diaphasic markings for idioms in Spanish phraseology dictionaries. The aforementioned dictionary will be onomasiological, semasiological, and will include a synonym and antonym thesaurus.

A Comparative Analysis of Definitions of Phrasal Verbs in Monolingual General-purpose Dictionaries for Native Speakers of American and British English

Perdek, Magdalena

8. Phraseology and Collocation

Short Paper

This paper is an attempt to analyze the definitions of phrasal verbs in monolingual general-purpose dictionaries for native speakers of English. Four dictionaries from Great Britain and four from the USA published in the last decade provide material for the study which includes a total of 100 phrasal verbs. Bearing in mind the specific semantic load of phrasal verbs, their limitation as to the choice of objects as well as the fact that they are commonly used, this study aims at finding whether there exist significant differences in describing phrasal verbs on both sides of the Atlantic. Three aspects are analyzed in particular: word choice with emphasis on the occurrence of difficult, very formal and rarely used words; precision in rendering the meaning, and inclusion of objects typical of a given sense of a phrasal verb. The analysis reveals that there are certain areas of correlation but also points of differences, not only between the two lexicographic traditions but within each of them separately.

Inclusión de los papeles semánticos de FrameNet en DiCE

Prieto González, Sabela

8. Phraseology and Collocation

Student Paper

The aim of our project is to enrich the actantial information of the *Diccionario de Colocaciones del Español* (DiCE) with labels about semantic roles. Since there are other projects which follow this line of research-such as FrameNet, we decided to include the existing information in the DICE. Although our database focuses on collocations, it also identifies their actants and it compiles a wide-ranging corpus of predicates with their arguments. Therefore, the entry for each lexical unit contains the proposicional form or argumental structure where a semantic description of the actants is given. For instance, in the entry for the noun *ira* we find: *ira de individuo X contra individuo Y a causa del hecho Z*. In this way, actants are also described semantically: *ira* is felt by a person against another person because of something. We are trying to add more semantical information to DICE, by linking the actants of each

lemma with the core elements of the relating frame, the same as FrameNet does. Taking up again the same example, the noun *anger* is set into the frame «Emotion_directed» and it presents four core elements: *experiencer*, *expressor*, *stimulus*, *topic*. These nuclear elements can be related to the actants that appear in the DICE, giving the dictionary some detailed semantic information, regarding not only the lemma but also the elements relating to this lemma. This process of connection will allow us to label the corpus compiled in the DICE to make the most of the data.

Colocaciones léxicas en diccionarios generales monolingües del español

Romero Aguilera, Laura



Poster

The purpose of this paper is to describe the way some of the Spanish general monolingual dictionaries published during the last twelve years have dealt with lexical collocations, that is, those combinations of words that present certain combinatorial restrictions in the norm, basically semantic restrictions, imposed by usage (Corpas 1996). These have been the analyzed dictionaries: *Diccionario Salamanca de la lengua española*, directed by Juan Gutiérrez (1996); *Diccionario del español actual*, by Manuel Seco, Olimpia Andrés y Gabino Ramos (1999); RAE's *Diccionario de la Lengua Española* (2001); and *Gran diccionario de uso del español actual. Basado en el Corpus Cumbre*, directed by Aquilino Sánchez (2001). We have based our research on a corpus of 52 lexical collocations, which has been built on the analysis of the subentries starting with *b* in the chosen dictionaries. After that, we have looked up the entries corresponding to each element that constitutes the collocation, in order to know if these dictionaries account for those same combinations in other parts of the lexicographical article. The analysis of the lexicographic information has focused on our aspects: a) the preliminary pages of each dictionary; b) the position of collocations in the lexicographic article; c) the inclusion of these units in a given article; and d) the grammatical category.

Una bella esperienza, una buona prova. A Corpus Analysis of Purely Evaluative Adjectives in Italian Russo, Irene

8. Phraseology and Collocation

Short Paper

It is questionable how much pragmatic information should be included in a dictionary entry. In a native-speaker's dictionary such information is considered unnecessary, but nevertheless, a certain amount of it could be included as multiword expressions-fixed and semi-fixed-that are regarded as holistic units rather than compositional strings. In this work a corpus analysis of two purely evaluative adjectives in Italian-bello, buono-will shed light on substitutability among them in noun phrases. Mutual Information (Church & Hanks1990) as a measure to compare and contrast the distribution of words in context highlights nouns for which bello and buono are interchangeable in NPs. We propose to manage adjectival polysemy clustering word senses according to similar evaluative functions. A dictionary entry for bello can be partially structured on the base of its strong similarity with buono in NPs contexts: bello and buono usages are informed by evaluative attitudes displayed by speakers.

