Proceedings of the XIV Euralex International Congress

Proceedings of the XIV Euralex International Congress

Basis foarmjouwing: Roelof Koster BNO, Mildaam
Opmaak: Daam de Vries
Foto omslach: Hindrik Sijens. WFT-kartoteek Fryske Akademy
Euralexkongreslogo: Gull design & media B.V. Seisbierrum

Redaksje: Anne Dykstra, Fryske Akademy en Tanneke Schoonheim, Instituut voor Nederlandse Lexicologie, Leiden.
Redaksje-assistint: Liesje Haanstra, Fryske Akademy

Afûk, Postbus 53, 8900 AB Ljouwert

NUR 616
ISBN 978-90-6273-850-3

www.afuk.nl
www.fryske-akademy.nl

ANNE DYKSTRA AND TANNEKE SCHOONHEIM (*eds.*),

Proceedings of the XIV Euralex International Congress

(Leeuwarden, 6-10 July 2010)

FRYSKE AKADEMY – Afûk, Ljouwert 2010

··· *Foreword*

On behalf of the organizing committee of the XIV EURALEX International Congress held July 6-10 2010 at the World Trade Centre in Leeuwarden, it is my pleasure to present the proceedings, which include the plenary papers, contributed papers, posters, and summaries of software demonstrations. The congress was organized by the Fryske Akademy (Frisian Academy), in cooperation with the Leiden Instituut voor Nederlandse Lexicologie (Institute for Dutch Lexicology).

The Fryske Akademy has its seat in Leeuwarden, the capital of the Dutch bilingual Province of Friesland. The Fryske Akademy does research on Frisian and Friesland. Over the years the Fryske Akademy has published a wide range of Frisian dictionaries. The Fryske Akademy considers it a great honour to host the XIV EURALEX International Congress. I could not have wished for a better platform for the presentation of the on line version of the scholarly *Wurdboek fan de Fryske taal* (Dictionary of the Frisian Language). The Mercator Research Centre at the Fryske Akademy studies the position of lesser-used languages in education, gives information on national and regional education systems and provides the latest statistics regarding lesser-used languages in education in the European Union. Frisian being a small non-state language, the organizing committee especially welcomed papers on any aspect of the lexicography of non-state or lesser used languages. I take great pleasure in saying that these proceedings contain ten contributions on that topic, a very satisfying number. I am also very happy with the response to the survey that we conducted to learn more about the state of affairs in the lexicography of non-state or lesser used languages. My warm thanks go to all the people that went to the trouble of completing the lengthy questionnaire. In one of the plenary lectures Anne Popkema (Rijksuniversiteit Groningen) discusses the outcomes of the survey. In another plenary lecture Sarah Ogilvie (University of Cambridge) also deals with the lexicography of non-state or lesser used languages.

We consider ourselves very fortunate that Anne Popkema and Sarah Ogilvie accepted our invitation to deliver a plenary paper at the congress, and this also goes for the other plenary speakers, whose papers are included in these proceedings. The subject matter of Anatoly Liberman's (University of Minnesota) paper is: 'The Genre and Uses of the Etymological Dictionary.'

Arleta Adamska (Adam Mickiewicz University) dwells upon 'Lexicographic equivalence' and Hornby Lecturer Paul Bogaards deals with the subject of 'Dictionaries and foreign language learning.' I would like to thank the Hornby Trust here for their generous support in sponsoring this lecture, which is in honour of A. S. Hornby, renowned for his work on learner's dictionaries for non-native speakers.

Following the example of the Barcelona organizing committee, we present the proceedings as a book with an accompanying CD-ROM. The book contains the five plenary papers presented at the congress, the abstracts of all the papers, posters, and software demonstrations accepted for presentation, and an index to this CD-ROM. The CD-ROM contains contributed papers, posters and summaries of software demonstrations, as well as the plenary lectures. The CD-ROM allows you to access a paper either by author or by title. In addition, you may also create a complete copy of the proceedings.

The proceedings of the XIII EURALEX Congress in Barcelona provided a broad range of the various contributions in nine different sections. The Leeuwarden proceedings have two more sections, one of them being Sign Language. Sign Language as a subject is attracting ever more attention in linguistics and I am happy that the growing interest in Sign Language is also reflected in lexicography. The three contributions on Sign Language in these proceedings, from three different countries, therefore deserve a separate section. The other section not included in the Barcelona Proceedings is: Lexicography of lesser used or non-state languages. The papers in these proceedings are organized in the following sections:

Computational Lexicography and Lexicology
The Dictionary-Making Process
Reports on Lexicographical and Lexicological Projects
Bilingual Lexicography
Lexicography for Specialised Languages – Terminology and Terminography
Historical and Scholarly Lexicography and Etymology
Dictionary Use
Phraseology and Collocation
Lexicological Issues of Lexicographical Relevance
Lexicography of Lesser Used or Non-State Languages
Sign Language

The papers are organized alphabetically by first author within each section on the CD-ROM included with this volume.

All submissions that were sent to the organizing committee were reviewed by two anonymous and independent referees. The referees evaluated the papers using the EASYABSTRACTS (EasyAbs) Abstract Submission and Review Facility made available free of charge by *The Linguist List* (http://linguistlist.org/confcustom/). EasyAbs guaranteed anonymous evaluation. The use of this double-blind peer review procedure has ensured the academic quality of the papers, posters, and demonstrations presented at EURALEX 2010. I am pleased to thank the following people who participated in this review process:

Elisenda Bernal	Universitat Pompeu Fabra
Paul Bogaards	International Journal of Lexicography
Anna Braasch	Centre for Language Technology, University of Copenhagen
Lut Colman	Instituut voor Nederlandse Lexicologie
Janet DeCesaris	Universitat Pompeu Fabra
Jesse De Does	Instituut voor Nederlandse Lexicologie
Katrien Depuydt	Instituut voor Nederlandse Lexicologie
Pieter Duijff	Fryske Akademy
Ruth Fjeld	Institutt for lingvistiske og nordiske studier
Pius ten Hacken	Swansea University
Frans Heyvaert	Instituut voor Nederlandse Lexicologie
Eric Hoekstra	Fryske Akademy
Robert Lew	School of English, Adam Mickiewicz University
Carla Marello	Università di Torino
Willy Martin	Professor Emeritus, Vrije Universiteit, Amsterdam
Geart van der Meer	Formerly Department of English, University of Groningen
Marijke Mooijaart	Instituut voor Nederlandse Lexicologie
Michael Rundell	Lexicography MasterClass Ltd
Tanneke Schoonheim	Instituut voor Nederlandse Lexicologie
Hindrik Sijens	Fryske Akademy
Rob Tempelaars	Instituut voor Nederlandse Lexicologie
Carole Tiberius	Instituut voor Nederlandse Lexicologie
Lars Trap-Jensen	Society for Danish Language and Literature
Willem Visser	Fryske Akademy
Alastair Walker	Nordfriesische Wörterbuchstelle, Universität Kiel
Geoffrey Williams	Université de Bretagne-Sud

The following people joined me on the EURALEX 2010 programme committee:

Janet DeCesaris	Universitat Pompeu Fabra
Ruth Fjeld	Institutt for lingvistiske og nordiske studier
Pius ten Hacken	Swansea University
Tanneke Schoonheim	Instituut voor Nederlandse Lexicologie
Hindrik Sijens	Fryske Akademy
Willem Visser	Fryske Akademy and Rijksuniversiteit Groningen

I thank them all for their commitment and their advice and encouragement.

Tanneke Schoonheim, Hindrik Sijens and Willem Visser were also members of the local organizing committee, which was completed by Liesje Haanstra, Linda Hoekstra, Joop Petter and Reinier Salverda. Without them EURALEX 2010 would never have been organized. A special word of thanks should go to Liesje Haanstra, who helped me to keep track of the situation.

I also would like to thank the Fryske Akademy, not only for allowing me to organize EURALEX 2010, but also for giving me support when I needed it. Furthermore, I would like to thank the Instituut voor Nederlandse Lexicologie for delegating Tanneke Schoonheim to the local committee. Congresbureau Friesland proved to be an indispensable partner in the organization of the congress, thank you, Jant van Dijk and Aranka Ties.

The organization of EURALEX 2010 in Leeuwarden received financial support from the following sources, which we gratefully acknowledge:

Stichting Woudsend Anno 1816
Province of Friesland
Municipality of Leeuwarden
Instituut voor Nederlandse Lexicologie
Boersma-Adema Stichting
Frysk Akademyfûns

In the process of editing the proceedings, my fellow editor Tanneke Schoonheim and I noticed that many contributions referred to papers in earlier EURALEX proceedings, which may serve as an illustration of the ever growing importance of the EURALEX congresses.

On behalf of everyone associated with the organization of EURALEX 2010 in Leeuwarden, I would like to thank all the contributors for submitting very interesting work, and for meeting the tight production schedule of these proceedings.

Anne Dykstra
Chair, XIV EURALEX International Congress
May, 2010

··· *Contents*

Section 2 – The Dictionary-Making Process

Section 3 – Reports on Lexicographical and Lexicological Projects

Section 4 – Bilingual Lexicography

Section 5 – Lexicography for Specialised Languages – Terminology and Terminography

Section 6 – Historical and Scholarly Lexicography and Etymology

Section 7 – Dictionary Use

Section 8 – Phraseology and Collocation

Section 9 – Lexicological Issues of Lexicographical Relevance

··· *Plenary Lectures*

> Lexicography and Endangered Languages: What Can Europe Learn from the Rest of the World?

SARAH OGILVIE

1 Introduction

In 1926, at the age of 24, the controversial American anthropologist Margaret Mead (1901-1978) was on her first field trip to Samoa. She wrote back to her PhD supervisor at Columbia University, the famous Professor Franz Boas (1858-1942), saying 'Through it all, I have no idea whether I am doing the right thing or not, or how valuable my results will be. It all weighs rather heavily on my mind'.[1]

This, I would like to suggest, is the sentiment of every fieldworker who is documenting a language for the first time, and, if the language is endangered, most probably for the last time. Linguistic documentation and description has traditionally entailed recording the language, transcribing the language, and writing a grammar of the language. Writing a dictionary of the language was more a stepping stone towards the grammar, rather than a goal in itself. Most dictionaries of endangered languages, therefore, are compiled by linguists or anthropologists who are not lexicographers. They learn the craft 'on the job', and most of these 'new lexicographers' – and I say this from personal experience – feel the same bewilderment as Margaret Mead: they have no idea whether they are doing the right thing or not, or how valuable their results will be.

This was certainly the case for me when, twenty years ago, I arrived in an Aboriginal community on the tip of Cape York Peninsula in remote northern Australia with a tape recorder, ten cassettes, a notebook and pencil, and a change of shorts and t-shirt. My task was to write a grammar and dictionary of an endangered Aboriginal language called Morrobalama. The language had never been recorded or written down before, and my task was to describe the language before the last two speakers died.

My only experience of dictionary writing at that time had been as pronunciation editor on the *Australian Concise Oxford Dictionary* 8th ed., responsible for creating the first Oxford dictionary with Australian rather than British pronunciations. That job did little to prepare me for

[1] Letter from Margaret Mead to Franz Boas, 16 Jan, 1926. 'The Correspondence Between Margaret Mead and Franz Boas Exchanged During Mead's 1925-26 Samoan Research Project (and related material)' <http://sociology.uwo.ca/mead/>

writing a dictionary of a language that 1) I had never heard before, 2) had never been written down before, and 3) was spoken by two last speakers who were rarely sober enough to teach me their language. The safe and comfortable offices of Oxford University Press were in stark contrast to life in a community where I slept on the ground each night on a burnt-out mattress, ate fresh-water turtle, and generally felt like the 'stranger' on every level: culturally, linguistically, and socially.

When one crosses the boundary between one's own language and culture into another, one can't help but be changed by it. Claire Bowern calls this a 'peculiar displacement' in which 'the fieldworker is displaced from their own community and culture, and is sent to think analytically about another social and linguistic system'.[2] Daniel Everett described his experience of documenting the Pirahã language in the Amazon as akin to 'becoming an alien'.[3] He warned other field workers:

> You could become a 'freak' instead of an attractive person; an incompetent, instead of a respected professional; ugly instead of lovely; fat instead of average; stinky instead of normal-smelling; and on and on. You may go from being articulate and witty in conversation to being perceived as an infantile dullard who can barely function in conversation. You will go from having many friends to having none. From enjoying good company, to stark loneliness.

Twenty years ago, field lexicography was lagging behind commercial lexicography on all levels, and my experience of dictionary making in the field as opposed to the office certainly provided a stark contrast to my future experiences in lexicography, as I went on to be Senior Editor on the *Macquarie Dictionary* and various Oxford Dictionaries in Australia, and more recently on the *Shorter Oxford Dictionary* and the *Oxford English Dictionary* (*OED*) in the United Kingdom. In the world of language description, there was barely any overlap between field linguists and commercial lexicographers, and I found myself in the unusual position of combining the two. In recent years however, linguists have started to do innovative work on collecting primary data and rethinking the principles, theories, and practice of documenting languages and cultures. Their concern not only for language preservation but also for its maintenance and revitalization has meant that field linguists have had to rethink how to write dictionaries.

[2] Bowern (2007:10)

[3] Everett (2006:6)

What can we – as members of Euralex and as writers of European mainstream dictionaries and dictionaries of minority languages – learn from dictionaries produced by documentation linguists? What is the potential relationship between linguists and lexicographers? When I was an editor on the *OED*, I was the only one of forty editors trained in linguistics. I remember being surprised by this, and one day raising it with the Chief Editor. He explained that he preferred it that way because linguists thought too much about things. It is true that thought can slow things down in lexicography… Seriously though, an historian or literary scholar can often contribute more to historical lexicography than the specialist in linguistics. However, there have been changes in the area of descriptive and documentary linguistics in the past decade that suggest that linguists might have something to teach lexicographers. How might we all share our expertise with one another? Can we make more of the relationship between lexicographers and linguists than we have in the past?

2 Lesser Used Non-state Languages

The theme of this paper is specifically dictionaries of endangered languages, but the theme of the 14th Euralex International Congress is dictionaries of 'lesser used or non-state languages'. While all endangered languages fall in this category, not all lesser used non-state languages are endangered, i.e. endangerment depends on the degree of language shift. Twenty years ago, Joshua Fishman identified eight steps toward reversing language shift.[4] The steps progressed from the ultimate goal of step 1 – making a language the language of national government – to the easiest goal of step 8 – reconstructing the language and designing language learning programmes. Where a language sat on this spectrum was considered a barometer of its chances of being saved and revitalized. Speakers of non-endangered languages that are lesser used and non-state probably take Fishman's step 8 as a given, and step 1 as a real desire and possibility. Speakers of endangered languages, on the other hand, may strive for step 8 and not even dream of the possibility of step 1.

But that was twenty years ago, and many linguists see things differently now. They follow the lead of Leanne Hinton who shifted the focus from the national to the domestic, from the ultimate goal of government recognition and sanctioned use to the realization that languages must first be spoken at home by children if they have a chance of being spoken

[4] Fishman (1991)

anywhere.[5] This change in scholarship has affected in fundamental ways the approach to linguistic description and the nature, focus, and quality of documentation and revitalization programmes. I would like to suggest that it has also changed the nature, focus, and quality of dictionaries of endangered languages in ways that all of us can learn from, regardless of whether our aims are to promote our language to national or domestic level.

The past ten years have seen the emergence of new lexicographic policies and practices around the world that can be characterized by an innovative exploitation of new technologies, predominant use of oral as well as written sources, incorporation of pedagogical materials, and collaborative involvement of members of the speech community. For these reasons, this paper will focus specifically on lexicography of endangered languages around the world.

3 Endangered Languages

There is no doubt that one of the most important issues facing humankind today is the rate at which our languages are dying. On present trends, the next century will see more than half of the world's 6800 languages become extinct, and most of these will disappear without being adequately recorded.[6] Current language distribution shows that 96% of the world's languages are spoken by only 4% of the world's population. David Crystal calculated in his book *Language Death* that one language dies on average every two weeks.[7] And, of course, more is lost than mere words. As vehicles for the transmission of unique cultural knowledge, local languages encode oral traditions that become threatened when elders die and livelihoods are disrupted. When a language disappears so does a culture and a speech community's unique way of seeing and ordering the world.

What kinds of languages are we talking about? Let me play for you a few words of Yurok, a North American Indian language, with six fluent speakers left. In this case, the language has been documented (and I will talk about that later) but it remains to be seen whether it will be revived successfully. If not, it will die out in the next decade. Likewise this video of one of the last ten speakers of Kayardild, a language of northern Australia. In twenty years time, unless the language is properly documented and

[5] See Hinton (1997) and Hinton and Hale (2001).

[6] Krauss (1992), Crystal (2002:19).

[7] Crystal (2002:19)

revived no one will be speaking Kayardild and cultural connections such as these will be lost.

4 Language Documentation and Description

Unless the academic community works swiftly with indigenous communities and NGOs in collaborative and innovative ways, most of this expressive diversity will disappear without being adequately recorded or given a chance of conservation and revitalization. An important first step in slowing down or reversing the process of language death is to document the language in the form of a dictionary. Using innovative lexicographic policies, practices, and technologies, the lexicographer is able to produce dictionaries that are useful to both communities and scholars; dictionaries that not only describe and preserve an endangered language – as was the goal of linguists in the past – but also help in the processes of maintenance and revitalization.

Writing dictionaries of this kind is important on a number of levels. On an immediate level, as lexicographers, we have a duty to speakers of a language to record and describe their words with precision, accuracy, and in a way that is most useful to them. For those of us who are linguists, our linguistic theories depend on linguistic diversity and the rigorous description of that diversity. But more important, for humanity in general, is the need to preserve cultural diversity and knowledge systems that can be encoded in a dictionary.

For many years in descriptive linguistics, academics wrote dictionaries of endangered languages that were merely by-products of their primary aim – which was to describe the grammar of a language. My work on Morrobalama certainly fell in this category. But linguists and indigenous communities now recognize the important role that dictionaries can play in the documentation, preservation, and revitalization of endangered languages, and the past decade has seen linguists and anthropologists begin to focus on dictionaries as important tools and products in themselves.[8] These changes have been accompanied by new trends in Documentary Linguistics and Anthropology as priority research areas that deal with the principles, theories, and practice of documenting languages and cultures that are at risk.[9] In 1998, in a landmark article in the journal *Linguistics,* Nikolaus Himmelmann formally distinguished

[8] See Frawley, Hill, and Munro (2002) for evidence of this.

[9] See Himmelmann (2002), Woodbury (2003), and Austin (2006).

between language documentation and language description. The aims of language documentation were to record the primary data of language study, e.g. spoken and written texts which are transcribed, annotated with metadata, and archived for posterity. Language description, by contrast, was concerned with the secondary data of language study, e.g. analysis of primary data in the form of dictionaries and grammars. Since then, however, dictionaries of endangered languages have begun to blur the boundaries between documentation and description. More and more, they have become repositories for primary data which include images, sound, and video. This development has coincided with innovations in technology and documentation techniques thereby opening up the field of lexicography beyond academia so that academics are joined in the task by indigenous communities, educators, and certain NGOs whose work involves language support.

5 Compiling Dictionaries of Endangered Languages

For the endangered-language speech community, the most useful and relevant research outcome of field linguistics is usually the dictionary. Articles and books on syntax, morphology, or phonology have little relevance to indigenous speech communities. Dictionaries, on the other hand, are not only useful and functional texts, but emblems and tools of prestige which many communities use to boost their sense of identity and political profiles.

For the lexicographer, the field situation often presents a complex set of challenges that have an impact on lexicographic policies and practices. On top of the challenging living conditions, an undocumented language presents challenges relating to audience (are you writing for scholars or the speech community?), format (will it be a print dictionary, web-based, or electronic with imbedded pictures, sound, or video? Will the dictionary be linked to learning materials?), and compilation (what orthography and writing system will you devise? How do you list words in a dictionary if the language does not really have separate lemmas but rather joins up all the units of meaning into one polysynthetic word that we would probably call a sentence? How will the compilation involve the speech community? What software will you choose to accomplish this?). All of these issues – the audience, format and mode of compilation of the dictionary – will depend on region; health of the language and degrees of endangerment; community attitudes towards language, literacy, and learning; and access to electricity and internet.

The collaborative dictionary-making efforts of academics, community members, and NGOs are producing dictionaries that are community-focused and collaborative in their compilation, content, and format. Currently, in response to different degrees of language endangerment, dictionary projects around the world fall into one of three categories: dictionaries for language preservation, dictionaries for language maintenance, or dictionaries for language revitalization. While this paper is not an exhaustive survey of projects around the world, I have chosen some examples of dictionary projects which have developed methodologies that nonetheless might have applicability to European dictionaries whether they be of minority or majority languages.

5.1 – Dictionaries for Language Preservation

In the Aslian (Mon-Khmer, Austroasiatic) languages of the equatorial forests of Malaysia, Niclas Burenhult is currently compiling dictionaries of Jahai, Semnam, Menriq, Batek, Lenoh, and Maniq. They focus on descriptions of unique ethnobiological knowledge about the forest and how to make a sustainable livelihood from it. In compiling the dictionaries, Burenhult faced tricky decisions relating to the order of entries, choosing not to order the headwords alphabetically but rather according to manner and place of articulation with left-to-right ordering rather than rhyming order, as is the tradition in many Austroasiatic dictionaries. At this stage, with no literate speakers, the dictionaries are primarily for preservation and scholarly purposes.

5.2 – Dictionaries for Language Maintenance

While access to computers and the internet is rare in many remote parts of the world, mobile phone access is not. In remote parts of Australia, for example, the presence of mining companies in the Outback has brought network access to areas that probably would not normally have been priority zones for telecommunication companies. Hence, perhaps surprisingly, people in remote Aboriginal communities currently own and use mobile phones more than any other form of technology. There has been a successful dictionary program by James McElvenny and Aidan Wilson at Sydney University, the Project for Free Electronic Dictionaries, to install dictionaries of endangered Australian Aboriginal languages on mobile phones.[10] Loaded on to a mobile phone via software called

[10] The dictionary software for mobile phones can be downloaded at http://www.pfed.info/wksite

Wunderkammer, a Java ME MIDlet, each dictionary entry has a spoken pronunciation and many entries have pictures. Currently, the Wagiman language, spoken in the Northern Territory of Australia, is on mobile phones, and further projects are currently underway for Tuva, a language of the Ivory Coast, and Whitesands, a language of Vanuatu.

In recent years there has been a trend in endangered-language lexicography to produce small dictionaries of semantic fields. These are particularly suited to language maintenance, in the sense that breaking down the mammoth overall task of compiling a comprehensive dictionary into 'mini dictionaries', provides the speech community with quick access to a dictionary of their language for use in schools and the community in general. Ulrike Mosel and Ruth Spriggs compiled mini dictionaries of Teop, a language spoken in Papua, which covered semantic fields such as House Building, Body and Health, Fish, Shells, and Trees. The mini dictionaries were collaborative efforts with older speakers who assisted with editing, young speakers who checked the clarity of the entries, and children who gave feedback on the dictionary's lexical coverage (e.g. Teop children collected shells which they found missing in the first draft of the shell dictionary). Mosel and Spriggs found that collaborative lexicographic activities such as these promoted language awareness and pride in young speakers, the targeted demographic for successful language maintenance or revitalization. Being able frequently to present the speech community with tangible results of lexicographic work, in the form of mini dictionaries, rather than wait years for the completion of a comprehensive dictionary, has the additional benefit of demonstrating the lexicographer's commitment to language maintenance and revitalization in the community, and their ability to produce results.

5.3 – Dictionaries for Language Revitalization

It is in the area of language revitalization that the most exciting lexicographic work is taking place. Dictionaries written for revitalization have to address quite a complex set of issues relating to the stage of endangerment, level of literacy, and opportunity for capacity building and empowerment of community members to revitalize their language.

Dictionaries of all endangered languages have the added pressure of having to be compiled quickly, or at least the materials must be collected quickly, before the last speakers die. The Iquito Dictionary Project in northern Peruvian Amazonia, led by Christine Beier and Lev Michael, advocates a team-based and community-participatory approach to dictionary writing which helps in fast collection of data.[11] The research team comprised 2-3 community linguists and 4-7 visiting linguists (professors and graduate students) who visited the field at the same time. The initial task of the visiting linguists is to help with capacity building and skills-transfer activities so that community members can be trained as 'community linguists', and work alongside the research team. In the case of Iquito, an Amazonian language with 25 speakers all of whom are over the age of 65 years, a few of the community members were immediately trained in basic aspects of descriptive linguistics and language documentation. Training of this sort is not always a straight forward process, as it is often the case that last speakers of endangered languages are not literate, and members of the community who are literate may not be proficient in any of the indigenous language. It is therefore important to incorporate literate adults as 'community linguists' and traditional speakers as 'language specialists'.[12]

The working schedule for team-based dictionary projects is highly structured. The collection of data for the dictionary takes the form of weekly data-gathering tasks or 'modules' – each task allocated to a different team member – the results of which are reported daily to the rest of the team in the form of a 'seminar'. In the case of the Iquito Dictionary project, this schedule of Module-and-Seminar continued every day during the two-month visits by the academics each summer for three years (2003-2006), and dictionary compilation continued throughout the year by the local community linguists. The language is now preserved in a printed bilingual *Iquito-Spanish Dictionary*, and the community linguists now teach Iquito language classes in the community's schools.[13]

Transfer of skills and capacity building are therefore responsible for turning what may have just been a language preservation dictionary project into a language revitalization dictionary project. The project trained a group of independent local experts – community linguists and language specialists – who could serve the community beyond the life of the dictionary compilation. The inclusion of graduate students in the research teams was also an ideal way of training and mentoring future lexicographers – all the while supporting their first experience of field lexicography with social, scholarly, and material infrastructure. Not only does this boost the numbers of linguists and anthropologists who learn

[11] See Beier and Michael (2006)

[12] See Beier (2009:4)

[13] In addition to the dictionary, the IDLP (Iquito Language Documentation Project) team also produced grammatical analyses and an extensive collection of audio, video, and written texts which are described further in Beier (2009) and Michael (2009).

the art of lexicography in the field, but it also increases the productivity and amount of dictionary work carried out in any one field trip.

From the perspective of the visiting lexicographer, the team-based approach to dictionary-writing has the additional benefit of providing social support in what can otherwise be an isolating situation. However, the lexicographer must be careful that her/his integration into the speech community is not jeopardized by the comfort of socializing solely with other members of the visiting team. In an oral culture, the field lexicographer's access to words and language is increased by his/her ability to integrate into a speech community. Hence, within the team-based model of documentation the lexicographer must be careful not to rely too heavily on the social support of other visiting members of the team, especially if such socializing would neglect relationships within the community.

Another capacity building strategy in the rest of the world that supports dictionary making is the BOLD (Basic Oral Language Documentation) initiative in Papua New Guinea, in which Olympus has donated hundreds of voice recorders for traditional speakers to record their languages. This project, organized by Steven Bird, has a strict schedule of voice recording and transcribing, all of which can feed into dictionary-building. With about 850 languages, Papua New Guinea is the most densely populated region for language diversity in the world. The BOLD project provides Olympus VN5200PC digital voice recorders to one hundred speakers of different languages. Over a period of one year (February 2010 – February 2011), participants commit to a three-stage process: first, participants record 10 hours of culturally-rich speech (e.g. conversation, personal narratives, and idiomatic speech). The next stage involves re-playing the recordings and re-speaking the oral translation with another speaker on another digital recorder. This recording therefore contains not only the original recorded text but also a commentary on it. The third and final step involves choosing one or two segments of the original recording that amount to six minutes of spoken language, and transcribing it.

On average, BOLD participants spend one hour transcribing one minute of recorded text. Many of these languages have not been written down before, so the process of transcription in stage three will prompt the speakers and participants to think about the written representation of sound and the challenges of devising an orthography for their own languages.[14] These transcribed texts and primary data will go towards writing dictionaries

[14] For more on native orthographies, see Harrison and Anderson (2006).

and grammars of the languages of Papua New Guinea, but, along the way, the process will have trained native speakers in the techniques of language documentation, created community interest and pride in their traditional languages, and, in many cases, prompted indigenous community members to think about their languages in a new way.

For critically endangered languages (those with no child speakers), it is not only necessary to record the language quickly, but it is important for the dictionary content to facilitate, or potentially facilitate, language revitalization. In addition to the resultant skills transfer from collaborative techniques of dictionary compilation, there are also mechanisms within the dictionary itself that can aid revitalization and make the text more appealing, functional, and useful to language learners, especially children. For example, for communities with computer and internet access, such as the Yurok North American Indian tribe in northern California, the dictionary entries can be linked to language memory tests and language learning exercises with dictionary audio files.[15]

Available free online, the Yurok Dictionary is similar in structure to the *Oxford English Dictionary* in that it is an historical dictionary which shows the use of Yurok words over time. It makes use of the fact that the Yurok language was recorded at different times throughout the twentieth century. Recordings of the language on wax cylinders have made it possible for Andrew Garrett at UC Berkeley to include a quotation paragraph after each definition showing how the word was used at different points of the twentieth century. For example, in 1902 and 1907, the language was recorded by the famous American anthropologist A. L. Kroeber (1876-1960); in 1927, it was recorded by the doyen of linguistics Edward Sapir (1884-1939); in 1958, it was recorded by the British linguist R. H. Robins (1921-2000); in 1980 and 1986 the same speaker whom Robins had recorded, Florence Shaughnessy, was recorded again by Paul Proulx and Jean Perry respectively; and finally in 2007, the last remaining speakers were also recorded. The Yurok Dictionary is able to supply each entry with recorded illustrative sentences from throughout the twentieth century (1902-2007), as exemplified at the entry *kwelekw,* adverb meaning 'well'. Illustrative sentences are linked to the larger texts in which they appear, and users see a picture of the original speaker and can read or listen to the original recordings of the entire stories, such as this recording of Mary Marshall telling Edward Sapir the story of 'Coyote Tries to Kill the Sun' in 1927 or Domingo of Weitchpec telling A. L. Kroeber the story of 'Buzzard's Medicine' in 1907.

[15] http://linguistics.berkeley.edu/~yurok/web/random.html

There is one important difference between the historical examples in the Yurok dictionary as opposed to those in the *OED*. The Yurok examples are predominantly based on spoken, rather than written, evidence. Dictionaries of endangered languages are based on oral more than print culture which thereby captures more words from different genres. In my own work as a lexicographer in the UK, I had responsibility for non-European words in the *OED*, so I am aware of the restrictive implications of inclusion policies that require a minimum of five published citations over five years. This policy, based on concerns for the unreliability of spoken or unpublished sources, is particularly difficult to satisfy for words in English from parts of the world without established publishing traditions. A word in Philippine English may not appear five times in print but it may be used in the English of 42 million speakers. Hence I am aware of the hundreds of words that did not get in to the dictionary because of the bias in our European lexicographic tradition toward printed sources. Inclusion policies based on the number of citations from written sources get increasingly difficult to defend as technology improves our ability to capture, reproduce, and verify natural speech in natural contexts. Perhaps this is an area in which mainstream lexicography will follow innovations in field lexicography.

Unlike most lexicographers of minority languages in Europe, who are frequently native speakers of the language they are describing, lexicographers of endangered languages must undergo the slow process of learning the language they are describing. If they are writing dictionaries for language revitalization, they face the added challenge of not only learning the language themselves but also facilitating the learning (and teaching) of language for others within the community. In addition to creating a text – like the Yurok Dictionary – that facilitates language learning, the lexicographer may be in a position to empower native speakers and young adults in the community to work together so that young members acquire conversational proficiency in the traditional language. By doing this, the lexicographer can help to ensure that language learning becomes a part of the community culture beyond the life of the dictionary project. As explained by Chief Harry Wallace, the elected leader of the Unkechaug Nation (Long Island): 'When our children study their own language and culture, they perform better academically. They have a core foundation to rely on'.[16] The Africanist Paul Newman, however, criticizes these efforts by lexicographers and linguists because he argues that they should not

[16] As quoted in 'Indian Tribes Go in Search of Their Lost Languages' *New York Times* 6 April 2010, C1.

become mere 'linguistic social workers' who waste their skills and time on the 'hopeless cause' of language revitalization.[17] Far from a hopeless cause however, there are numerous examples of lexicographers around the world who successfully negotiate a balance between dictionary work and revitalization work, and for dictionaries written with revitalization as one of the outcomes, many would argue that the two are inseparable. Indeed, many field lexicographers successfully facilitate language revitalization, and in turn these efforts result in increased dictionary use and ultimately a reinforcement of the lexicographer's raison d'etre.

One proven and successful methodology for bringing native speakers together with language learners is the Master-Apprentice Program, originally devised by Leanne Hinton, Nancy Richardson, and Mary Bates Abbott for revitalization of Californian languages.[18] By instituting this method while compiling the dictionary, the lexicographer lays the foundation for other one-on-one relationships between traditional speakers (the Masters) and language learners (the Apprentices). Hence, at the same time that the lexicographer learns the language from the Master, s/he also sets up a facility for language learning that can be replicated by other members of the community. The program advocates five main principles: 1) The Master and the lexicographer must not speak together in the dominant language (i.e. the language which is replacing the endangered language); 2) only oral (not written) language must be transmitted; 3) the lexicographer must be at least as active as the Master in deciding what is to be learned and in keeping communication going in the language; 4) learning must take place in real-life situations and traditional activities e.g. collecting food, going hunting, cooking, and doing crafts; 5) it must all be recorded or videoed for later analysis and use in the dictionary.

Advocating and practising a lexicographic methodology that facilitates the maintenance and revitalization of endangered languages is only part of the process. Ultimately, of course, whether or not a language survives – and the role that a dictionary plays in this process – will depend on the speakers themselves i.e. their attitudes towards the language in general and their willingness for inter-generational language transmission.

Activists for preservation of endangered languages often stress the urgency

[17] Newman (1999) and Newman (2003).

[18] See Hinton (1997) and Hinton (2001) for more information on the Master-Apprentice Program.

of capturing and saving languages before they disappear, arguing that it is literally a matter of life or death. Is it? The logical extreme of dictionaries for revitalization, of course, are those that are written from direct contact with no speakers at all. It is possible to revive a language from written sources alone (e.g. modern Hebrew) and every field lexicographer must hold in their mind the possibility that their own work may one day be used for such a purpose. In 1791, when the third President of the United States and the principal author of the Declaration of Independence, Thomas Jefferson (1743-1826), collected a wordlist from the last three speakers of Unkechaug, he probably had no idea that their descendents would be using his wordlist to revive the language on Long Island in 2010.[19]

Indeed, the current work by lexicographers of endangered languages will surely provide materials for language programs of the future. The exact sound, form, and structure of that language may not be exactly the same as that recorded by the lexicographer but the dictionary maker must be mindful of the possible future uses of her/his work. Unlike dictionary work on languages with established literary traditions, like those in Europe, the stakes are particularly high with endangered languages. The accuracy with which a lexicographer describes the sound, form, meaning, history, and usage of words from endangered languages may be the only lasting record of a language and culture, and future generations will depend on it in unforeseen ways: 'Would someone from 200 years ago think we had a funny accent?' asked Robert Hoberman, organizer of the Unkechaug revitalization, 'Yes. Would they understand it? I hope so.'[20]

Similarly, Natasha Warner and Quirina Luna are currently writing a dictionary of Mutsun, the language traditionally spoken south of San Francisco, California. It has been extinct, or 'dormant' as Warner and Luna prefer to describe it, since 1930, but the lexicographers are hoping that their dictionary will enable 'all interested members of the community to achieve reasonable fluency in (the revitalized form of) the language, at which point it is likely that some Mutsuns would be raising their children in Mutsun'.[21] The dictionary was compiled using original notes and materials by the early nineteenth-century Roman Catholic missionary, Felipe Arroyo de la Cuesta, and the early twentieth-century anthropologist, J. P. Harrington. In the 1920s, the eccentric Harrington

[19] See 'Indian Tribes on Long Island Go in Search of Their Lost Languages' *New York Times* 6 April 2010 C5.

[20] Robert Hoberman quoted in 'Indian Tribes on Long Island Go in Search of Their Lost Languages' NYT 6 April 2010 C5.

[21] Warner, Butler, and Luna-Costillas (2006:259)

collected 36,000 pages of notes on Mutsun (within a two year period) from the last fluent speaker, an elderly Mrs Ascension Solorsano. These have been collated into a dictionary of headwords with a uniform orthography. The lexicographers were also faced with the task of inventing new Mutsun terms for the modern world, e.g. restaurant 'ammamsa' = eat+locative nominalizer.