From Subdomains and Parameters to Collocational Patterns: On the Analysis of Swedish Medical Collocations Sköldberg, Emma; Toporowska Gronostaj, Maria

8. Phraseology and Collocation

Full Paper

This paper presents a study on Swedish collocations in an electronic medical lexicon, currently under construction at the University of Gothenburg, Department of Swedish Language. There are two strands discussed in the paper. The first one is about a knowledge-based, onomasiological, approach to detecting and analysing medical collocations and their patterning. The second one deals with the representation of these collocations in both a general lexicon module and a collocational lexicon module. In the latter module, there are some advanced search options made available which enable selective access to the content of the lexicon. It is assumed that the onomasiological approach to the analysis of medical collocations complements the semasiological one and that the fusion of the two paves the way for a more consistent and exhaustive description of medical collocations and their patterns.

Aspectos de fraseografía bilingüe español-alemán: la equivalencia frente a la definición Torrent-Lenzen, Aina

8. Phraseology and Collocation

Short Paper

Aspects of bilingual phraseography Spanish-German-phraseological equivalence versus definition. In my paper I would like to discuss some problematic aspects of Spanish-German phraseography which frequently arise for both phraseologists and dictionary users. The document will focus, among other things, on the problems conveyed by the phraseological equivalents, the treatment that the contextual and partial phraseological equivalents should receive, and in some cases, the benefit of introducing definitions. In addition to this, it will also deal with how some verbal phraseological units of the German language should be mentioned if they constitute equivalents to units in Spanish. The practical experience that allows me to perform the analysis of the above-mentioned questions is the Spanish-German Dictionary on Idioms, currently being compiled by our team in association with the University of Applied Sciences in Cologne (Fachhochschule Köln).

A Multilingual Electronic Database of Distributionally Idiosyncratic Items

Trawiński, Beata; Soehn, Jan-Philipp; Sailer, Manfred; Richter, Frank

8. Phraseology and Collocation

Poster

We present a multilingual electronic database of lexical items with idiosyncratic occurrence patterns. Currently, our database consists of: (1)a collection of 444 bound words in German; (2)a collection of 77 bound words in English; (3)a collection of 58 negative polarity items in Romanian; (4)a collection of 84 negative polarity items in German; and (5)a collection of 52 positive polarity items in German. Our database is encoded in XML and is available via the Internet, offering dynamic and flexible access.

For an Extended Definition of Lexical Collocations

Tutin, Aanès

8. Phraseology and Collocation

Short Paper

Restricted lexical collocations have now been studied and encoded in dictionaries for over twenty years, and stable definitions have been provided for this notion by numerous scholars working on collocations (e.g. Hausmann 1989, Mel'cuk 1998, Heid 1994). They are roughly defined as recurrent combinations of two linguistic elements which have a syntactic relationship. One of the elements of the collocation, called base, keeps its usual meaning-autosemantic words (Hausmann 2004)-while the other, the collocate, is dependent on the other-synsemantic words-and usually has a less transparent meaning. Even though such a definition is nevertheless operational for a large number of lexical associations, it raises several problems. The first problems has to do with the binary status of the collocation and the unequal status of the two parts of the collocation, which has been questioned by several linguists (inter alia Siepmann 2006, Bartsch 2004) who suggest expanding the definition to associations of three or more elements. A second problem concerns the grammatical status of the collocations. Should functional words-and to what extent-be included in the definition of collocation? For example, in expressions such as for fear of, the whole combination can be analysed as a preposition, and not as a phrase contrary to prototypical collocations such as pay attention-verb phrase, major problem-noun phrase, seriously injured-adjective phrase. However, fear in for fear of can be considered as relatively transparent, and according to us, it should be considered a collocation. In this paper, we study these two issues in detail and call for an extended typology of restricted collocations. We examine the lexicographical consequences of such an extended definition.