The Mutsun dictionary initiative and the Unkechaug revitalization efforts both came out of a workshop organized by Leanne Hinton at UC Berkeley called 'Breath of Life'. Every two years, the Breath of Life workshop brings 60 people who identify as North American Indians to UC Berkeley for one week. They are united by one similarity: their traditional languages are extinct, but each person is accompanied by two mentors who are lexicographers or linguists. They spend the week receiving intensive training each morning in the basics of lexicography and linguistics. Each afternoon, they are shown how to use the rich linguistic and anthropological archives housed at UC Berkeley, and each evening the participants work on their own projects which might include writing a poem or song in their traditional languages, or beginning to compile a dictionary. At the end of the week, each person presents their project to the larger group. The Breath of Life workshop has provided descendants of North American Indian tribes with the tools to produce dictionaries out of the silence of archives, libraries, and extinct languages. It is being replicated else where in the world, e.g. early this year there was a Breath of Life workshop in Outback Australia and in the Canadian Arctic Archipelago of Nunavut.

6 Lexicography as a Means of Skills Transfer and Capacity Building

As seen with the Iquito Dictionary Project, the BOLD initiative, and the Teop Dictionary Project, the advent of language documentation as a field in itself has opened new opportunities for ensuring that dictionaries of endangered languages are community-focused and collaborative. New technologies and software allow dictionaries to imbed sound, video, and texts. They also allow multi-user access during the compilation process i.e. indigenous dictionary makers are jointly able to edit dictionaries with linguists living elsewhere in the world, thereby forming a dictionary team that can simultaneously work on the dictionary from different parts of the world. Such web-based collaboration is possible via an open-source software application called *Wesay*, produced by Summer Institute of Linguistics (SIL) in Papua New Guinea and Thailand. Intended for rugged low-power hardware, such as notebooks, *Wesay* specially caters to the needs of indigenous dictionary makers by providing them with a simple and easy interface that requires minimal training. The software was developed especially for the speakers of endangered languages so that they can create their own dictionaries.

There are currently efforts to put this dictionary-making software on thousands of laptops being distributed to the world's poorest children via the One Laptop Per Child (OLPC) program. This initiative provides each child with a rugged, low-cost, low-power, connected laptop which is able to take photos and record sound. The software and learning materials currently provided on the laptops are in the dominant languages of the regions (e.g. mainly in Spanish or English) which of course increases the risk of endangering indigenous languages. Hence, there are currently initiatives to add Wesay dictionary-making software, along with lexicographic pedagogical materials, so that speakers of endangered languages in countries such as Peru, Rwanda, and Uruguay can document their languages and be introduced to dictionary-making activities at school level. It is hoped that classes will be able to compile their own mini dictionaries of community languages, thereby not only recording the languages but also increasing the children's use and pride in them.

The most sophisticated new dictionary-making technologies which enable speech communities to be involved in the documentation of their own languages are called Lexus and ViCoS. They are web-based tools by which lexicographers, both in the field and outside the field, can create (simultaneously) dictionaries that include sound, video, and immediate links to the relevant video segment where any word occurs. They also allow a dictionary to capture the indigenous view of the world by including a kind of visual thesaurus that presents indigenous semantic networks, i.e. networks that display the way speakers order and conceptualize semantic categories. These are particularly useful for communities that are not largely literate, and for dictionary users who rely more on visual and auditory than textual features. For example, in Gaby Cablitz's dictionary of the Polynesian language, Marquesan, a user can look up the meaning of a verb and see it in action. A user can look up the entry *kae,* a transitive verb meaning 'to cut or split off bark of a trunk or branch with a knife', and press a video to see how *kae* is performed.

The advent of documentary linguistics has encouraged lexicographers to integrate documentary materials into the text so as to create multimedia dictionaries which are more like cultural encyclopedias in their range. And, as we saw in the Yurok Dictionary, multimedia dictionaries can also combine new lexical data with older archive material, allowing diachronic perspectives.

The inclusion of multimedia materials, and the desire for dictionaries of endangered languages to include socio-cultural information, opens the lexicographer to new considerations of ethical issues. The interests of the speakers are primary in the lexicographer's mind. In addition to negotiating extra issues with the speech community such as informed consent, payment for language consultants, and sharing outcomes, lexicographers of endangered languages must be mindful of cultural sensitivities surrounding the material they are documenting, i.e. access to sacred songs, taboo words, or the voice or image of Elders who may soon be dead (and whose name, voice, or image must not be uttered, heard, or seen for a certain period of time). Hence, many parts of the Yurok Dictionary are password protected. During dictionary work by Marina Chumakina on Archi, a north-east Caucasian language spoken by 1200 people in southern Dagestan, Russia, sound files were recorded for every word in the dictionary by member of the community. At the end of the project, it became apparent that in such a small community, where everyone knew each other's voice, the speaker was embarrassed that the rest of the community would hear her saying words considered taboo, such as intimate parts of the body. She asked for those files to be excluded, and her wish was respected. Similar issues surround illustrative sentences based on recorded speech that includes gossip or private stories which would be easily recognized within small speech communities.

Software such as Lexus and Wesay enable a dictionary to be compiled over the internet in a wiki-like fashion, and software such as ViCoS and Protégé enable the speech community to have a linguistic resource linked to the dictionary that represents their own intuitions and ontologies. For example, the dictionary of Yami, a language of Taiwan, includes links to ontologies which represent indigenous semantic connections between fish names, e.g. the Yami tripartite distinction between edible fish for young men, edible fish for women, and edible fish for old men.[22] The Yami Dictionary used Protégé software to show the semantic connections between the fish, but there is other software available, the most well-known being Kirrkirr. Kirrkirr pioneered work in semantic networks and was developed originally to work with the Warlpiri Dictionary, an Australian Aboriginal language, published by Mary Laughren and David Nash in 1983. Since then the software has been developed further by scholars at

[22] See Rau et al (2009).

Sydney University and Stanford. By creating a semantic network view, the lexicographer presents the user with a network in which words in the dictionary that are semantically related are connected together by coloured lines – each colour represents a different relationship e.g. same meaning or alternate forms. By creating a semantic domain view, the lexicographer presents the user with nested nodes that represent semantic domains. Given the current limitations of remote places (lack of electricity, computers, and internet access), these online ontology tools are still a little way off being used to their full potential, but they are certainly indicative of the direction in which field lexicography is heading.

One issue to consider with dictionary software is that of archiving, which is neither reliable nor guaranteed especially as software is updated and changed. Therefore some field lexicographers avoid dictionary-making proprietary software because they are concerned about the longevity and archiving of their data. For example, the datafiles of the Yurok Dictionary and the Hupa Dictionary are XML documents and the interface is run via an XSL style sheet. This is wise when you consider that dictionary work on an endangered language may be the final record of the language, so it is imperative that it is stored in ways that are flexible, enduring, and easily accessible.

7 Conclusion: the Impact of Language Documentation on Lexicography

The emergence of the field of language documentation in the past decade has clearly had an impact on dictionary writing. And this paper has provided a glimpse of dictionary projects around the world that are creating methodologies that might be relevant or insightful to European lexicographers. The lexicographer cannot ignore the new focus on primary data; the new recognition of the importance of collaboration and involvement of the speech community in the dictionary-making process; the new concerns for accountability and ethics; the new concern for storage and accessibility of archived dictionary materials; and the new possibilities that technology brings to both the content of dictionaries and their compilation.

On the macro level, language documentation has increased creation of, and access to, innovative dictionary technologies. It has also increased the opportunity for lexicographers to engage in capacity building, transfer of skills, and empowerment of community members to share the responsibility of dictionary making. On the micro level, the impact of language documentation on lexicography is perhaps even more tangibly obvious. These dictionaries of endangered languages comprise a wider inventory from a variety of speech genres, with sophisticated multimedia materials, and new ways of preserving cultural memory and representing semantic and cultural ontologies. Content is linked to learning materials which facilitate language revitalization so that the dictionary becomes more than just a means of language preservation; it becomes the catalyst and focus for living language. These dictionaries challenge traditional types of dictionaries because they are everything in one. They combine aspects of the learner's dictionary, historical dictionary, encyclopaedic dictionary, talking dictionary, pictorial dictionary, video dictionary, and visual thesaurus. Consequently, the field lexicographer wears many hats. Her/his lexicographic methods and practices incorporate aspects of all genres of dictionary writing, and her/his mode of dictionary compilation is collaborative in nature. This paper has presented ways that lexicographers around the globe are able to preserve, maintain, and revitalize endangered languages. While Europe created and shaped the art of dictionary writing as we know it today, the rest of the world is taking it in new directions.

> Bibliography

Austen, P. K. (2006). 'Data and Language Documentation.' In J. Gippert, N. Himmelmann and U. Mosel (eds.) *Fundamentals of Language Documentation* Berlin: Mouton de Gruyter, 87-112.

Beier, Christine and Lev Michael (2006). 'The Iquito Language Documentation Project: Developing team-based methods for language documentation.' In *Linguistic Discovery* 4(1).

Beier, Chris (2009). 'Implementing Community Involvement in Collaborative Language Documentation: The Iquito Case.' *First International Conference on Language Documentation and Conservation.*

Bowern, Claire (2007). *Linguistic Fieldwork* Palgrave Macmillan

Crystal, D. (2002). *Language Death* Cambridge: Cambridge University Press.

Everett, D. (2006). *Linguistic Fieldwork: A Student Guide* draft 3 April unpubl. ms.

Fishman, Joshua (1991). *Reversing Language Shift: Theoretical and Empirical Foundations of Assistance to Threatened Languages* Multilingual Matters.

Frawley, William, Hill, Kenneth, and Pamela Munro (2002). *Making Dictionaries: Preserving Indigenous Languages of the Americas* Berkeley: University of California Press.

Harrison, K. D. and Anderson, Gregory (2006). 'Na(t)ive Orthographies and Language Endangerment: Two Cases from Siberia.' In *Linguistic Discovery* 4 (1).

Himmelmann, N. P. (1998). 'Documentary and Descriptive Linguistics.' In *Linguistics* 36:165-191.

Himmelmann, N. P. (2002). 'Documentary and Descriptive Linguistics.' In O. Sakiyama and F. Endo (eds.). *Lectures on Endangered Languages* 5 Kyoto: Endangered Languages of the Pacific Rim 5:37-84.

Hinton, Leanne (1997). 'Survival of Endangered Languages: The California Master-Apprentice Program.' In *International Journal of the Sociology of Language* 123:177-191.

Hinton, Leanne (2001). 'The Master-Apprentice Language Learning Program.' In Hinton and Hale (eds.). *The Green Book of Language Revitalization in Practice* San Diego: Academic Press: 217-226.

Hinton, Leanne and Ken Hale (eds.). (2001). *The Green Book of Language Revitalization in Practice* San Diego: Academic Press.

Krauss, M. (1992). 'The World's Languages in Crisis.' In *Language* 68:4-10.

Michael, Lev (2009). 'Developing Infrastructure for team-based language documentation and description: the Module-and-Seminar Model.' In *First International Conference on Language Documentation and Conservation* 1 (1):1-6.

Newman, Paul (1999). 'We have seen the enemy and it is us: the endangered language as a hopeless cause.' In *Studies in the Linguistic Sciences* 28(2):11-20.

Newman, Paul (2003). 'The endangered language issue as a hopeless cause.' In M. Janse & S. Tol (eds.). *Language Death and Language Maintenance: Theoretical, practical and descriptive approaches* Amsterdam: John Benjamins, 1-13.

D. Victoria Rau, Meng-Chien Yang, Hui-Huan Ann Chang, and Maa-Neu Dong (2009). 'Online Dictionary and Ontology Building for Austronesian Languages in Taiwan.' In *Language Documentation and Conservation* 3 (2):192-212.

Warner, Natasha, Butler, Lynnika, and Quirina Luna-Costillas (2006). 'Making a Dictionary for Community Revitalization; the case of Mutsun.' In *International Journal of Lexicography* 19 (3):257-285.

Warner, Natasha, Luna, Quirina, Butler, Lynnika, and Heather Van Volkinburg (2009). 'Revitalization in a scattered language community: Problems and methods from the perspective of Mutsun language revitalization.' In *International Journal of the Sociology of Language* 198:135-148.

Woodbury, T. (2003). 'Defining Documentary Linguistics.' In Peter K. Austin (ed.). *Language Documentation and Description* vol. 1 SOAS: University of London: 35-51.

> The Genre of the Etymological Dictionary

ANATOLY LIBERMAN

1 Etymological dictionaries among other comprehensive dictionaries

Etymological dictionaries are stepchildren of lexicography. In surveys, at best a few pages are devoted to them. Even lists of the etymological dictionaries of English, German, Dutch, and the Scandinavian languages, that is, of languages having a strong tradition of producing such reference works, were impossible to find until I compiled and published them (Liberman 1998; 1999; 2005; the Dutch list still awaits publication). But years after I began this work I keep running into old, not necessarily worthless, works that fell between the cracks and wonder how many more I have missed (cf. the postscript to my 1999 paper added in proof and Liberman 2009b). One does not have to look far for the reason complicating this search. The common habit of depending on the latest products, which, allegedly, contain more pieces of distilled wisdom than their predecessors, severed our ties with the past, and few experts consult the first editions of Kluge (1883, etc.) or any of the four editions of Wedgwood (1859-1865, etc.), let alone dictionaries by less distinguished authors. As a result, references to them are rare. Theory of lexicography and excellent instructions to lexicographers exist (the market seems to be always ready for new encyclopedias and voluminous 'handbooks'), but no one except Yakov Malkiel (1975) has taken the trouble to analyze the practice of etymological lexicography or look at the multitude of etymological dictionaries written in the post-medieval period. Those who know his book may agree that despite its scope it is not a fully satisfactory guide to the subject, partly because of Malkiel's penchant for baroque style and partly because he was preeminently an expert in Romance linguistics, which made his opinions about Germanic and Slavic lexicography less valuable. Nor can a bird's eye view of any subject replace a series of more specialized works.

It is easy to see why etymological dictionaries have been pushed to the margin of theoretical lexicography. Ever since people became literate, they have been putting together glossaries and 'lexicons.' Travelers, merchants, statesmen, and officials had to communicate with foreigners, and in every epoch some language had the status of the most prestigious one, whether Egyptian, Hittite, Greek, Latin, French, or English. It is therefore no wonder that we have bilingual texts from the dawn of human civilization and thousands of medieval glosses. The collapse of the Tower of Babel provided language teachers and lexicographers with permanent employment. As time went on, culture gave an impetus to the compilation of dictionaries of synonyms, homonyms, antonyms, slang, and so forth. By contrast, etymology, though excellent for lexicographic dessert, cannot pretend to be anyone's main course. To speak, read, and write well, we do not need information on word origins.

The study of language history is inseparable from etymology, but the public can thrive without knowing where words came from, and two factors keep this branch of scholarship afloat: the inertia of historical linguistics as an academic subject and humanity's natural curiosity. People love etiological tales ('just so stories'): they want to know how the big bang occurred, why the bat hunts at night, when and under what circumstances language originated, and, among other things, how sign (the form of any given word) and meaning are connected. This thirst for knowledge is almost instinctive (at least it is ineradicable), as shown by the popularity of word columns and countless books with titles like 'Why Do We Say So?' Etymological dictionaries purport to satisfy both professionals and the uninitiated, but, as regards their appeal, they cannot compete with explanatory, bilingual, and pronouncing dictionaries. To put it bluntly, they occupy the place they deserve, but without them the world would have been poorer; so may they live and multiply.

2 The reception of etymological dictionaries

In literary studies and art history, reception theory has been a major topic for decades. In lexicography, it hardly exists at all. Scandal once resulted in the appearance of books featuring and commenting on the main reviews of a dictionary (Sledd [and] Ebbitt [1962]; Morton [1994]). Of course, I mean *Webster's Third International*..., and how misspent those passions appear to us today! Dictionary wars have been documented. The reaction of the public to Samuel Johnson and the *OED* has been traced up to a point, but on the whole, as I said, reception of dictionaries by lay users and professionals is an almost nonexistent area. The authors of etymological dictionaries are even worse off than other lexicographers, for reviews of such dictionaries have never been collected or analyzed. Sometimes I wonder who reads them. Even the authors often disregard sensible suggestions while preparing later editions. Probably they have never seen the reviews.

I am speaking from experience. More than twenty years ago, I began work on a new etymological dictionary of English. My goal was to write entries in which the literature on the origin of words would be discussed as fully as possible, various conjectures sifted, and reasonable conclusions drawn from the data. The models were many: Walde-Hofmann for Latin (1938-1954), Feist for Gothic (1939), Vasmer for Russian (1950-1958), and quite a few other etymological dictionaries (of Hittite, Classical Greek, French, Spanish, Old High German, Lithuanian, several Slavic languages, Old Icelandic, and Old Irish). An English dictionary of this type does not exist. Skeat (1882/1910), like his predecessors and followers, gave almost no references, so that someone who decides to study the etymology of an English word in depth starts practically from scratch. To what extent the project on which I embarked in the early eighties is feasible, given the resources at my disposal, is beyond the point in the present context, but the task I faced could not be clearer: it was necessary, for the first time ever, to collect the enormous literature on the origin and history of English words, summarize the findings, and offer convincing solutions.

To find the relevant articles not only in the most visible journals but also in countless fugitive periodicals (with minor exceptions, word columns and letters to the editor in newspapers remained untapped by my team of about a hundred volunteers and meagerly paid assistants) was a formidable task. I had no illusions about the completeness of the final product (one cannot read everything; besides, new articles and books appear every day), especially because etymology is based on a good knowledge of cognates. One should screen the literature in and on all the Indo-European languages (and occasionally on the languages of other families) in the hope of finding the sought-for answers outside English (for example, someone might have guessed the origin of German *gleiten*, and this would solve the etymology of Engl. *glide*, or perhaps a preliminary agreement has been reached on Dutch *big*, which would then shed light on its connection with Engl. *pig*; the importance of works on borrowings and on words belonging to the Indo-European stock needs no proof). Reviews were among the sources I studied with great care. All the publications used for the database have been copied, and more than 20,000 of them are kept in my office. At least a thousand of them are reviews.

The database, as well as the introductory ('showcase') volume of the dictionary, has now been published (Liberman 2008 [dictionary] and 2009a [bibliography]). Since the reviews that ended up on my desk could be put to use only insofar as they contained discussion of words, some, however insightful, were not included in the bibliography, but I excerpted and preserved the rejects. My acquaintance with them (brief and long, devoted to minutiae and attacking general questions) justifies my statement that reviews of etymological dictionaries have not been used for any conclusions about the genre of the etymological dictionary and exercise minimal influence on the authors. I hope to write a book on the etymological dictionaries of the Germanic languages and in addition to a survey and analysis of all of them, discuss their reception. Over the years, reviewers have been asking and often answering the same questions that interest me. Perhaps this chapter will even expedite the birth of reception theory in lexicography.

3 The readership and the market of etymological dictionaries

Above I said that etymology stays alive (or afloat) because it is the foundation of historical linguistics and because the public wants to know where words come from. Every dictionary has a certain user in view. Although the authors of etymological dictionaries cannot disregard this circumstance, they do not always know what to make of it, for their idea of their audience is dim. It is instructive to compare introductions to etymological dictionaries. In Western Europe, the earliest of them appeared in 1599 (Kilian; Dutch). Kilian's work was followed by similar dictionaries of German, English, French, and other languages. Front matter sometimes ran to more than a hundred pages and offered the author's views on the origin of language and the derivation of words (a tradition that was upheld by Wedgwood and Skeat among many others, who in this respect did not differ from Samuel Johnson and Webster). It was not deemed necessary to justify the production of such a book since the uses and benefits of etymology were taken for granted.

The first dictionaries were sometimes sold by subscription, and the lists of subscribers are long and impressive, from dukes down. Occasionally the first edition would be brought out by the author, who would break even or make a profit, so that the next edition would be undertaken by a commercial publisher. This is what happened to Minsheu, for example (1617; 1627). As late as the nineteen-eighties, in the days of Skeat and Kluge, etymological dictionaries still had a respectable market: every gentleman was likely to own a copy, and country squires read them like fiction. Nowadays publishers depend almost entirely on libraries. Outside the circle of historical linguists, most people have a lively but perfunctory interest in word origins, which seldom goes beyond exotica, slang and family names. Even if etymological dictionaries were less helpless in dealing with the origin of *cocktail*, *dodge*, *scalawag*, and their likes, what they say on the subject can be found in any other 'thick' dictionary. One would have thought that this state of affairs would have stopped the production of vapid etymological dictionaries, but the stream flows on unimpeded.

Books are published to be sold. Hence the tendency to coax the reader into purchasing etymological dictionaries by emphasizing their novelty (a new edition is supposed to be an improvement on the previous one by definition), accessibility (no knowledge of linguistics or any other 'prerequisite' is expected), and increased bulk (the more words are included, the better—also by definition). Every now and then an additional incentive is mentioned. Under the Nazis, Kluge's classical work was advertised as 'a German dictionary for the German people.' The preface to a relatively recent (serious and scholarly) etymological dictionary of Icelandic celebrated the fact that it was the first etymological dictionary of Icelandic written in Icelandic. At nearly the same time an etymological dictionary of English brought out in the United States proudly announced that it was the first American work of its kind. This focus had predictable negative consequences. Presses churn out ever new dictionaries that recycle trivial information. Fortunately, the sources of academic subsidies have not yet dried up, and from time to time we witness the appearance of real etymological dictionaries, rather than their pompous digests.

The truth of the matter is that an etymological dictionary requires a prepared user. Since grammar is not considered to be 'fun,' our college graduates have trouble distinguishing nouns from adjectives and subjects from objects. (A recent handbook of linguistics for literary scholars provides its readers with the definitions of such terms as *vowel* and *consonant.*) The origin of slang is hard to discover but easy to explain. All the other cases are more complicated. No one without previous exposure to special courses can appreciate the methods of etymological analysis. The role of cognates (and the concept of a related form), their choice, the difference between a cognate and a borrowing, sound correspondences, the idea of a protoform, the periodization of language history (Old/Middle/early Modern English, archaic Latin, and so forth), their confusing nomenclature (Germanic, German; Baltic; Old Prussian, which has nothing to do with the language spoken in modern Prussia; Old Saxon, Anglo-Saxon, German Saxon dialects; Old and Middle German, which is sometimes 'High' and sometimes 'Low'; Anglo-French and northern French as opposed to Parisian French; Middle Dutch contemporaneous with Old Frisian, the absence of 'Old Dutch'; Old Norse: what is Norse?), the basic facts of history (the Scandinavian invasion, the Norman Conquest, the epoch of Humanism) are subjects most of which the so-called general reader rarely knows. Elmar Seebold supplied his revision of Kluge's dictionary with a long list of terms, including *umlaut*, *ablaut*, *vriddhi*, and many others, quite unlike *vowel* and *consonant.* Yet that same semi-mythical general reader who wants to learn the origin of a German word will hardly agree to wander in the thicket of special terms and ponder their meaning. Nowadays we expect instant satisfaction or 'money back.' Etymology does provide satisfaction, but it is not instant, and there is not much money in it.

Reputable, especially academic, publishers have to choose between profit and excellence. In theory, they strive for both, but they cannot afford big losses: once they are out of business, there will be neither profit nor excellence. Good etymological dictionaries are doomed to attract only specialists. In principle, this conclusion does not spell disaster. Books on laryngeals in Indo-European, Verner's Law, palatalization in Dutch dialects, Scandinavian accents, and even the style of Shakespeare's sonnets are written for those who study accents, palatalization, and the rest. Our society is still rich enough to support the interest of a chosen few in such esoteric subjects. But a feeling prevails that dictionaries are different: allegedly, they must be 'popular.' This feeling may be justified in many cases, but not in etymology. Etymology is no less special than organic chemistry, and etymological lexicography has to resign itself to this fact.

Since I work on an etymological dictionary of English, I will confine myself to the field I know best. The latest special etymological dictionary of English was published almost exactly a hundred years ago (Skeat 1910). It was not special enough (see section 5, below), but it lived up to the expectations of its readership. The dictionary was addressed to language historians and to those who had a good deal of Latin and Greek driven into them at school and after. Although Skeat never tired of berating his countrymen for their laziness, ignorance, and inability to understand what etymology is all about, he relied on their familiarity with the rudiments of grammar in its Latin guise (a luxury none of us can afford) and understanding that languages develop and change. His concise dictionary was just that: concise (the entries were shorter but not simplified in comparison with the full opus). Even if sometimes he despised his reader, he does not seem to have been worried by the idea that he was talking over his head. The next etymological dictionary of English, and the last written by a serious researcher (Weekley 1921), is a watered down version of Skeat and the *OED*, though Weekley had many original ideas, especially about words of French origin and words derived from names. In the English speaking world, specialists as the main target group were forgotten (not in theory but in practice), for they could not make such dictionaries profitable. Since roughly the First World War English etymological dictionaries have been written only for 'everyman.'

The results of this attitude were disastrous. Despite the abandonment of the idea that *Prinzipien der Sprachgeschichte* is a synonym of *Introduction to Linguistics* and the reorientation of language sciences toward synchrony, etymological studies continued. The 20th century witnessed outstanding progress in the investigation of early Indo-European. Great etymological dictionaries of the living Romance, Slavic, and Baltic, as well as of several dead and reconstructed languages, were written, and numerous publications clarified the origin of both common and obscure words in and outside English. But the authors of new English etymological dictionaries paid minimal or no attention to them. In some cases they did not know those works (if I am not mistaken, the indefatigable Eric Partridge, the author of the widely used dictionary *Origins*, could not read German), in others they did not care to do sufficient research. Collecting books and articles on etymology is a time-consuming occupation (see above), and 'everyone' had no intention to 'dive below' or watch lexicographers perform such aquatic tricks. Recycling and repackaging the information in Skeat and the *OED* guaranteed safety (for even their outdated opinions were clever) and satisfied the public that could not distinguish between original work and a rehash of the classics. That is why not a single dictionary of English etymology contains more than one volume, while the great dictionary of French occupies a whole shelf, and in this respect von Wartburg, its author and editor, was not alone, though no other project acquired such gigantic dimensions (that is, if we stay with a dictionary of a single language rather than a group).

Surprisingly, with regard to etymology, two 'thick' English explanatory dictionaries went beyond what we find in Weekley and even Skeat. Since the days of Blount (1656) every comprehensive English dictionary has included information on word origins. As long as etymology remained guesswork and every conjecture seemed to be thought provoking, this custom made sense, but the appearance of the first edition of Skeat (1882), and before him of Wedgwood (1859-1865), though Wedgwood never wielded so much authority, dictionary makers could only copy from the recognized masters: inventing more sophisticated derivations became a dangerous enterprise. An etymologist or a consultant remains a familiar figure on the staff of some great dictionaries. However, this person's duty is not to outdo Skeat and the *OED* but to replace their conjectures if those have been shown wrong (an unusual occurrence) and trace the origin of the most recent words. Even that task is hard to perform, and not too rarely we find the latest dubious solution replacing the old one only because it is brand-new and published in a prestigious journal by an illustrious author (as happened to *boy*, *girl*, and *filch*, among others; a typical example of 'haste makes waste' or 'don't jump on the bandwagon'), along with the absolutely secure but irritating verdict 'of unknown origin.'

The etymologies in Webster, even in the Collegiate digest, are not easy. They presuppose a user aware of the things mentioned at the beginning of this section, that is, someone who knows the difference between Old and Middle English, appreciates the role of cognates and protoforms, understands that English is a language of the Germanic group and of the Indo-European family, and has been taught that etymology can seldom be absolutely certain. But those who open Webster more often look up meanings and can skip the etymological introduction. The answer to the fateful question—has such a dictionary been written for the expert or for 'everyman'?—depends on the etymologies to a minimal degree, while in a specialized etymological dictionary the information on the origin of words is all that matters. In any case, the two dictionaries, mentioned above, bravely offered detailed and highly professional etymologies. They are *The Century Dictionary* and Wyld's *UED*. As far as I can judge, both are hardly ever used by modern etymologists, who rely only on Skeat and the *OED* (my opinion is based on the absence of references to them in the books and articles I read). Yet both have a great deal to offer, and some of their suggestions do not recur elsewhere. Depending on the word we investigate, they may be more useful than any post-Skeat and post-*OED* etymological dictionary of English.

Serious etymological dictionaries, such as are worthy of their name, should be written for specialists and in this respect share common ground with books on mechanical engineering, calculus, and a host of others. This probably means that they can be published only by presses depending on institutional support and that in the first year hardly more than two or three hundred copies will be sold. Skeat's days are over: an etymological dictionary is no longer a status symbol and cannot rival an easel with an unfinished picture or a piano with the score of Brahms's variations, even though no one around ever painted or played. After the appearance of a dictionary written for the expert, producing a simplified concise version is an excellent idea, but first the profession should be served. Otherwise there will be nothing to simplify or abridge and we will forever stay with books on 'why do we say so?' Nowadays the word *elitist* is a term of abuse. Let this semantic somersault remain on the conscience of those who coined it. We should not be bullied into the belief that the only possible variety of an etymology dictionary is the popular one.

4 The stratification of vocabulary in etymological dictionaries. The words to be included

In what follows I will concentrate on dictionaries for the expert. Their genre also needs clarification. One of the main questions here is about the depth of etymologies. At the moment, the prevalent tendency in Indo-European etymological lexicography is to stress distant reconstruction. The thrust of the Leiden project is a good example of what can be expected in the future, and the Norwegian experience points in the same direction. This approach has its advantages and disadvantages. In the modern Indo-European languages, most words, if we exclude borrowings from Greek, Latin, and partly French, cannot be traced to hoary antiquity, so that the scope of the inherited element in their vocabulary is limited. In the Germanic, Slavic, and Celtic languages, hundreds of words, even the older ones, have likewise at best vague Indo-European connections despite their age. In the attempt to derive as many words as possible from the roots in Walde-Pokorny (1927-1932) and Pokorny (1959) or relegate them to the pre-Indo-European substrate compilers pay less attention to late medieval, early modern, and recent words. The *cocktail-dodge-scalawag* group is a distraction to them, and they prefer to ignore its existence or confess that they have nothing to say about its history.

Lexicographers never stop bothering about the number of words to be included in the dictionaries they edit. Considerations of size interest the authors of etymological dictionaries no less than their colleagues, but I am not sure that they have ever been debated. As pointed out, all one can find is an occasional blurb promising more words than ever. Since modern 'thick' explanatory dictionaries regularly feature etymologies (with the result that someone who wants to know the origin of *come*, *go*, *take*, *door*, *wall*, and other common words will find a brief but reliable answer in any non-etymological dictionary, whether on paper or online), it seems that at present the main effort should be directed toward the derivation of the words passed over or insufficiently explained. I see no virtue in writing a long entry on *brother* and *eight* in an etymological dictionary of English, as opposed to two uninformative lines on *scalawag* and omitting *dodge* altogether. Obviously, *brother* and *eight* cannot be excluded, but the level of our knowledge is such that the entries on them often contain only lists of cognates. Since knowing those cognates is not tantamount to understanding their origin, I would argue for giving them limited, perhaps even minimal space in prospective etymological dictionaries, unless the author has new ideas on how they were coined.

In my opinion, the tradition of writing Romance dictionaries has everything to recommend it. Wherever possible, their authors trace the word to Latin, and the rest is left out. The implication seems to be that anyone who is enlightened enough to understand an entry in a French, Italian, or Spanish etymological dictionary and wants to know more should turn to a dictionary of Latin. Nor is Latin our last destination, but in Walde-Hofmann the oldest reconstructable form in the spirit of Walde-Pokorny is also given. The message is: If you want to go as far as the present state of the etymological science allows us to go, use several dictionaries. Obviously, my proposal runs counter to the ideas of those who are mainly concerned with the Indo-European sources of our vocabulary. Yet my and their efforts are not at cross-purposes; they are rather complementary. I am only saying that even an ideal etymological dictionary cannot be all things to all people.

Etymologists venerate archaisms. It is their pleasure to walk among ruins, and their predilections should be treated with understanding and a measure of sympathy. A great favorite of English etymological dictionaries used to be the adjective *nesh* 'soft.' The word is regional, and the reason it has been favored over thousands of other local words is its ascertained old age (it was recorded in Old English, and its cognate turned up in Gothic, Old High German, and West Frisian) and its relative transparency (see *hnasqus** in Feist). Very few people will look up *nesh* in Skeat, but those who will may not know Gothic *hnasqus** or Old Engl. *hnesce* and will be grateful for finding an entry on it.

A more revealing example is the adjective *loom* 'moderate, gentle' (said of a breeze). This is also a local (northern English and Scots) word, and only the most detailed dictionaries of Modern English include it (for example, it will be found in the third edition of *Webster's International…*, but not in the latest edition of *The Shorter Oxford…*). According to Webster's dictionary, the origin of this adjective has not been discovered. It would be a waste of space to feature it in an etymological dictionary for the sole purpose of saying 'of unknown/uncertain/debatable origin.' But the history of *loom* (adj.) is not a blank. Frisian dialects have *luum* 'lazy, depressed,' *loom* 'thin, tired, lazy, *etc.*', and other similar forms. It is not clear how, if at all, they are related to Dutch *loom* 'slack, slow, *etc.*', Old High German -*luom* occurring in several compounds, German *lau*- in *lauwarm* 'tepid, lukewarm', the Germanic words for 'lame' (Engl. *lame*, German *lahm*, etc.) and Engl. *loom* 'appear indistinctly' (possibly of Low German or Dutch origin), but, in any case, the English adjective has been removed from its isolation, and now it is the etymologists' business to disentangle the knot. I knew neither the English adjective nor its Frisian cognates until I read about them in Faltings (1996:106-107). Unfortunately, I missed the article while working on my database, and it does not appear in the published bibliography (Liberman 2009).

This example shows that a consultant in the employ of a great dictionary has no chance of revising obscure etymologies. How can anyone confronted with the question about the origin of the English adjective *loom* find the elucidating passage in *NOWELE* 28/29? And who has enough time to begin searching for cognates in the multiple dialectal dictionaries of Frisian and Low German on the off-chance of running into something useful? Lexicographers are always 'caught in the web of words' and cannot afford spending long hours on what may become a wild goose chase, for they suspect that if the origin of a hard word is still unknown, there must be a good reason for it. (Don't we remember Meillet's unkind and unfair remark that all the good etymologies have been discovered, while new etymologies are usually bad? I think this was said at least a hundred years ago.) My team and I screened all the available philological and popular periodicals in two dozen languages for more than three centuries and 'by chance' unearthed countless important but forgotten publications whose titles frequently had no bearing on etymology; yet I managed to overlook Faltings's article despite its promising title. Another lesson I can draw from this example concerns the choice of words. Even if a word is rare or local but if its origin is worthy of note (such are, to my mind, the English adjectives *nesh* and *loom*), it should be featured in an etymological dictionary. Including them is a luxury, but an etymological dictionary has been conceived as a feast for hungry minds, however Micawberian this statement may sound.

In deciding how many words to include in a dictionary, etymologists will be well-advised to show restraint. Not only the users' expectations but the state of the art and common sense should guide their hand. Given the overabundance of competing presses and the easy access people (at least in developed countries) have to the Internet, an etymological dictionary has become a reference book for a limited, mainly professional readership (advanced students and their teachers). Those who want to learn the origin of *antelope*, *papaya*, *baritone*, *algebra*, *samovar*, and *Schadenfreude* will hardly buy or even open an *English* etymological dictionary, for their curiosity can be satisfied in an easier and cheaper way (just Google for *papaya*: etymology or *algebra*: etymology, or look it up in the pocket edition of *The American Heritage Dictionary of the English Language* and be informed: '*papaya* <Cariban [sic],' '*algebra* <Ar[abic] *al-jebr, al-jabr* 'the (science of) reuniting"; do many people want to go further?). This does not mean that

borrowings should be ignored. *Bigot, ghetto, fiasco, rascal,* and others like them should be dealt with in detail, even though an English scholar hardly has enough expertise to risk an independent opinion about their history. Skeat included about 12,000 words in his dictionary, and many of them could have been dispensed with. At present, seven or eight thousand non-derived words will be quite enough to answer the main questions about the origin of English vocabulary.

Before concluding this section, I would like to explain to what extent I do what I preach. The English etymological dictionary on which I have been working for more than twenty years won't be comprehensive, like Skeat's or Weekley's, or the *ODEE.* At the very outset I realized that I would have little or nothing to contribute to explicating words like *brother* and *eight.* Not only the basic numerals and kin terms but also many other words with broad Indo-European connections, such as *hear* and *break,* have been the object of numerous profound articles, dissertations, and books. It would have been presumptuous to expect that I was able to offer innovative suggestions in this area, and I did not want to spend the rest of my life only writing summaries of other people's opinions. The same holds for loanwords from Romance languages. I can perhaps shed some light on the etymology of Germanic words without established cognates in the rest of Indo-European, but even here one has to tackle *bride, dwarf, God,* the notorious maritime vocabulary (*ship, sea, sail,* etc.), and the rest. So I decided to concentrate on the dregs of English etymology, the words lacking unquestionable cognates outside English. It is these words about which dictionaries usually say: 'Origin unknown.' Predictably, I ended up with *scalawag-cocktail-dodge.* My database is all-encompassing (whatever came to my mill was called grist), but, if my rough estimation is accurate, the dictionary will feature about 1,000 words like those three. Some of them have been around for centuries. Before looking at every word like *nesh* and *loom,* I cannot say which of them I will include. I will avoid volatile and exotic slang, but, other than that, slang will occupy a noticeable place in the final product.