SciE-Lex: A lexical database of collocations in scientific English for Spanish scientists

Verdaguer, Isabel; Poch, Anna; Laso, Natalia Judith; Giménez, Eva

8. Phraseology and Collocation

Poster

As a result of the widespread use of English in science and scholarship, there is an increasing need of reference tools which provide accurate information to non-native-especially junior-researchers on the correct use of lexico-grammatical patterns of non-technical words when writing their scientific papers in English and on the conventionalized phraseological characteristics of the genre. Our aim is to present SciE-Lex, a lexical database which provides information to help Spanish researchers to write

research papers in English accurately. Whereas there are specialized monolingual and bilingual dictionaries with specific terminological information, there is a shortage of reference tools supplying information on the correct use of syntactic and collocational patterns of non-technical words in the scientific register and on the conventionalized phraseological characteristics of the genre. Based on the analysis of a 3+ million word corpus of scientific English, in its first stage, SciE-Lex displays information on: word class, morphological variants, equivalent(s) in Spanish, patterns of occurrence, list of collocations, examples of real use, and notes to clarify usage. In a second stage we plan to include lexical bundles, compositional recurrent sequences of words, since several studies have confirmed the difficulties that learners have with them. Further research will provide SciE-Lex with information about the distribution of lexical bundles across the different sections and/or moves of the academic research article as well as their function in discourse.

Database of Bavarian Dialects (DBÖ) Electronically Mapped (dbo@ema). A System for Archiving, Maintaining and Field Mapping of Heterogeneous Dialect Data for the Compilation of Dialect Lexicons

Wandl-Vogt, Eveline; Kop, Christian; Fliedl, Günther; Nickel, Jost; Scholz, Johannes

8. Phraseology and Collocation

Software Demo

dbo@ema is a system for the archiving, handling and mapping of heterogenous dialect data for dialect dictionaries. Within this software presentation:

- a. the users should get known to the general project aims of dbo@ema, that are:
 - developement of a webbased, interactive data base
 - development of a webbased, interactiv tool to map dialect data and background information of a dialect dictionary
 - developement of a specific, free font for the phonetic transcription of dialect data in digital surroundings (further information see http://www.wboe.at)
- b. the users should get known to special tools of the software developed to compiling a dialect dictionary
- c. the users should get known to how geoinformation aids the compilation of a dialect dictionary
- d. the users should get known to the project Wörterbuch der bairischen Mundarten in Österreich (WBÖ) (Dictionary of Bavarian dialects in Austria) and
- e. the project Datenbank der bairischen Mundarten in Österreich (DBÖ) (Data base of bavarian dialects in Austria) that are both mother-projects to the project dbo@ema.

Incomprehensible Languages in Idioms: Functional Equivalents and Bilingual Dictionaries Woźniak, Monika

8. Phraseology and Collocation

Student Paper

Phraseology is a source of interesting information on the speakers' world view and different fixed expressions are used in different languages when one does not understand the message. Lack of understanding of what is said or written is often associated with the inability to comprehend the language, which is proved by the use of idiomatic expressions containing names of different foreign languages considered to be particularly difficult in a given society. In this paper several bilingual dictionaries are consulted in order to: 1) find equivalents of some expressions of that kind in English, Polish and Spanish; 2) review their lexicographical treatment; and 3) see how the recorded parallels correspond with the functional view of idiom equivalents proposed by Dobrovol'skij (2000a, b).

Prepositions in Dictionaries for Foreign Learners: A Cognitive Linguistic Look Adamska-Sałaciak, Arleta

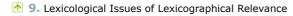
9. Lexicological Issues of Lexicographical Relevance

Full Paper

The paper is an attempt to look at the problems faced by lexicographers compiling prepositional entries in dictionaries for foreign learners, and to suggest ways in which these problems could be alleviated. After discussing some of the reasons why prepositions are difficult to deal with in a dictionary, and reporting on the results of metalexicographic studies examining the treatment of prepositions in monolingual English learners' dictionaries and in three bilingual English-Polish dictionaries, Cognitive Linguistics is suggested as a source of important insights which could be of assistance in solving practical lexicographic problems. Among those insights are: the idea that the linguistic structuring of

space functions as a mental template for other domains; recognition of the polysemic sense network of prepositional meanings; preference for *principled polysemy* over earlier *unrestricted polysemy* approaches; introduction of rigid criteria for the recognition of separate senses; recognition of the fact that the overwhelming majority of spatial senses of prepositions are related through metonymy. Drawing on the cognitive linguistic analyses of the semantics of English prepositions offered by Tyler and Evans (2003), some practical recommendations are made regarding ways in which prepositional entries in dictionaries for foreign learners could be made more informative and useful. These include a considerable reduction of the number of senses and examples of usage, an introduction of semantic 'profiles' at the beginning of entries, and supplementing verbal illustrations with simple graphics, highlighting the salient meanings of particular prepositions, the links between different senses, and the differences between semantically close and therefore frequently confused items.