So far, my experience has been entirely positive. As soon as about two-thirds of the bibliography had been assembled, it became clear that even conscientious researchers are unaware of some valuable publications to the field. Jacob Grimm read all there was to read, and so did (presumably) Benfey, Feist, and perhaps von Wartburg, but hardly anyone else (not Skeat, and certainly not Onions). The standard verdict 'origin unknown' about words like *cocktail* often does not reflect the state of the art. It is rather a comment on lexicographic practice: dictionary makers copy from one another, and endless repetition produces the illusion of consensus. The origin of *cocktail* (to stay with the same example) was clarified decades ago, but this clarification has found its way into very few dictionaries. I assume that bibliographies do not belong to consultants' everyday reading, for the titles could not be more revealing ('The Origin of *Cocktail*'; five of them). The same is true of many other words. Even when no one could offer a fully convincing solution, I found that good suggestions abound. This happened in my investigation of the etymology of *dwarf* belonging to the Germanic protolanguage (in this case my mythological studies required visits to several foreign lands): a tentative suggestion in the first edition of Kluge's dictionary (under *Zwerg*) gave me a clue; the rest, as chess players say, was a matter of technique. Most of the words I will write about will probably remain to some extent obscure, but their origin will become partly 'known' (compare *loom*, above).

5 The depth and breadth of entries in an etymological dictionary

Entries in etymological dictionaries tend to be short (five and more per page; only the best Romance dictionaries are an exception to this rule). Skeat treated some words at greater length, but, in principle, he managed to say all he wanted in the concise version of his great work. The meaning and the pronunciation of a word in a living language can be discovered by turning to native speakers, whereas etymology depends on reconstruction and from the nature of the case is seldom 'final.' The main part of an etymological entry should be discussion, and this is what we find in Walde-Hofmann, Feist, *EWA*, and many other works. However, some dictionaries, including the earliest editions of Kluge, only state the opinions of their authors. Skeat sometimes explained why he disagreed with his predecessors or who inspired his solutions, but, as a rule, he avoided polemic and 'the history of the question.' In English studies, only such dictionaries exist to this day.

Skeat and Kluge were the first authors of reliable etymological dictionaries of English and German. In 1882 and 1884 the solid tradition at their disposal was a few decades old (this holds especially for Skeat's experience, even though the impact on him of early scholarship should not be underestimated). The 'pioneers'' reticence had good reasons. Since the eighties of the 19th century, dozens of dictionaries and innumerable articles and books on word origins in the Indo-European languages have been written, and suggestions on where this or that word came from are countless. Some of them are too speculative, but most merit attention. I believe that a modern dictionary of a language with a rich tradition of etymological research should contain a summary of the views on the origin of every word featured in its pages.

The breadth and depth of the discussion poses various questions, and the answers to them depend on the type of the dictionary and the nature of the material. Feist, the author of an incomparable Gothic etymological dictionary, worked with a closed corpus containing a limited number of items. Since the study of the Gothic language forms the foundation of Germanic and to a certain extent Indo-European philology, every recorded word of that language counts. This is a dictionary oriented toward reconstruction, and Feist showed the place Gothic vocabulary occupies in the entire panorama of Germanic and Indo-European. In performing such a formidable task, he could not always decide where to stop and gave superfluous references. In examining well-preserved dead languages (Sanskrit, Classical Greek, and Latin are the best examples), an etymologist encounters the problems similar to those confronting a student of Modern English, French, or Russian. The stock is huge, with some words belonging to slang and others having extremely low frequency. Yet the idea is that all of them should be given some space, for despite the bulk we face a closed corpus. The vocabulary of a living language is inexhaustible, especially if technical terms (for instance, the names of diseases and drugs, plants, animals, and mechanical gadgets), regional words, and the slang of all epochs are taken into account. I have touched on the problem of choice above. Every lexicographer selects what he or she finds indispensable. The important thing is not to fill the dictionary with what a smart 19th-century reviewer called obstructive rubbish (here I mean not the words of the *papaya* class but an excessive number of dispensable references), though one man's trash is another man's treasure, as evidenced by the dust heaps immortalized by Dickens.

In my work I call dictionaries like Skeat's and Kluge's dogmatic and those by Feist and Walde-Hofmann analytic. Until the middle of the 19th century all etymological dictionaries were analytic: that is, every entry in them contained a summary (polemical, oftentimes vituperative, or neutral) of what has been said about the origin of the word in question. Reliable criteria for tracing a word to its etymon did not exist. Perhaps a Hebrew look-alike would provide a clue, or the source might be Greek or Latin; when those respectable languages failed to yield desired forms, Old English, Old High German, or Dutch came to the rescue. For a long time etymology remained an exercise in imaginative, moderately intelligent guesswork, and every conjecture, however improbable, aroused interest and excitement. The advent of comparative linguistics and the discovery of sound laws made the ‘prescientific’ period in the study of word origins obsolete. Skeat would sometimes criticize Wedgwood or quote Skinner (1671) with approval (Skinner made many good suggestions, some of which even the cautious James A. H. Murray accepted), but, other than that, he did not find it necessary to refer to his predecessors. Kluge and his French contemporaries were even more ‘dogmatic.’

Very soon it became clear that sound laws had their limitations. Onomatopoeic, symbolic, and jocular formations; blends, baby words, taboo and anagrams, inexplicable residual forms (*Restformen*) and hybrid forms (*Mischformen*); suspicious borrowings and substrate words, seemingly invulnerable to sound laws, and quite a few others challenged but did not abolish Neogrammarian algebra. Also, within the framework of that algebra solutions vary widely. Any entry in Feist or von Wartburg looks like a record of a military campaign: all scholars swear by sound laws, but their results are different. For this reason, the 20th century returned to the analytic format. The only philology still recycling dogmatic entries is English, so that the post-Skeat English etymological dictionaries are the least substantive in Indo-European. Presses advertise ever new books in which the information culled from the *OED* is presented as particularly ‘fascinating.’ But the *OED*, all its brilliance notwithstanding, is a historical rather than an etymological dictionary and cannot do for the English language what Vasmer did for Modern Russian or Jan de Vries did for Modern Dutch. My project was motivated by the wish to make a first notch in the dogmatic tradition of English etymology.

I think entries in analytic dictionaries should be of unequal length and breadth. In dealing with *brother* and *eight*, it is sufficient to give a succinct overview of the existing theories (those are numerous!) and a list of the main works in which the reader will find further references. My database contains close to a hundred citations for *God*. It does not mean that a hundred conjectures on the origin of this word have been offered (though there have been more than the two routinely repeated in our dictionaries). The entry should contain a summary of the type suggested above and the titles of the original works. Since, as a matter of principle, I look up every word and its cognates with which I deal in all the editions of all the dictionaries, I mention them in my text, just to alert the reader to the fact that nothing new can be found there in comparison with what has already been said. But when we approach *bird*, *boy*, *girl*, *lad*, *lass*, *cub*, *Cockney*, *ivy*, *oat*, *heifer*, *slang*, *witch*, *yet*, *ever*, and their likes (those are some of the words in English featured in Liberman 2008), that is, seemingly isolated words with unclear connections and of debatable structure (native or borrowed?

simple or compound? ancient or late?), the overview has to be exhaustive. This is where even the opinions of the 'prescientific' etymologists matter, for to break the spell laid on those words by their capricious history, we need all the help there is, and in solving such puzzles modern linguists have few advantages over a resourceful scholar who lived three or two centuries ago.

An analytic etymological dictionary does not run the risk of degenerating into an annotated bibliography, because a specialist who has read everything on the origin of a hard word, thought of what has been said about it, and considered numerous proposals will undoubtedly draw conclusions that will be valuable to other researchers. Such an author will be able to reject obviously wrong connections, point out mistakes in earlier reasoning, wherever possible, combine bits and pieces of previous solutions, and weave them into a coherent whole. Some riddles will defy the most strenuous efforts, for it would be naïve to hope that any single person, even endowed with the talents of Jacob Grimm, Antoine Meillet, or Karl Brugmann, can puzzle out all the inscrutable etymologies. We often lack the required data to come to a persuasive result. Also, etymology is both a science and an art. A good deal in it depends on the knowledge of an obscure dialectal form, on an unusual association, and on serendipity. Every serious article on the origin of a difficult word begins with a glance at previous scholarship, and this allows others, regardless of the solution offered, to pick up where the author has left off. An analytic dictionary is called upon to gather several thousand such articles but in congested form. It may take decades to complete, but the effort is worth the trouble.

6 Conclusion

The author of an etymological dictionary needs a clear view of the work's readership, of the vocabulary to be included, of the balance between the most ancient inherited words and those that have emerged in the full light of history, and of the state of the art and of the market. All those points sound trivial, but this impression is false. The methodology of etymological dictionaries and their reception have been discussed too rarely. The genre of the etymological dictionary has not yet been defined.

> References

Blount, Thomas. (1656). *Glossographia: or, A Dictionary, Interpreting All Such Hard words….* London: Printed by Theo. Newcomb.

The Century Dictionary: An Encyclopedic Lexicon of the English Language. William D. Whitney, ed. New York: The Century Co.

EWA = Albert L. Lloyd, Rosemarie Lühr, and Otto Springer, *Etymologisches Wörterbuch des Althochdeutschen.* Göttingen, Zürich: Vandenhoeck & Ruprecht, (1988-).

Faltings, Volkert F. (1996). 'Bemerkungen und Nachträge zu Frank Heidermanns *Etymologischem Wörterbuch der germanischen Primäradjektive* aus der Sicht des Firesischen.' *NOWELE* 28/29, 103-18.

Feist, Sigmund. (1909). *Vergleichendes Wörterbuch der gotischen Sprache.* Halle a. S.: Max Niemeyer. 2nd ed., (1924; 3rd ed., 1939; 4th ed., by W.P. Lehmann, 1986 (The last two editions, Leiden: E.J. Brill).

Kilianus, Cornelius. (1599). *Etymologicum teutonicae linguae.* Antverpiae: Ex Officina Plantaniana, apud Joannem Mortum.

Kluge, Friedrich. (1883). *Etymologisches Wörterbuch der deutschen Sprache.* Straßburg: Trübner. (Kluge took care of the first ten editions of the dictionary until 1924. The latest edition to date is by Elmar Seebold, 2003.)

Liberman, Anatoly. (1998). 'An Annotated Survey of English Etymological Dictionaries and Glossaries.' *Dictionaries* 19, 21-96.

Liberman, Anatoly. (1999). 'An Annotated Survey of German Etymological Dictionaries and Glossaries.' *Dictionaries* 20, 49-89.

Liberman, Anatoly. (2005). 'Scandinavian Etymological Lexicography.' In Henrik Gottlieb, Jens Erik Mogensen, and Arne Zettersten, (eds.). *Symposium on Lexicography XI: Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002, at the University of Copenhagen.* Lexicographica. Series Maior 115. Tübingen: Niemeyer, 351-67.

Liberman, Anatoly. (with the assistance of J. Lawrence Mitchell). (2008). *An Analytic Dictionary of English Etymology: An Introduction.* Minneapolis, London: University of Minnesota Press.

Liberman, Anatoly (with the assistance of Ari Hoptman and Nathan E. Carlson). (2009a). *A Bibliography of English Etymology.* Minneapolis, London: University of Minnesota Press.

Liberman, Anatoly. (2009b). 'English Etymological Dictionaries.' In A. P. Cowie, (ed.). *The Oxford History of English Lexicography*, vol. 2. Oxford: Oxford University Press, 269-289.

Malkiel, Yakov. (1975). *Etymological Dictionaries: A Tentative Typology.* Chicago: Chicago University Press.

Minsheu, John. (1617). *Ductor in Linguas. The Guide into the Tongues….* London: Published by the author. 2nd ed., London: John Haviland, 1627.

ODDE = C. T. Onions, ed., with the assistance of G. W. S. Friedrichsen and R. W. Burchfield, *The Oxford Dictionary of English Etymology.* Oxford: Clarendon Press, (1966).

OED = James A. H. Murray et al., (eds.). *The Oxford English Dictionary.* 2nd ed., by J. A. Simpson and E. S. C. Weiner. Oxford University Press, (1992). 1992-, online.

Pokorny, Julius. (1959). *Indogermanisches etymologisches Wörterbuch.* Bern and München: Francke.

Skeat, Walter W. (1882). *An Etymological Dictionary of the English Language.* Oxford: Clarendon Press. 2nd ed., 1884; 3rd ed., 1897; 4th ed., 1910.

Skinner, Stephen. (1671). *Etymologicum Linguae Anglicanae….* London: Typis T. Roycroft.

Sledd, James [and] Wilma R. Ebbitt. [1962]. *Dictionaries and That Dictionary: A Casebook on the Aims of Lexicographers and the Targets of Reviewers.* Chicago: Scott, Foresman.

UED = *The Universal Dictionary of the English Language.* Henry C. Wyld, ed. London: Routledge & Kegan Paul Limited, (1932).

Vasmer, Max. (1950-1958). *Russisches etymologisches Wörterbuch.* Heidelberg: Carl Winter.

Vries, Jan de. (1971). *Nederlands etymologisch woordeboek.* Leiden, *etc.*: E. J. Brill (three later editions).

Walde-Hofmann = Walde, Alois. (1938-1954). *Lateinisches etymologisches Wörterbuch.* 3rd ed. by Johann B. Hofmann.

Walde-Pokorny = Alois Walde, *Vergleichendes Wörterbuch der indogermanischen Sprachen.* Julius Pokorny, (ed.). Berlin: Walter de Gruyter, (1927-1932).

Wartburg, Walther von. *Französisches etymologisches Wörterbuch….* Leipzig, Berlin: B. G. Teubner, etc; Basel: Zbinden, (1934-1998).

Wedgwood, Hensleigh. (1859-1865). *A Dictionary of English Etymology….* London: Trubner. 2nd ed., 1872; 3rd ed., 1878; 4th ed., 1888.

Weekley Ernest. (1921). *An Etymological Dictionary of Modern English.* London: John Murray.

> State of the Art of the Lexicography of European Lesser Used or Non-State Languages

ANNE TJERK POPKEMA

'The people who chronicle the life of our language (...) are called lexicographers' (Martin Hardee, blogger in Cyberspace, 2006)

0 Introductory remarks[1]

Language codification and language elaboration ('Ausbau') are key ingredients for raising a lesser used language to a level that is adequate for modern use.[2] In dictionaries (as well as in grammars) a language's written standard may be laid down, 'codified'.[3] At the same time dictionaries make clear what lexical gaps remain or arise in a language. The filling of such gaps – part of language elaboration – will only gain wide acceptance when, in turn, it is codified in a dictionary itself. Thus, both prime categories of language development – codification and elaboration – are hats worn by the same head: the lexicographer's.

Bo Svensén begins the opening chapter of his recent handbook on lexicography by stating that 'dictionaries are a cultural phenomenon. It is a commonplace to say that a dictionary is a product of the culture in which it has come into being; it is less so to say that it plays an important part in the development of that culture.'[4] In the case of lesser used languages, language development may lead to (increased) use in domains that were formerly out of reach because of the dominance – for any number of reasons – of another language. In such instances, language development equals language emancipation. An emancipating language takes on new functions, enters new domains of society and is therefore in need of new terminology. As a result, the language needs new or revised dictionaries that in their turn strengthen the ongoing language emancipation – a virtuous circle with the lexicographer at its very heart. Such a circle may turn vicious just as easily, when emancipatory efforts are unsuccessful

[1] I am grateful to Anne Dykstra, Willem Visser (both: Fryske Akademy Leeuwarden, the Netherlands) and Alastair Walker (Nordfriesische Wörterbuchstelle, Christian-Albrechts-Universität Kiel, Germany) for their comments on an earlier version of this paper.

[2] Cf. Haugen (1966:931)

[3] Cf. Inoue (2006): 'One concrete strategy of minority language revitalization is formal codification through practices of literacy, orthography, dictionary, grammar, or census.'

[4] Svensén (2009:1)

or even absent, and domains and functions are firmly held or taken over by the dominant language. Yet in this case, too, the lexicographer finds himself at the heart of the circle, He documents as much as he can of the language in decline, objectively executing his scientific, descriptive task (although he may well have silent hopes of stopping or slowing down the downwards spiral).

In both capacities, the lexicographer is a language emancipator – whether he wants to or not. For even the fiercest denier of language-ideological influence on his lexicographic activities cannot prevent his strictly objective language description – his dictionary – becoming a tool of language emancipation. Even more so: the more objective, the more 'scholarly' a dictionary of a language is, the more up to date and elaborate it may become, and the more it may be deployed by language movements in their efforts to elevate a language's social status: 'It is certainly a real language – look at this enormous dictionary!' It's an example of the classic paradox of influencing an object of research by studying it objectively.

The capacity of lexicography as a prime emancipatory tool for lesser used languages makes the lexicographer a key figure in the play of language life – and, alas, at times a close spectator of the tragedy of language death. For undeniably, a lexicographer of an endangered language runs the increased risk of finding himself at the language's death bed, meticulously documenting its final gasps of air. Under such circumstances, the lexicographer's task is of equal importance as it is for revitalizing purposes: a language contains centuries, possibly millennia of cultural and ecological information that may be forever lost if it is not documented properly. In 2003, UNESCO language experts stated that 'a language that can no longer be maintained, perpetuated, or revitalized still merits the most complete documentation possible. This is because each language embodies unique cultural and ecological knowledge in it. It is also because languages are diverse. Documentation of such a language is important for several reasons: 1) it enriches the human intellectual property, 2) it presents a cultural perspective that may be new to our current knowledge, and 3) the process of documentation often helps the language resource person to re-activate the linguistic and cultural knowledge.'[5]

Witnessing language extinction certainly is a gloomy perspective, yet for

[5] UNESCO (2003), par. 3.5, at http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf.

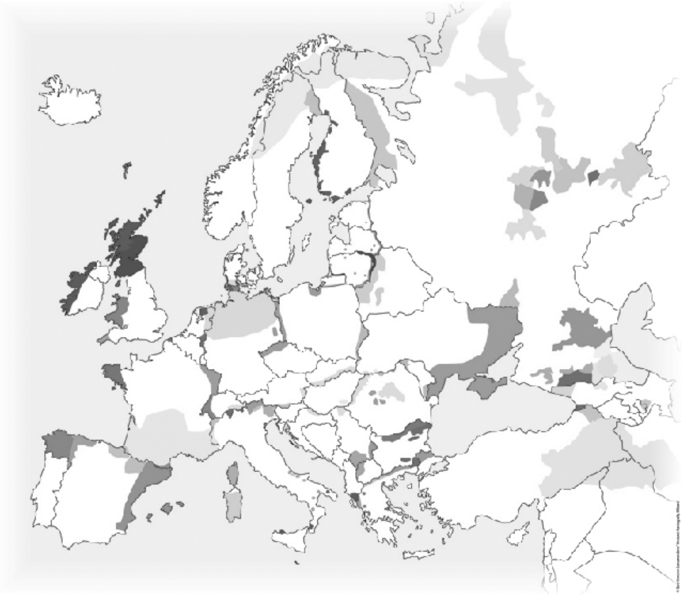
many lexicographers of endangered languages across the globe it is not an entirely unlikely professional destiny. Such lexicographers often operate in relative isolation – the poorer the patient, the less doctors he's likely to see. And, like any friendly small town physician, the lexicographer is often not left unmoved by the decline. For although lexicography of lesser used languages is a full-grown scientific discipline and most lexicographers are well capable of objectively documenting the language concerned – still, in my experience many lexicographers of a lesser used language tend to not to be fully objective towards the language. Not uncommonly, they are either native speakers – and who wishes one's mother tongue to become extinct? – or they have come to appreciate the language they document – and who remains unmoved by a dear friend's good or misfortune?

In my opinion, however, a certain degree of subjectivity with respect to a language's vitality does not imply an unscientific attitude by definition. It may well often be such personal involvement that allows the lexicographer to get close enough to the language speakers to be able to document their language in its entirety. An objective, hard-core scientific attitude may be perceived as 'cold-hearted' by the native speakers, and might lead to less than full openness, which in turn may lead to less than optimal research results – a poorer dictionary.

It would, however, appear that relative professional isolation and personal involvement are more of a rule than an exception in the lexicography of lesser used languages, making exchanging information and experience with fellow-lexicographers useful, even vital at times. In the light of such considerations and, indeed, responsibilities, it is fortunate that the organizing committee of the Euralex International Congress chose to make the lexicography of lesser used or non-state languages the central theme of their 14th gathering.

One important feature of the conference is a survey of the state of the art of the lexicography of such languages. To this end, the organizing committee sent out questionnaires to dozens of institutions and individuals involved in the lexicography of lesser used or non-state languages. In this paper, some of the results of this Euralex 2010 Survey of the Lexicography of European Lesser Used or Non-State Languages (henceforth: Survey) are presented.

In the first section of this paper, I will discuss several methodological aspects of the questionnaire and the Survey. The second section is the core of this paper, providing facts and figures, including indispensable sociolinguistic information on the languages in the Survey and summarized results of various lexicographical aspects. In the third section some implicational statements are made. Since offering and analyzing all the data the Survey has culled in the field of lexicographical practice (and there is quite a considerable amount – it will hopefully be made generally available via the Euralex website) was simply impossible because of the limited space for this paper, I chose to concentrate on some of the results I found particularly interesting in the third section. The fourth section consists of a few final remarks.



Map 1: Geographical distribution of all lesser used or non-state language areas of Europe

1 The questionnaire[6]

European lesser used or non-state languages

In order to give an impression of the scale of European lesser used or non-state languages, perhaps I might quote some figures.[7] Except for Iceland, every European state has at least one linguistic minority (cf. Map

[6] On language surveys – including those on minority languages – and their pitfalls, cf. De Vries (2006).

[7] All numbers of speakers and percentages mentioned in the following sections are approximate, without explicitly stating so each time.

1). The total number of lesser used or non-state European languages is approximately 60, representing 55 million European citizens.[8] Keeping these numbers in mind as well as the above-mentioned importance of lexicography for language development and language emancipation, it seems rather odd that up till now there has been no comparative overview of the lexicographical situation of lesser used or non-state languages in any lexicographical or (socio)linguistic handbook on European languages.[9] There are numerous overviews of the lexicographical state of affairs of individual lesser used or non-state languages as well as overviews of the lexicography of such languages within a single state's borders,[10] yet an overview transcending the state level has, to the best of my knowledge, never been attempted.

Terminology I: 'lesser used or non-state languages'

A particular minefield is the nomenclature in the semantic range of

[8] Cf. Facts and Figures on the website of Mercator European Research Centre on Multilingualism and Language Learning (henceforth: Mercator), at http://www.mercator-research.eu/minority-languages/facts-figures.

[9] Such (socio)linguistic handbooks include: Hinderling/Eichinger (1996); Janich/Greule (2002) (which includes paragraphs on the lexicography of individual languages, in which the most important dictionaries are mentioned); Åkermark e.a. (2006). The 16th online edition of Ethnologue Languages of the World, at http://www.ethnologue.com/web.asp) also provides a lot of sociolinguistic information on virtually all languages of the world. Sociolinguistic data on dozens of European minority languages are available in the Introductions to the Mercator Regional Dossiers. Among these are many on languages represented in this Survey (viz. Asturian (in Spain), Basque (both in France and in Spain), Catalan (both in France and in Spain), West Frisian, Galician (in Spain), Scottish Gaelic, Latgalian, North Frisian, Sami (in Sweden), Sorbian (including Lower Sorbian), Võro and Welsh). All of Mercator s Regional Dossiers are digitally accessible at http://www.mercator-research.eu/research-projects/regional-dossiers).

[10] The outstanding lexicographic handbook Hausmann e.a. (1989-1991) offers excellent overviews for some of the languages present in the Survey (cf. Table 1), viz. Galician (article nr. 181a), Catalan (184), Romansh (190), the Frisian languages (202), the Sorbian languages (210), Basque (226) and the Sami languages (228a). The lexicography of Romance languages is well covered in the monumental LRL (the Romance languages in the Survey that are treated in the LRL are: Friulian (LRL III, art. 217), Romansh (III 233b), Catalan (V, 2 358b), Asturian (VI, 1 408, pp. 688-689) and Galician (VI, 2 417)). Also, the equally monumental ELL offers overviews on the lexicography of many languages in the Survey (Welsh, by Hawke (2006), Frisian, by Bremmer (2006), Nynorsk, by Kulbrandstad/Veka (2006, par. 'Norwegian'), Galician, Asturian, Catalan and Basque (combined in Saurí Colomer (2006)). ELL also contains an overview of lexicographical topics and issues by Hanks (2006) including a short paragraph on 'Dictionaries of Rare and Endangered Languages', providing merely a few general remarks on tools and aims of such dictionaries.

‘language and dialects’ – the linguist’s approach to what constitutes a dialect may differ entirely from that of a sociolinguist or a language policy maker.[11] In the Survey as well as in this paper, the term ‘lesser used or non-state languages’ is used in a sense similar to the term ‘regional or minority languages’ as defined in the *European Charter for Regional or Minority Languages* (henceforth: Charter).[12] Such lesser used or non-state languages include: unique languages in one state (e.g. West Frisian in the Netherlands); unique languages spread over more than one state (e.g. Basque in Spain and France); transfrontier languages that are both minority and majority languages, depending on the state (e.g. Sweden Finnish); non-territorial languages (e.g. Romani or Yiddish); official languages that are lesser used on the whole or part of the state’s territory (e.g. Romansh in Switzerland).[13]

Terminology II: ‘dictionary’ vs ‘wordlist’

Special attention should also be paid to the problem of defining the concepts ‘dictionary’ and ‘wordlist’, which are both key concepts in the questionnaire. The grey area between the two types is both vast and treacherous. Not every informant will distinguish equally sharply or consistently between the two, which is partly due to the fact that ‘dictionary’ does not only denote a specific lexicographical type, but also

[11] Cf. Haugen (1966), esp. pp. 926-927.

[12] See: Charter, Part I art. 1a and art. 3 sub 1, at http://conventions.coe.int/treaty/en/Treaties/Html/148.htm. The term ‘regional or minority languages’ is avoided here since the organizing committee deemed ‘lesser used or non-state languages’ to be less politically charged. Also, seemingly, there is somewhat of a paradox in the Charter definition, as it aims at protecting languages that are ‘different from the official language(s) of that State’. However, it does explicitly protect some lesser used languages that are in fact official state languages. This is the case, for example, with Romansh and Italian in Switzerland and with Swedish in Finland, which all are official state languages, yet nevertheless have been brought up for protection under the Charter by the Swiss resp. Finnish government. Cf., however, footnote 13.

[13] For the latter category, cf. paragraph 51 of the *Explanatory Report* on the Charter (at http://conventions.coe.int/treaty/en/Reports/Html/148.htm): ‘The wording of Article 3 takes account of the position in certain member states whereby a national language which has the status of an official language of the state, either on the whole or on part of its territory, may in other respects be in a comparable situation to regional or minority languages as defined in Article 1, paragraph a, because it is used by a group numerically smaller than the population using the other official language(s).’ For comments on terminology, also cf. the *Explanatory Report*, art. 18. I am grateful to Auke van der Goot, employee of the Dutch Ministry of Interior, for pointing this out to me.

is an umbrella term for several lexicographical types.[14] The latter usage explains the tendency among many publishers (and authors!) to name any lexicographical product 'dictionary', which does not, however, help untangle matters. It is in fact somewhat of a paradox that the number of wordlists (or genres related in magnitude and scope like glossaries, lexica, vocabularies) by far exceeds that of actual dictionaries, yet the title words 'wordlist/glossary/vocabulary/lexicon' are by no means as numerous as the title word 'dictionary', which is obviously all too often used in its umbrella-term capacity.[15] Thus it takes quite a theoretical lexicographer to draw a sharp and consistent line between the lexicographical types, and there is no harm in acknowledging that not all informants have been equally successful in doing so.

Aim and set-up

The main aim of the Survey is to ascertain the current state of affairs in the lexicography of European lesser used or non-state languages and their social and linguistic situation.[16] To this end a questionnaire was compiled by the organizing committee (see Appendix).[17]

The questionnaire consists of three main parts. In the first part, contact data for the individual or organization responsible for filling out the questionnaire are gathered. The second part consists of questions on the sociolinguistic position of the language concerned (e.g. questions on the numbers of speaker (2.3), regions/states in which the language is spoken (2.2), related state languages (2.4), the level of recognition by the national government (2.5), the existence of an official spelling (2.11) and grammar (2.13)). The third part is divisible into two subparts. The first subpart (3.1-

[14] Cf. Hartmann/James (1998), s.v. *dictionary*. For anyone not quite familiar with the distinction between dictionaries proper vs wordlists and the like, I would recommend reading Simpson (1993, esp. pp. 124-125), who provides a very accessible introduction for laymen, taking lexicography of Aboriginal languages as a starting point.

[15] A quick search on international bookselling site Amazon.com rendered 180,327 titles containing the word 'dictionary' and a mere total of 78,290 titles containing either 'wordlist' (385) or 'vocabulary' (32,232) or 'glossary' (13,522) or 'lexicon' (7,185). A Google quick search for the same words (which renders a completely different range of hits, since not only book titles are found), still shows 'dictionary' as the more popular term over the other ones combined (139,000,000 vs 125,430,000 hits).

[16] The Survey was set up in close collaboration with Mercator, which is affiliated to the Fryske Akademy in Leeuwarden, The Netherlands.

[17] The questionnaire in the Appendix represents a slightly adapted version of the original questionnaire. The adaptations merely concern matters of lay-out; the phrasing of the questions has not been altered. The questionnaire can also be downloaded via the Euralex website (http://www.euralex2010.eu/).

3.23) contains questions on aspects of the lexicographical situation of the language (e.g. the existence of monolingual or bilingual dictionaries or wordlists, the way in which they were published, numbers of sales, the existence of lexicographic tools in education). The second subpart (3.24-3.35) is aimed at gathering information on the lexicographic infrastructure (e.g. questions on subsidies for compiling dictionaries (3.26) or on support for setting up and maintaining a lexicographical infrastructure (3.27), the embedding of lexicography in other linguistic research (3.30), the way in which primary sources of the language are organized and processed (3.32-3.33)).

Response

The organizing committee put a lot of effort into contacting as many institutions or individuals involved in the lexicography of lesser used or non-state languages as possible. To this end the Mercator Database of Experts proved especially useful.[18] In the event of languages for which experts were lacking in the Mercator Database of Experts, the committee tried complementing the list of informants by making use of the Mercator Regional Dossiers,[19] which often contain useful contact information.

These efforts led to fully satisfactory results. As mentioned above, European lesser used or non-state languages are extant in all European countries, except for Iceland. Their total number is approximately 60, representing a total of 55 million European citizens. The committee received a total of 26 completed questionnaires on European languages (as well as 4 on non-European languages, see below). The total amount of European languages present in the Survey is 22, covering 19 Mio speakers. Hence, the Survey covers well over 1/3rd of all European lesser used or non-state languages and of the total number of their speakers.

The response shows that, although questionnaires were sent out to well over a dozen of lesser used languages that are a dominant state language elsewhere, only one response came in, namely from Sweden Finnish. By comparison, the response ratio for languages that are not dominant elsewhere is approximately 2:1, with two responses for every lack of response.

[18] At http://www.mercator-research.eu/minority-languages/database%20of%20experts. Tjallien Kalsbeek (Mercator) was especially helpful in compiling the extensive list of informants, and I am grateful to Liesje Haanstra (Fryske Akademy) for collecting and forwarding the completed questionnaires.

[19] Cf. footnote 9.

Although the Survey primarily aimed at gathering data on European lesser used or non-state languages, several informants of non-European lesser used languages were also approached. Response was received for Nivkh (Russian peninsula of Sakhalin, from two informants, independent of each other), for a group of minor East Iranian languages and for South Efate (Polynesian isle of Vanuatu).

The committee received a total of 30 completed questionnaires. Unavoidably, in order to achieve a survey of European lesser used or non-state languages, it proved necessary to make a selection of the questionnaires. Firstly, since a survey of European languages is concerned here, questionnaires on non-European languages were not taken into account (these are, as already mentioned: Nivkh, a group of minor East Iranian languages and South Efate).[20] Also, for some European languages duplicates were handed in, namely on Galician and Catalan. The data from such parallel questionnaires were combined into one new, 'optimized' version. Furthermore, in two instances, questionnaires were filled out for varieties of a language that was already present in the response (viz. Helgolandic (North Frisian) and Algherese (Catalan)). The data from these questionnaires were incorporated into the questionnaire of the language of which they are varieties. Sweden Finnish was also left out. As mentioned earlier, it is the only language which is a dominant state language elsewhere for which a questionnaire was filled out. For its lexicographical infrastructure and output Sweden Finnish may rely on the lexicographical infrastructure of its 'homeland' Finland. This renders it less useful for the purpose of comparing languages that do not have such an infrastructure to rely on.[21]

As a result of the selection criteria, in total 22 questionnaires were deemed suitable for processing in the Survey (see Map 2 for their geographical distribution).

Processing the data

The questionnaire contained over a hundred questions (cf. Appendix). As it turned out, some questions were hard to answer, either because the answering options were insufficient, or because the question itself did

[20] The data these questionnaires contain will not be neglected. In due time Euralex hopes to be able to present the questionnaires on their website.

[21] Actually, Catalan may also count as a language that is a dominant state language elsewhere: it is the only official state language of Andorra. However, the lexicographical epicentre of Catalan is located in Spanish Catalonia. For this reason, Catalan was not excluded.

not (fully) apply.[22] Also, several questions were skipped by a large portion of the informants since they did not have relevant data at their immediate disposal (this applies particularly to several questions in the third section, on URL's and on printing and sales numbers) or since answering them would simply take up too many resources (e.g. extensive bibliographical information). As a result, there were several questions that (might have) rendered unreliable data. They were filtered out for that reason.

As it turned out, some questions were answered incompletely, by a small number of informants or in a heterogenic way. Such questions include the ones on the manner in which dictionaries and wordlists are published (3.25), on their goal or background (3.20) and on (the lack of) non-governmental subsidies (3.26). Making general statements on the basis of such results would not render reliable information. Hence, such questions were left out.

It was rather unfortunate that in some cases the questionnaire did not offer sufficient opportunity for specifying information for individual dictionaries and wordlists. For example, following a list of bilingual dictionaries, the informant was asked whether they are online, which may differ per dictionary. Yet, by offering merely one 'yes/no' choice box (3.6), the informant was unfortunately not provided the opportunity to specify per dictionary. The overall decision was made to regard such questions as answered with 'yes' if the answer applied to any of the listed publications, and with 'no' if nothing was indicated at all.

There is also the above-mentioned matter of terminology: not every informant will use or define key-concepts like 'wordlist' or 'dictionary' in the same way. Quite often, 'dictionary' was used as an umbrella term for any lexicographical type. Of course, questions containing these concepts can hardly be filtered out in a survey on lexicography. Here a rather pragmatic approach was followed: where a questionnaire was deemed to be falsely indicating a wordlist as a dictionary or vice versa (which occurred much less frequently), the answer was corrected silently. It should be noted, however, that the questionnaires were not systematically checked with respect to the dictionaries and wordlists mentioned.

[22] For example: the question on the tendency of distancing (question2.17) is often not relevant for languages that are genetically distant to the dominant language. In the case of (Slavonic) Lower Sorbian in Germany, for instance, this is the case, and therefore distancing is not a relevant issue here (Gunter Spiess (formerly) of the Sorbisches Institut in Cottbus, Germany, in personal correspondence with the author).

There seems to have been some confusion as to whether the questionnaire should be filled out for lexicographical projects the informant is personally involved in or for all projects of the language of the informant's expertise. The latter was the aim of the committee and, although this was not explicitly indicated anywhere, fortunately most informants seem to have filled out the questionnaire accordingly.

Information on protection under the Charter as well as on the level of endangerment according to the UNESCO Atlas of Languages in Danger (henceforth: UNESCO Atlas) was added by myself at a later stage.[23]

2 Facts and Figures

The main part of the questionnaires consisted of questions on lexicographic infrastructure and lexicographical output of the language. In order to evaluate such topics, the languages need to be placed in some sociolinguistic perspective. Numbers of speakers of the languages, protection under the Charter, level of endangerment, countries in which they are spoken: all of these may influence a language's social status, which in turn may have an effect on lexicographical practice and possibilities.
Therefore, the following pages contain several sets of thematically linked results of the Survey. The initial two sets consist of sociolinguistic data. The first set (A) is on the states in which the languages are spoken, the state borders they cross and their numbers of speakers. The second set (B) is concerned with governmental recognition (including protection by the Charter) and the level of endangerment. After this, two sets of facts and figures concerned with lexicography follow: the first one (C) focuses on lexicographic infrastructure and the latter (D) on actual lexicographical output. Section (E) contains two tables showing, respectively, the level of diversity of lexicographical output per language, and the level of use of modern technology in lexicographic practice.

The information offered is restricted to mere facts. What the data might tell us, what they mean or why they are what they are, is the content of paragraph 3 ('Implicational Statements').[24]

[23] At http://www.unesco.org/culture/ich/index.php?pg=00206.