Lexicographie historique, noms de métier, féminisation: quelle méthodologie? Baider, Fabienne



Full Paper

This article investigates the way trade names in the feminine form are presented in the French etymological and historical lexicographic discourse. Several French decrees in 1986, 1994 and 2000 were issued to promote use of the feminine form of trade names in reference to women. Working from two different corpuses-one before the feminization policy and one after, the analysis establishes whether progress had been made in such usage: feminine forms have increased in the past 30 years, even though their presentation remains incomplete and sometimes even marginal. However, the study of the presence or absence of these feminine forms could provide insight into what the linguistic function of gender is for various lexicographers. For some, a different gender and a different form of a trade name-ex. boulanger and boulanger-do not justify the inclusion of the feminine form, since they are derived morphologically and semantically from the masculine word, even though this case is not necessarily true. On the semantic level, this reasoning presupposes that grammatical gender does not fulfill any relevant function for nouns denoting animates. If it is impossible to conclude that these different lexicographical discursive practices support an asymmetrical representation of the sexes because of their different treatment of grammatical gender, it is nevertheless certain that such reasoning deprives all feminine forms from etymological information, hence truncating the history of words.

Scale-free Networks in Dictionaries

Fóris, Ágota



9. Lexicological Issues of Lexicographical Relevance

Poster

The aim of this paper is to show, through the application of the mathematical model of scale-free networks, how the scale-free network of language is represented in the information contained in dictionaries. Research conducted in the last few decades has proven that every phenomenon of nature and society-the relations of so many various systems-is organised into a complex system of networks. Research has also proven that complex networks can be analysed with the help of a common network model, and that the application of this network theory allows us to discover features of the analysed system that are not observable by other methods. After the discovery of the significance of networks, broad experimental and theoretical studies were launched to reveal the nature of networks and to apply the findings, and research on scale-free networks is the most outstanding among these. If we accept that the three components of the terminological unit may be modelled with the scale-free terminological network model (Fóris 2007), and that the language network is made up of at least these three networks, then we may suppose that dictionaries select and present various parts of this complex network from different approaches. The lexicographers' task-to put it simply-is to collect, record and make the data necessary for language use easily accessible. In order to meet this aim, dictionaries need to follow the three-sided structure of language networks. The various types of dictionaries compiled for different purposes developed a practical structure that reflects the structure of the language network. In the paper, I briefly touch upon the main characteristics of the scale-free network model that can be widely applied in linguistic research, and point out the lexicographic aspects of the model. Based on the network model we can draw conclusions concerning the practical structure of dictionaries. I demonstrate that the complex scale-free network structure of language containing

three sub-systems enables us to use the language quickly and completely. I also illustrate and support the features of the language network model and its application with figures.

La place du métalangage dans la définition lexicographique: l'exemple des définitions des mots syncatégorématiques dans le TLF

Frassi, Paolo

:X: 0 .

9. Lexicological Issues of Lexicographical Relevance

Poster

The studies on lexicographic definitions connected with the French tradition take charge eminently of typology and leave aside the question of metalanguage. So, in lexicography, the metalinguistic definition is often considered in the typological frame. This is because the above-mentioned studies are mostly based upon definitions either of nouns or verbs. In my presentation I shall attempt to demonstrate, from defining statements of the syncategorematic words drawn from the Trésor de la langue française, that the metalinguistic definition is indeed a category of the definitions but that, when compared to the other categories, it requires a different criteria of analysi, due to its nature. In order to do this, I shall present, first, the different nature of this issue from a typological approach on one side and a metalinguistic approach on the other. I shall expose, then, the main typological studiesin particular the unpublished document which is stored in the archives of the Laboratory ATILF ["Pour un nouveau cahier de normes...", 1979] as well as Martin (1983) and Rey-Debove (1998)-in which the question of the metalanguage is dealt with inside and following the example of typology to demonstrate that, if a definition such as aiguillette-nom populaire de l'orphie-is metalinguistic and a definition such as chaise-siège à dossier sans bras-is perifrastic, nom et siège are both hyperonyms, so that the typological criteria are not enough to distinguish between mealinguistic and perifrastic definition. Thus, I will establish, in accordance with Rey-Debove (1997), in which the definition is considered from a metalinguistic point of view-according to the sintactic relation between a lexical entry and its lexicographical definition, the principles which govern the metalinguistic analysis. The results will lead to three different categories of metalinguistic definitions of the syncategorematic words:

- 1. the definition refers to both infralinguistic and extralinguistic reality-in this case two sub-categories are possible:
 - a. the hyperonym refers to the infralinguistic reality while the specific semes refer to the extralinguistic reality;
 - b. the hyperonym refers to the infralinguistic reality while the specific semes, among which there is at least an autonym with "schize" (cf. Rey-Debove 1997: 116-118), refer to the extralinguistic reality;
- 2. the definition refers to the only infralinguistic reality;
- 3. the definition refers to the only extralinguistic reality.

Verbal Aspect and the Frame Elements in the FrameNet for Polish

Linde-Usiekniewicz, Jadwiga; Derwojedowa, Magdalena; Zawisławska, Magdalena

9. Lexicological Issues of Lexicographical Relevance

Short Paper

This paper deals with theoretical and practical problems involved when describing a language from the morphological aspect within the FrameNet. In terms of aspect and in lexicographical description of the Polish language, there is a tendency to treat pairs where the aspectual distinction is marked by suffix as a single lexical unit. Where the aspectual distinction is marked by a prefix, pairs represent different units, e.g. kaszlnac (pf. to give a cough) -kaslac (impf. to cough repeatedly) vs. pisac (impf. to write, to be writing) -napisac (pf. to have written). More complex sense relations between perfective and imperfective verbs complicate matters even more. In addition, aspectual pairs differ in terms of what constitute their core frame elements. Many perfectives differ from their imperfective counterparts since they transform temporal quantification from a non-core to a core element of the frame, e.g. Przesiedzial w bibliotece dwie godziny, studiujac rekopisy. (He sat for two (solid) hours in the library, poring over the manuscripts) vs. Siedzial w bibliotece (przez) dwie godziny, studiujac rekopisy (He sat in the library for two hours, pouring over the manuscripts). Because of this, in the Polish version of FrameNet, each member of an aspectual pair will be initially given a separate description. Once the respective frames and frame elements for each perfective and imperfective member of an aspectual pair are established independently, the two putative frames will be compared in order to see if they can be conflated into a single frame.

Verbos que traban discurso: implicaciones lexicográficas para el DAELE

López Ferrero, Carmen; Torner Castells, Sergi

9. Lexicological Issues of Lexicographical Relevance

Short Paper

Our work falls within the framework of the Project for the Elaboration of a Dictionary for Learning Spanish as a Foreign Language-Diccionario de aprendizaje del español como lengua extranjera, ref. HUM2006-06982, in progress at the Universitat Pompeu Fabra. In particular we analyse the syntactic and discursive behaviour of five semantic classes (as set by Bosque 2004), since they amount to clusters of verbs which share both the same meaning in context as a grammatical behaviour is similar so it seems and there is a high frequency of use in each type of verb of the syntactic structures and patterns. These semantic classes are the following:

- 1. Verbs for introducing, unaccusative verbs of existence and apparition: ocurrir, suceder, existir, aparecer, resultar, etc. (cfr. Bosque y Demonte 1999);
- 2. Metalinguistic verbs or verbs expressing ways of talking: decir, afirmar, asegurar, explicar, referir, etc.;
- 3. Verbs that convey to what extent the information they introduce is relevant: destacar, detallar, especificar, mostrar, sobresalir, etc.;
- 4. Verbs of comparison and contrast: comparar, contrastar, distinguir, diferenciar, oponer, etc.;
- 5. Cause-consequence verbs: causar, concluir, confirmar, consequir, depender, etc.