[24] I stress that the data that are listed on the next few pages represent only part of what the combined questionnaires have to offer: as a result of the limited space only a selection of the results can be presented here.

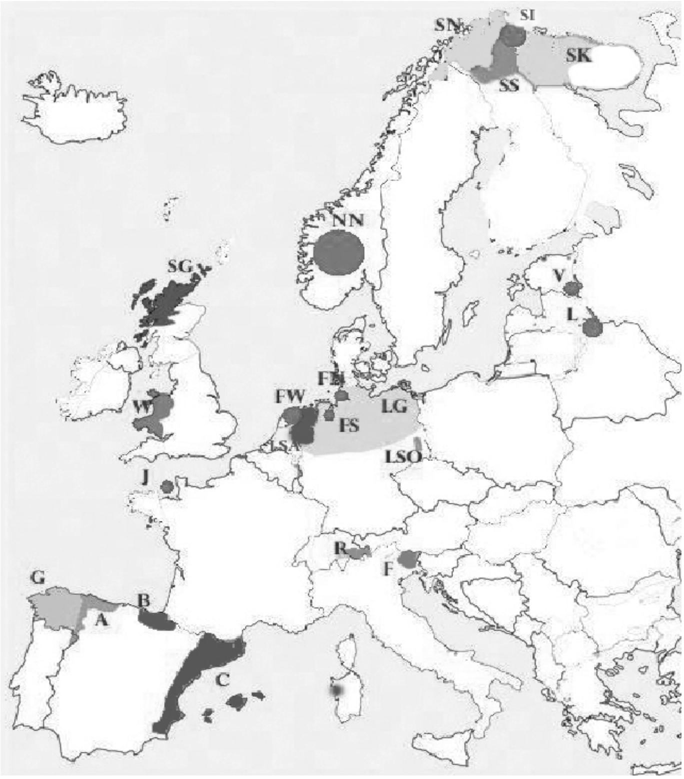
A – Sociolinguistic: States, Transfrontier Languages, Numbers of Speakers

Language (state(s))	**No. of speakers**
(A) Asturian (Spain/Portugal)	350,000-500,000
(B) Basque (Spain/France)	1 Mio
(C) Catalan (Spain/France/Italy/Andorra)	7-9 Mio
(F) Friulian (Italy)	600,000
(FN) North Frisian (Germany)	8,000
(FS) Sater Frisian (Germany)	2,000
(FW) West Frisian (Netherlands)	450,000
(G) Galician (Spain/Portugal)	2 Mio
(J) Jèrriais (UK)	2,600
(L) Latgalian (Latvia)	150,000-200,000
(LG) Low German (Germany)[1]	2.6 Mio
(LSa) Low Saxon (Netherlands)[2]	2.15 Mio
(LSo) Lower Sorbian (Germany)	7,000
(NN) Nynorsk (Norway)	500,000
(R) Romansh (Switzerland)	60,000
(SG) Scottish Gaelic (UK)	60,000
(SI) Inari Sami (Finland)	300
(SK) Kildin Sami (Russia)	300-700
(SN) North Sami (Norway/Finland/Sweden)	30,000
(SS) Skolt Sami (Finland/Russia)	300
(V) Võro (Estonia)	50,000-70,000
(W) Welsh (UK)	600,000

Table 1: languages represented in the Survey (cf. Map 2)
(in italics: protected under the Charter)

A1. Lesser used or non-state languages are extant in all European countries, except for Iceland (cf. Map 1). Their total number is approximately 60, representing a total of 55 Mio European citizens. The total number of languages in the Survey is 22, covering 19 Mio speakers in 15 different countries.

A2. There are 7 transfrontier languages in the Survey: Asturian (Spain/Portugal), Basque (Spain/France), Catalan (Spain/France/Italy/Andorra), Galician (Spain/Portugal), Low German/Low Saxon (Germany/Netherlands),[25] North Sami (Norway/Finland/Sweden) and Skolt Sami (Finland/Russia).



Map 2: geographical distribution of European languages present in the Survey
(for the abbreviations, cf. Table 1)

[25] In the Survey, Low Saxon and Low German are both considered to be transfrontier languages. This does, however, call for some elaboration. Due to increasing influence in the course of the 20th century of Dutch (west of the state border) and High German (east of the border) (cf. Niebaum (2008a) and (2008b, esp. 437-438), a formerly relatively coherent language area diverged into two sets of dialect groups: a Dutch one, Low Saxon, and a German one, Low German. As a result, the Dutch-German state border does constitute a linguistic border nowadays: it divides the larger coherent language area, stretching from the Veluwe region at the heart of the Netherlands all the way east to the German-Polish state border. Nevertheless, these two larger dialect groups do still undeniably constitute a transfrontier dialect continuum, be it with a sharper linguistic division than some decades ago. To illustrate the taxonomical confusion: The UNESCO Atlas (cf. footnote 23) considers 'Low Saxon' (with the alternate name: Low German) to be a language spoken both in the Netherlands and in Germany, as does Ethnologue (cf. footnote 9).

A3. Out of the 19 Mio speakers covered by the Survey, almost 60% are Spanish citizens. With 8 Mio speakers (42% of the total), Catalan is by far the largest language in the Survey. Taking Catalan out of the equation would still mean that 32% of the remaining population represented by the Survey is of Spanish nationality.

A4. Skolt Sami (300 speakers), Inari Sami (300) and Kildin Sami (500) are the languages in the Survey with the smallest numbers of speakers.[26]

A5. In the Survey, the two largest language groups are Germanic and Romance languages: both categories are represented by 6 languages, covering 5.7 Mio (Germanic) and 11 Mio speakers (Romance).[27]

B – Sociolinguistic: Governmental Recognition and Level of Endangerment

B1. All languages in the Survey have in some way been officially recognized by their national governments the only exception is Võro (Estonia).[28] All languages also receive financial support from the government.[29]

B2. Out of the 15 states represented in the Survey by one or more languages, 7 have not ratified the Charter.[30] As a result, Friulian (Italy), Latgalian (Latvia), Kildin Sami (Russia) and Võro (Estonia) are not under its protection. Furthermore, two languages (Jèrriais (UK) and Asturian

[26] The fourth Sami language in the Survey, North Sami, is spoken by 30,000 people.

[27] Germanic: Low German, Low Saxon, West Frisian, Sater Frisian, North Frisian, Nynorsk. Romance: Catalan, Asturian, Galician, Jèrriais, Friulian, Romansh.

[28] However, Võro does receive governmental financial aid for lexicographical purposes, which is seemingly a bit of a paradox. On enquiring about this matter, I received the following comments: 'Võro is the most vivid variant of what is now known as South Estonian. Lately, South Estonian is officially indicated as 'a special modification of Estonian', but not as a language. The Estonian government does acknowledge that South Estonia and the West coast islands are culturally special and in need of support. The Ministry of Culture hosts several programs for cultural support of South Estonian, out of which one program is especially meant for Võro. Furthermore, Võro is supported by funds from the Ministry of Education (as part of the program 'Estonian and National Memory') for lexicographical work. Also, the Võro Institute is a state institution, which receives state funds from the Ministry of Culture.' (Mariko Faster of the Võro Institute in Võru, Estonia, in a personal email to the author, February 2010).

[29] In some cases, this may be a local or regional rather than the national government.

[30] Viz. Andorra, Estonia, France (signed, not ratified), Italy (signed, not ratified), Latvia, Portugal, Russia (signed, not ratified).

(Spain/Portugal)) are spoken in countries which did ratify the Charter,[31] yet were not brought up for protection under the Charter, bringing the total of languages not protected by the Charter to 6.[32] The combined number of speakers of these 6 languages is 1.2 Mio.

Languages	**Level of**	**Level of**	**Languages**
(no. of speakers)	**endangerment**	**endangerment**	**(no. of speakers)**
Asturian (350-400,000)	Definitely endangered	(Not listed)	*Catalan* (7-9 Mio)
Basque (1 Mio)	Vulnerable	(Not listed)	*Galician* (2 Mio)
Catalan (7-9 Mio)	(Not listed)	(Not listed)	*Low German* (2.6 Mio)
Friulian (600,000)	Definitely endangered	(Not listed)	*Nynorsk* (500,000)
Galician (2 Mio)	(Not listed)	Vulnerable	*Basque* (1 Mio)
Inari Sami (300)	Severely endangered	Vulnerable	Latgalian (150-200,000)
Jèrriais (2,600)	Severely endangered	Vulnerable	*Low Saxon* (2.15 Mio)
Kildin Sami (300-700)	Severely endangered	Vulnerable	*Welsh* (580,000)
Latgalian (150-200,000)	Vulnerable	Vulnerable	*West Frisian* (450,000)
Low Saxon (2.15 Mio)	Vulnerable	Definitely endangered	Asturian (350-400,000)
Low German (2.6 Mio)	(Not listed)	Definitely endangered	Friulian (600,000)
Lower Sorbian (7,000)	Definitely endangered	Definitely endangered	*Lower Sorbian* (7,000)
North Frisian (8,000)	Severely endangered	Definitely endangered	*North Sami* (30,000)
North Sami (30,000)	Definitely endangered	Definitely endangered	*Romansh* (60,000)
Nynorsk (500,000)	(Not listed)	Definitely endangered	*Scottish Gaelic* (60,000)
Romansh (60,000)	Definitely endangered	Definitely endangered	Võro (50,000-70,000)
Sater Frisian (2,000)	Severely endangered	Severely endangered	*Inari Sami* (300)
Scottish Gaelic (60,000)	Definitely endangered	Severely endangered	Jèrriais (2,600)
Skolt Sami (300)	Severely endangered	Severely endangered	Kildin Sami (300-700)
Võro (50,000-70,000)	Definitely endangered	Severely endangered	*North Frisian* (8,000)
Welsh (580,000)	Vulnerable	Severely endangered	*Sater Frisian* (2,000)
West Frisian (450,000)	Vulnerable	Severely endangered	*Skolt Sami* (300)

Table 2: Level of endangerment of languages in the Survey according to the UNESCO Atlas (in italics: under protection of the Charter)

[31] For Asturian, this only holds true for Spain. Portugal has neither signed nor ratified the Charter.

[32] In the case of Jèrriais, the following applies: 'The situation for Jèrriais is slightly unclear at present. It seems that all technical obstacles and objections have been answered, and Jersey has received no objection on constitutional grounds from the UK to a ratification on Jersey's behalf. What remains is the political situation in Jersey (...) However, in September 2009 the Minister for Education, Sport and Culture signed an agreement to formalize the remit of L'Office du Jèrriais in terms of the promotion of Jèrriais and of drawing up language plans with States authorities. L'Office du Jèrriais is therefore now tasked with acting as though ratification has been completed.' (Geraint Jennings of L'Office du Jèrriais, in a personal email to the author, February 2010).

B3. Asturian is the only transfrontier language in the Survey that is not protected by the Charter.

B4. Whereas all Germanic languages in the Survey are protected by the Charter, only half of the Romance ones are (viz. Romansh, Catalan and Galician. Not protected are: Asturian, Friulian, Jèrriais).

Not included in the questionnaire was the language's level of endangerment according to the UNESCO Atlas.[33] It contains 5 different levels of language endangerment: vulnerable, definitely endangered, severely endangered, critically endangered, extinct. Table 2 shows how the languages in the Survey are categorized according to the UNESCO Atlas.

B5. Table 2 shows that 9 languages in the Survey are considered either vulnerable or are not listed,[34] whereas the remaining 13 languages are considered to be endangered. Out of these 13, 6 are severely endangered[35] No languages in the Survey are considered critically endangered.

B6. Except for Latgalian, all languages that are not under protection of the Charter are considered either definitely of severely endangered.[36]

C – Lexicographical: Infrastructure

C1. Although all languages in the Survey receive governmental support (cf. B1), for Latgalian (Latvia) and Asturian (Spain/Portugal) this does not

[33] Cf. footnote 23.

[34] Not listed are: Nynorsk, Catalan and Galician, which probably means that the UNESCO Atlas does not consider these to be in any danger. Low Saxon and Low German are treated as one – vulnerable – transfrontier language, which the UNESCO Atlas calls 'Low Saxon'. Its total number of speakers is estimated in the UNESCO Atlas at 4.8 Mio speakers (3 Mio in Germany, 1.8 Mio in the Netherlands). Cf. also footnote 27.

[35] They are: North Frisian, Sater Frisian, Skolt Sami, Kildin Sami, Inari Sami and Jèrriais. For Lower Sorbian a status of 'severely endangered' rather than 'definitely endangered' would seem more suitable; the UNESCO Atlas mentions 'Sorbian', by which both Upper and Lower Sorbian are indicated. However, Upper Sorbian is by far the more vital of the two languages (cf. the article on Sorbian, in: Janich/Greule (2002:290) Lower Sorbian has a number of speakers that is comparable to North Frisian, a language considered to be severely endangered in the UNESCO Atlas.

[36] Quite surprisingly, Latgalian is considered merely vulnerable in the UNESCO Atlas, thus placing it at a level comparable with relatively vital languages like West Frisian, Welsh and Basque. Considering the number of speakers of Latgalian as as well as the (less than optimal) state of governmental recognition according to the Mercator regional dossier on Latgalian (at http://www.mercator-research.eu/research-projects/regional-dossiers/090603.regional_dossier_latgalian_in_latvia.pdf, esp. p. 7-13) , a status of 'definitely endangered' would seem more realistic.

include support for lexicographical work. Since both of these are not protected by the Charter (cf. Table 1), this means that all languages in the Survey that are under the Charter's protection receive governmental support for lexicographical work.

C2. All languages in the Survey use an electronic corpus of primary sources. The only exception is Latgalian.[37]

C3. Languages in the Survey that use dictionary writing software (12 in total) do not use a card index system (anymore), with the exception of West Frisian, Low Saxon and Nynorsk.

C4. There are 10 languages in the Survey that do not use dictionary writing software. With the sole exception of Asturian, they all stem from the UK, from Germany or from the Baltic states. *Mutatis mutandis* all languages from Spain, the Scandinavian region (including Kildin Sami in Russia) and the Netherlands do use dictionary writing software.

D. Lexicographical: Output[38]

D1. All languages in the Survey have bilingual dictionaries and bilingual wordlists proper. Also, all languages have bilingual dictionaries of which the source language is the lesser used or non-state language.

D2. None of the languages in the Survey with less than 300,000 speakers (12 in total) have monolingual dictionaries, with the sole exception of Scottish Gaelic, which with 60,000 speakers does have a monolingual dictionary. Conversely, and therefore again with the exception of Scottish Gaelic, all languages that do have monolingual dictionaries are languages with 300,000+ speakers.

[37] Actually, current projects in Latgalian lexicography do not make use of a card index system either. For Latgalian, two ongoing dictionary projects were reported: one in which former dictionaries are digitalized and converted into a database format (cf. http://www.lu.lv/filol/latgalistica/index_en.htm), and another, private project in which Latgalian texts and audio fragments available to the editors are manually analyzed and processed (http://www.vuordineica.lv/). The second project was pointed out to me by Aleksey Andronov of St Petersburg State University, St Petersburg, Russia, who is personally involved in the first project only.

[38] Naturally funds and, as a result, manpower, constitute (the most) important factors in matters of lexicographical output. Although in the questionnaire there were no questions on these factors, one should bear in mind that differences in lexicographical output are to a large extent influenced by them.

D3. All languages in the Survey that have monolingual dictionaries also have bilingual dictionaries in which the source language is the dominant or official language.

D4. None of the languages that merely have the type of bilingual dictionary in which the source language is the lesser used or non-state language, have monolingual dictionaries.

D5. Languages for which no meta-lexicographical literature is reported in the Survey, include all languages from the UK (Welsh, Scottish Gaelic and Jèrriais), both languages from the Baltics (Võro and Latgalian) and all four Sami languages (Kildin, Inari, Skolt and North Sami). The two remaining languages for which no meta-lexicographical literature is reported, are Sater Frisian and Asturian.

D6. All languages that have dictionaries for pre-school as well as for elementary and secondary school are protected by the Charter.

Some of the results of the Survey have been linked together in order to obtain a better view on overall trends. The following section, E, contains two tables, in which several content-related questions have been brought together.

E – I: Level of Diversity of Lexicographic Output

Naturally an overview of lexicographical output in terms of the number of produced lexicographical products would render a useful impression of the state of the art of the lexicographical situation of each language. However, the questionnaires simply did not offer sufficient data for such an overview: too many informants were not able to provide the extensive bibliographical information that is necessary to produce such an overview.

A nice alternative in order to obtain an impression of the state of the art of the lexicography in lesser used language is to determine the *diversity* of lexicographical output. In order to provide such an overview, I linked together questions that are concerned with the following matters:

- monolingual dictionary (cf. Appendix, question 3.1)
- online monolingual dictionary (3.2)
- online bilingual dictionary (3.6)
- bilingual dictionary in which source language = dominant language[39] (3.9)

[39] Bilingual dictionaries in which source language = lesser used language are extant for all languages in the Survey (cf. D1), thus taking these into consideration is pointless.

- monolingual wordlist (3.11)
- online monolingual wordlist (3.12)
- meta-lexicographical literature (3.31)
- pre-school/elementary school/secondary school dictionaries (0, 1, 2 or 3 points) (3.21-3.23)

  Meeting each of the first 7 criteria renders 1 point, meeting the last criterion renders maximally 3 points. As a result, the maximum total of points is 10, the minimum is 0. The results are listed in Table 3.[40]

Points	**Language**	**No. of speakers**
10	*Basque*	1 Mio
10	*Galician*	2 Mio
9	*Catalan*	7-9 Mio
8	*West Frisian*	450,000
7	Asturian	350,000-500,000
7	*Nynorsk*	500,000
7	*Welsh*	600,000
6	*Lower Sorbian*	7,000
6	*Romansh*	60,000
6	*Scottish Gaelic*	60,000
4	Friulian	600,000
4	Latgalian	150,000-200,000
4	*North Frisian*	8,000
3	Jèrriais	2,600
2	*Low German*	2.6 Mio
2	*Low Saxon*	2.15 Mio
2	*Skolt Sami*	300
2	Võro	50,000-70,000
1	*Inari Sami*	300
1	*North Sami*	30,000
1	*Sater Frisian*	2,000
0	Kildin Sami	300-700

Table 3: level of diversity of lexicographical output (in italics: protected by the Charter)

[40] NB: a relatively low level of diversity in lexicographical output does by no means imply a low level of lexicographical output *per se*. For example, dozens of Low German, Low Saxon or North Frisian dictionaries have been compiled in the course of time, yet all of them bilingual, mostly published on paper, and merely with the lesser used language as the source language, all of which makes for a relatively low score on lexicographical diversity for these languages.