All of them are verbs that have been defined by text linguistics as explicit marks of textual connection in several lexicological works. The purpose of our analysis is to define the syntactic patterns and the discursive values of these groups of verbs that have such a close meaning. The constructions and combinations akin to them where the different types of verb intervene have been described in detail (Bosque y Demonte 1999 and Bosque 2004, to mention two recent works); these descriptions should be completed with the information that shows quite specifically the shared meaning and the specific syntactic meaning of two verbs of each of the semantic classes considered. This information may be systematized to carry out the lexicographical description so that it may contribute to avoiding the mistakes foreign students learning Spanish may make when using units that are semantically similar.

Verb Class-specific Criteria for the Differentiation of Senses in Dictionary Entries Proost, Kristel

9. Lexicological Issues of Lexicographical Relevance

Full Paper

This contribution deals with the representation of verbs with multiple meanings or senses in general monolingual dictionaries. Criteria for differentiating senses in dictionary entries have traditionally been formulated with respect to the vocabulary in general. This paper argues that, while some criteria do indeed apply to the entire lexicon, many of them are relevant only to specific semantic classes. This will be demonstrated considering two selected verb classes: speech-act verbs and perception verbs. Like verbs of other classes, speech-act verbs and perception verbs may be ambiguous in different but recurrent ways. Since recurrent patterns of ambiguity are always typical of particular semantic classes, class-specific semantic criteria are formulated to decide whether a particular ambiguous speech act or perception verb should be treated as being polysemous or homonymous in dictionary entries. In addition to these class-specific semantic criteria, the semantic-syntactic criterion of identity or difference of argument structure is suggested for the lexicographical representation of verbs which may not be considered to be polysemous or homonymous on the basis of semantic criteria alone. According to the suggested argument-structure criterion, these verbs should be treated as polysemous when their senses correlate with identical argument structures and as homonymous when their senses correlate with different argument structures properties. As opposed to the semantic criteria suggested, the semantic-syntactic criterion of identity vs. difference of argument structure applies to verbs of different semantic classes. However, as will be illustrated by the discussion of the different senses of smell, it may sometimes force us to treat different but related senses as corresponding to two distinct lexical items. In order to solve this problem, the criteria suggested are supplemented by a preference rule stating that semantic criteria apply prior to the semantic-syntactic criterion of identity vs. difference of argument structure...

Les dictionnaires québécois et le problème de la norme linguistique

Schafroth, Elmar

9. Lexicological Issues of Lexicographical Relevance

Short Paper

This paper deals with dictionaries of French in Quebec and the problem of language norm. In 2008, Quebec will celebrate its 400th anniversary. The publishing of the first *Dictionary of Standard French in Quebec* (Dictionnaire FRANQUS)-*Français Québécois Usage Standard*, announced as an online version for autumn 2008 and supposedly available in its printed version in 2009, will mark a new and important step in the history of Canadian French lexicography. It will be the fifth dictionary of French published within the last 20 years in Quebec, each of them conveying its own normative point of view. The article deals with these four dictionaries: the *Dictionnaire du français Plus à l'usage des francophones d'Amérique* (DFP), 1988; the *Multidictionnaire de la langue française* (MLT), 4th edition 2003; the *Dictionnaire québécois d'aujourd'hui* (DQA), 1992/1993; the *Dictionnaire québécois-français. Pour mieux se comprendre entre francophones* (DQF), 1999.

After discussing the problem of linguistic norm in general and then, especially with regard to Quebec, each of the four dictionaries will be analyzed according to a set of criteria in order to reveal the items indicating normativity. As a matter of fact, there are different types of normativity, such as the maximum orientation towards the standard of European French or the adherence to a more "Quebecist" attitude-legitimating a Quebec variety of French. The criteria are:

- the dictionaries' prefaces and introduction
- their labels indicating the value or the "correctness" of a word or a meaning
- any normative comment-the lexicographical description of English loan words-anglicisms being one of the major problems of language planning in Quebec.

On Connotation, Denotation and All That, or: Why a *Nigger* Is Not a 'Black Person' Van der Meer. Geart

9. Lexicological Issues of Lexicographical Relevance

Short Paper

In my paper I intend to demonstrate that it is, in the case of monolingual dictionaries, preferable to incorporate usage labels like *formal*, *vulgar* etc. in the sense definitions themselves instead of making them almost invisible by hiding them in the margins of the entries. I will also argue that it should be attempted to make clear exactly why a lexical item is said to be e.g. *humorous*.

Definición lexicográfica y orden de la información y de las palabras: el caso del euskera

Alberdi Larizgoitia, Xabier; García García de los Salmones, Julio; Ugarteburu Gastañares, Iñaki

10. Other Topics

Short Paper

The aim of this paper is to show how every language is determined by its syntactic structure when arranging information and words in lexicographic definitions. In the first two sections we analyze the difficulties we find in Basque with hyperonymic definitions of nouns: the informative structure. This kind of definition imposes goes from the general-hyperonym-to the particula-specific characteristics; but this order contrasts with the expansion of nominal nucleus modifiers towards the left, often found in Basque. Bearing in mind that Basque history regarding monolingual lexicography is very recent, in the third section we discuss and evaluate the solution to the aforementioned problem that Ibon Sarasola provides in his *Euskal Hiztegia* (Basque Dictionary, 1996). Basically, the solution resorts to appositional structures that get us closer to an analytical and more communicative model of definition: this way, the informative nucleus-hyperonym-precedes the specifications-modifiers in apposition. Finally, we extract the following conclusions of general-lexicography-and individual nature-Basque lexicography:

- 1. Every language is determined by its syntactic structure regarding the model of hyperonymic definition and, because of that, it must look for its own syntactic-discursive strategies.
- 2. In the Basque dictionary Euskal Hiztegia, Sarasola specifies some syntactic-discursive strategies that have been proven adequate for the definition, which converge nicely with written tradition: one of the keys for these strategies is a moderate use of appositional structures to add specifications to the informative nucleus-hyperonym.
- 3. The conclusions we have arrived at are also valid for the terminographic definition written in Basque, which similarly cannot differ in excess from the paradigm *hyperonym* (informative nucleus) + specific characteristics.

Dictionaries for University Students: A Real Deal or Merely a Marketing Ploy? Kosem, Iztok

10. Other Topics

Short Paper

Universities in English-speaking countries have experienced a sharp rise in the number of students in the past decades. One of the biggest problems students face is learning how to communicate in academic English, a language they have not experienced before. One of the tools that students often use to tackle language-related problems during their study is a dictionary. There are many dictionaries on the market, but only a few claim to be designed specifically for university students. This paper takes a closer look at these few dictionaries, and attempts to identify their unique features by comparing them with general dictionaries. The analysis reveals that the only real difference lies in the additional material-e.g. sections on academic writing, and not in the dictionary macrostructure or microstructure itself. The second part of the paper focuses on some of the features that dictionaries for university students share with general dictionaries, such as being based on corpus data, and discusses why many of these features cannot actually be acknowledged as student-friendly. The final remarks point out that publishers, researchers, and lexicographers need to acknowledge that students are a specific group of dictionary users-users that need help, not only with regard to general language, but also with academic language.

El tratamiento lexicográfico de de toute façon, de quelque façon y d'une certaine façon en el DEC Llopis Cardona, Ana

10. Other Topics

Poster

In this paper, I will examine the entries of the discourse markers de quelque façon, d'une certaine façon and de toute façon of the Dictionnaire explicatif et combinatoire du français contemporain (V.III) that Igor Mel'cuk leaded in the University of Montréal. I will propose to apply the existing format of the lexies non descriptives found in the IV volume to the definition of these discourse markers.

I will start with a linguistic description of these discourse markers in different levels. I will also explain the main aspects about macrostructure: de quelque façon and d'une certaine façon are recorded in different entries, but the first will take sends to the second, so there is only one definition for both markers. On the other hand, a semantic building can be established between d'une certaine façon and de toute façon.

Afterwards, I will analyze the aspects related to microstructure, namely: definition, lexical function and examples. The existing definition is a synonymous expression which is different from the typical definition of the DEC not only for the lexies descriptives but also for the lexies non descriptives found in the IV volume. With regard to the lexical function, several kinds of synonymy and antonymy are pointed out in the entries d'une certaine facon and de toute facon. The given expressions given of the lexical functions provide copious material to explore the differences between these discourse markers.

In conclusion, I note that the DEC is more appropriate for units that work in sentences, not in the discourse level; because the theoretical framework of the DEC was built to get the replacement of words and idiomatic expressions, but this system cannot deal with discourse markers. In addition, this lexicographical treatment doesn't include any pragmatic or communicative features, crucial to achieve a good description of these units.

euralex2008@upf.edu. Euralex 2008. Universitat Pompeu Fabra. La Rambla 30-32. 08002 Barcelona