Considering a level of 6 points or higher as indicative of a (relatively) high level of diversity of lexicographical output, and 5 points or less of a (relatively) low level, leads to the following conclusions:

E1. Languages that are under protection of the Charter all show a (relatively) high level of diversity of lexicographical output. Asturian, again, is the exception, scoring relatively well at lexicographical diversity (7 points), although it is not under the Charter's protection.

E2. Lower Sorbian is the only language with less than 50,000 speakers which scores (relatively) high on lexicographical diversity.

E3. Lower Sorbian, with 6 points, is also the only language from Germany that scores (relatively) high on lexicographical diversity.

E4. None of the languages that show (relatively) low lexicographical diversity have a monolingual dictionary or wordlist, with the exception of Latgalian, which does have a monolingual wordlist.

E. II: Level of Use of Modern Technology in Lexicographical Practice

In order to give an impression of the use of modern technology in the lexicographic practice of the languages in the Survey, I devised a system comparable to the one used for Table 4. Here, the presence/absence of the following criteria was taken into account:

- online monolingual dictionary (3.2)
- online bilingual dictionary (3.6)
- online monolingual wordlist (3.12)
- online bilingual wordlist (3.16)
- use of dictionary writing software (3.33)
- use of an electronic corpus (3.32)

The criteria can be divided into two subcategories: publishing (the first 4) and production (the latter 2). Again, the presence of each criterion merits a single point. As a result, the maximum total of points is 6 and the minimum is 0. The results are in Table 4, which also shows the subdivision of the criteria.

If a score of 4 to 6 points indicates a (relatively) high level of use of technology in lexicographic practice, and 3 or less points is indicative of a (relatively) low level, Table 4 leads to the following conclusions:

Language	Publishing	Production	Total
Galician	4	2	6
Basque	3	2	5
Catalan	3	2	5
Friulian	2	2	4
Welsh	3	1	4
Asturian	3	1	4
Low Saxon	2	2	4
Inari Sami	1	2	3
North Sami	1	2	3
Romansh	1	2	3
Skolt Sami	1	2	3
Võro	2	1	3
Nynorsk	1	2	3
Sater Frisian	2	1	3
West Frisian	1	2	3
Kildin Sami	0	2	2
Jèrriais	1	1	2
Lower Sorbian	1	1	2
Low German	1	1	2
Scottish Gaelic	1	1	2
Latgalian	1	0	1
North Frisian	0	1	1

Table 4: level of use of modern technology in lexicographical practice (in italics: protected by the Charter)

E5. All Iberian languages show a (relatively) high level of use of technology in lexicographical practice, while all languages in Germany show a (relatively) low level.

E6. A total of 13 languages in the Survey score either 1 (11 languages) or 0 (North Frisian and Kildin Sami) out of a possible 4 points on online publishing. Galician is the only language with the maximum score of 4. The remaining three Iberian languages score 3 out of 4 on online publishing.

E7. None of the languages in Germany, in the UK or in the Baltic states use dictionary writing software. Except for Asturian, all other languages do make use of dictionary writing software

3 Implicational statements

At this point, I would like to comment briefly on some of the results that have been listed above, by discussing what they might imply. The results to be commented on reflect matters that I found especially striking or for which I assume that they might be of special interest to lexicographers of lesser used or non-state languages. Anyone else might have picked other topics to comment on – and everyone is encouraged to do so. The data of the questionnaires will hopefully be made available via the Euralex website, in order to put the data to the widest use possible.

First of all, it was interesting – although rather unfortunate – to see that, with the exception of Sweden Finnish, there was no response whatsoever from languages that are a dominant state language elsewhere (cf. *Response*). I suppose this will have something to do with the fact that such languages may rely on the lexicographical infrastructure of the language's 'homeland'. The number of lexicographical products especially designed for the unique situation of such languages is often quite small or even absent. There will, however, be linguistic peculiarities in such languages that should be lexicographically documented, and information on this would have enabled an interesting comparison with non-state languages. Nevertheless, participating in this Survey by filling out a rather extensive questionnaire may not have seemed relevant or worthwhile for languages that are a dominant state language elsewhere.

Another matter I should like to draw attention to is the level of use of modern technology in lexicographical practice. It seems to me that modern-day technology provides extremely helpful tools for compiling and publishing dictionaries. If nothing else, publishing costs can be cut down to the bare minimum in the case of online dictionaries and wordlists. Lexicographers of lesser used languages frequently have to make do with a rather humble amount of money – and this especially holds true for very small language communities. Online publishing therefore seems like a welcome addition to traditional, paper publishing. It is quite surprising, then, to see that among the top 10 of languages in the Survey with respect to online lexicographical publishing, not one of the 8 languages with less than 10,000 speakers can be found.

Online publishing may occur in many ways.[41] Uploading a PDF-file

[41] At this point, I should like to draw attention to the Online Bibliography of Electronic Lexicography (OBELEX), set up by the Institut für Deutsche Sprache (http://hypermedia.ids-mannheim.de/pls/lexpublic/bib_en.ansicht).

of a short wordlist is one of these, which primarily has the benefits of wide accessibility and low publishing costs over its printed equivalent. The user of such an online wordlist will, however, hardly benefit – in fact, for reasons I will not elaborate on at this time, he might even prefer the printed version. A completely different way of online lexicography, however, is compiling a scholarly dictionary online from scratch and publishing it with a keen search engine on a website. In this case the user is offered many advantages (although, admittedly, again also some disadvantages) to printed dictionaries, including easy access to the corpus on which the dictionary bases. Factors like magnitude of the language community and governmental recognition will be of influence on what medium a lexicographer chooses, since such factors for a considerable part determine the quintessential factor for any lexicographical endeavor: funds. Smaller languages that have not been officially recognized may only have modest funds at their disposal and therefore, quite understandably, be forced to use the less innovative method. Nevertheless, one must also acknowledge that not every lexicographer or lexicographical institute is equally keen on innovation.

Furthermore, it turns out that the lexicographic situation of the Iberian languages in the Survey (viz. Asturian, Basque, Catalan and Galician) is, in comparison to most other languages in the Survey, very well-developed. Both in Table 3 (on lexicographical diversity) and in Table 4 (on the use of modern technology), all four Iberian languages are in the top 6. The reason for this will be the fact that after decades of oppression by the Franco regime, in 1978 the Spanish government drew up a new constitution in which autonomous communities were created. These communities had the right to make a lesser used language co-official along with Spanish (or 'Castilian', as it is often called). Several language movements in Spain ceased this opportunity and, as a result, the lexicography of lesser used languages benefited from these developments in the post-dictatorship period.

In contrast to the Iberian languages, Table 4 (on the use of modern technology) shows relatively low scores for all lesser used languages in Germany. This is surprising to me since, for one, Germany has a venerable lexicographic tradition, which is also expressed by a vast number of excellent dictionaries that have been compiled for the languages in Germany that are in the Survey; and also since in my experience German lexicography (that is: lexicography of modern High German and historical lexicography in Germany) is actually quite keen on innovation.[42] Therefore, the relatively low scores on use of modern technology can hardly be a matter of lack of interest in innovation in the German lexicographical field. Could perhaps the fact that, except for North Frisian, the lesser used languages in Germany that are in the Survey do not receive any subsidies for setting up and maintaining a lexicographical infrastructure account for something? At any rate, one must conclude that, notwithstanding the massive lexicographical output of the lesser used languages in Germany and the high average quality of this output, the lexicographical practice of these languages is not very innovative.

I was also struck by the (relatively) low level of diversity of lexicographical output of languages with a strong lexicographical tradition such as Low Saxon and, again, North Frisian or Low German (cf. Table 3). Both Low Saxon and Low German are quite large language communities, both with over 2 Mio speakers. In the case of many smaller language communities, a lack of lexicographical diversity may, quite prosaically, be due to a low lexicographical output, which in turn is due to lack of lexicographers, which is probably ultimately caused by lack of funds. In the case of the three languages mentioned, however, the lexicographical output in itself is quite large. What may have caused the relatively low level of lexicographical diversity in this considerable output?

I suppose an important underlying reason for this may be that these languages have strong dialectal differentiation. As a result, they are at risk of having standardization issues: what dialect should become the (basis of the) language's written standard? Unless there is a dialect among them which is quite clearly the most frequently used one, or which undeniably has the highest social status, each dialect may be expected to desire becoming the (basis of the) standard. The lack of a standard makes producing a monolingual dictionary a complicated, delicate matter, as a comparable question arises: which dialect should be both the source language and the metalanguage of the dictionary?

In the case of bilingual dictionaries in which the dominant language is the source language, two strategies are possible when a commonly accepted standard for a lesser used language with strong dialectal differentiation is lacking: either the target language comprises of several or even all (main) dialects, showing some of or the complete range of dialectal variation for each entry; or the target language is a selected dialect of the language.

[42] Cf. Popkema (2010).

In the first event, the listing of some or even all dialectal variants would involve huge amounts of time and (therefore) money, practical problems which often cannot be overcome. In the second event, again, there's the issue of what dialect should be the target language.

This problem is the more relevant since it influences overall text production in the lesser used language. In the case of lesser used languages, bilingual dictionaries in which the dominant language is the target language are often used by the language community as production dictionaries: as a rule, speakers of a lesser used language are insufficiently capable of writing it, yet do have sufficient writing skills for the dominant language. As a result, the strategy they use for writing a text in the lesser used language – their mother tongue – is looking up the word they want to write in a bilingual dictionary of which the source language is the dominant language, and then see how it is written in the lesser used language – a spell checker, if you like. Therefore, in the case of lesser used languages a bilingual dictionary in which the dominant language is the source language and the lesser used language is the target language is of vital importance for language production. Quite interestingly, in the case of speakers of a lesser used languages the strategy for writing their mother tongue is exactly the same strategy they would embrace when writing in a foreign language. At any rate, in this way the lack of a standard blocks the compiling of bilingual dictionaries in which the dominant language as the target language, which in turn blocks overall text production in the lesser used language. Thus a further diminishing of the language's social status is set into motion, placing the lesser used language in a vicious circle or sustaining the downward spiral of language decline.

Naturally, the lack of a standard in the case of languages with many sub-dialects does not mean no dictionaries are made. Quite the opposite: the existence of the sub-dialects merits the production of separate bilingual dictionaries of which each (main) dialect is the source language. This may lead to a considerable number of such bilingual dictionaries, yet not to much lexicographical diversity. And this is exactly what we see in the case of North Frisian, Low Saxon and Low German: all of them have several (main) dialects, but not a unified written standard, and all (main) dialects have their own bilingual dictionary, yet no monolingual ones or bilingual ones in which the dominant language is the source language and the lesser used language is the target language.

Conversely, there are several examples of languages that likewise have numerous sub-dialects, yet do not show the same lack of lexicographical diversity and are at the very top of the list in Table 3: Basque, Catalan and Galician. These are languages that have strived for a written standard that supersedes the language's sub-dialects. The scope of this article does not allow me to elaborate on why or how these languages were successful in establishing a written standard while others were not – suffice to say that because of various efforts by the language movements, a written standard was established, which opened the door for a wide range of monolingual and bilingual dictionaries and wordlists. These in turn make text production in the lesser used language easier, which may further enhance the language's emancipation – a virtuous circle as opposed to the vicious circle of some of the languages without a widely accepted written standard.

In short, a (relatively) low level of diversity of lexicographical output may be caused by strong dialectal differentiation that, in turn, may obstruct the acceptance of a widely accepted written standard, which is an important factor in language emancipation.[43]

4 Final remarks

'One of the most important issues facing humankind today is the rate at which our languages are dying. On present trends, the next century will see more than half of the world's 6800 languages become extinct, and most of these will disappear without being adequately recorded. An important first step in slowing down or reversing this process is to document the language in the form of a dictionary.'[44] In the light of such considerations, the lexicographer's contribution is of vital importance to documenting and preserving global cultural heritage. Therefore, it is my sincere hope that the results of the Survey – not just the results presented in this paper, but all information the combined questionnaires offer – will be of help and support to lexicographers across Europe and around the globe. Many problems lexicographers run into are of a more or less universal nature in the lexicography of lesser used languages and stories of success or failure may provide useful guidelines for lexicographical

[43] Cf. Haugen (1966:931): 'These categories [i.e. standardization and utilization in writing, ATP] suggest the path that 'underdeveloped' languages must take to become adequate instruments for a modern nation.'

[44] Sarah Ogilvie (University of Cambridge) in her introductory remarks to the *Endangered Languages and Dictionaries Survey* that was recently set up (at http://www.lucy-cav.cam.ac.uk/pages/the-college/people/sarah-ogilvie/elad1.php). Let me take this opportunity to encourage all readers to take note of this project, and all lexicographers of lesser used languages to fill out this survey, which really is for their own benefit.

planning. Yet I also hope that the Survey may offer some form of mental support for lexicographers of severely endangered languages who are personally moved by the language's decline. For them, their work is often as tragic as it is important.

Naturally, there is a lot more to the lexicography of lesser used languages than the Survey or this paper can treat. For example, aspects of language movement and ideology are undeniably of great influence on a language's social status, on the funds the language has at its disposal and, as a result, on its lexicographical situation. Studying the interaction of language movement and lexicography would no doubt render interesting results. Furthermore, the historical background of the lexicographical tradition and infrastructure of the languages deserves proper attention. And of course, comparing the results of the Survey with the lexicography of non-European languages would be extremely interesting.

These are all matters that deserve proper attention, for which the limited space for this paper is not sufficient. I encourage every scholar that is interested in the subject matter of the Survey to visit the Euralex website, study the results of the questionnaires and use them as a basis for their own research. Hopefully, on the 15th Euralex Congress the Survey will be the subject of a paper once more. I'd like to end this Survey by thanking all lexicographers that took the time to fill out the extensive questionnaire. It is because of their efforts and input that this Survey came to be, and its results are meant to support them in their important work.

> References

Åkermark, Sia Spiliopoulou, e.a. (eds.). (2006). *International Obligations and National Debates: Minorities Around the Baltic Sea.* Mariehamn.

Bremmer Rolf H. Jr. (2006). 'Frisian Lexicography.' In *ELL* (web-based).

De Vries, John de (2006). 'Language Surveys.' In *ELL* (web-based).

Brown, Keith (ed.). *Encyclopedia of Language and Linguistics*, 2nd Edition. Oxford (2006) (Also at: http://www.sciencedirect.com./science/referenceworks/9780080448541)

Hanks, Patrick (2006). 'Lexicography: Overview.' In *ELL* (web-based).

Hartmann, Reinhard K.K.; Gregory James (1998). *Dictionary of Lexicography*. London (1998).

Haugen, Einar (1966). 'Dialect, Language, Nation.' In *The American Anthropologist* 68/4 (1966). 922-935.

Hausmann, Franz. J. e.a. (eds.). (1989-1991). *Wörterbücher/Dictionaries/Dictionnaires. Ein internationales Handbuch zur Lexikographie/An International Encyclopedia of Lexicography/ Encyclopédie internationale de lexicographie* (3 Vols.). Berlin/New York (1989-1991) (*Handbücher zur Sprach- und Kommunikationswissenschaft* 5/1-3).

Hawke, Andrew (2006). 'Welsh Lexicography.' In *ELL* (web-based).

Hinderling, Robert; Ludwig M.Eichinger (1996). *Handbuch der mitteleuropäischen Sprachminderheiten.* Tübingen.

Holtus, Günter; Michael Metzeltin; Christian Schmitt (eds.). (1998-2005). *Lexikon der romanischen Linguistik* (8 Vols.). Tübingen.

Inoue, Miyak (2006). 'Standardization.' In *ELL* (web-based).

Janich, Nina; Albrecht Greule (2002). (eds.). *Sprachkulturen in Europa. Ein internationales Handbuch.* Tübingen.

Kulbrandstad, Lars A.; Olav Veka (2006). 'Scandinavian Lexicography.' In *ELL* (web-based).

LLR. See Holtus, e.a.

Niebaum, Hermann (2008a). 'Het Oostnederlandse taallandschap tot het begin van de 19e eeuw.' In Jurjen van der Kooi, e.a. (eds.), *Handboek der Nedersaksische Taal- en Letterkunde.* Assen (2008), 52-64.

Niebaum, Hermann (2008b). 'Het Nederduits.' In Jurjen van der Kooi, e.a. (eds.), *Handboek der Nedersaksische Taal- en Letterkunde.* Assen (2008). 430-447.

Popkema, Anne T. (2010). 'Eine Perspektive der altfriesischen Lexikographie? Zu einem Online-Belegwörterbuch des Altfriesischen.' In Piter Boersma e.a. (eds.), *Philologia Frisica anno 2008.* Leeuwarden (expected in 2010).

Saurí Colomer, Roser (2006). 'Spain: Lexicography in Iberian Languages.' In *ELL* (web-based).

Simpson. J. (1993). 'Making Dictionaries.' In Michael Walsh/Colin Yallop (eds.), *Language and Culture in Aboriginal Australia.* Canberra (1993). 123-144.

Svensén, Bo (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge.

> Web-based (April 2010)

ELL
Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd Edition. Oxford (2006).
http://www.sciencedirect.com./science/referenceworks/9780080448541
Endangered Languages and Dictionaries Survey
http://www.lucy-cav.cam.ac.uk/pages/the-college/people/sarah-ogilvie/elad1.php
Ethnologue Languages of the World (16th Edition)
http://www.ethnologue.com/web.asp
European Charter for Regional or Minority Languages
http://conventions.coe.int/treaty/en/Treaties/Html/148.htm
European Charter for Regional or Minority Languages: Explanatory Report
http://conventions.coe.int/treaty/en/Reports/Html/148.htm
Euralex
http://www.euralex2010.eu/
Mercator Facts and Figures
http://www.mercator-research.eu/minority-languages/facts-figures

Mercator Regional Dossiers

http://www.mercator-research.eu/research-projects/regional-dossiers)

Mercator Database of Experts

http://www.mercator-research.eu/minority-languages/database%20of%20experts

OBELEX

http://hypermedia.ids-mannheim.de/pls/lexpublic/bib_en.ansicht

UNESCO Atlas of Languages in Danger

http://www.unesco.org/culture/ich/index.php?pg=00206

UNESCO (2003)

UNESCO, 'Language Vitality and Endangerment'. UNESCO Ad Hoc Expert Group on Endangered Languages. Document submitted to the *International Expert Meeting on UNESCO Programme Safeguarding of Endangered Languages* Paris, 10–12 March 2003

http://www.unesco.org/culture/ich/doc/src/00120-EN.pdf

> Appendix

Questionnaire concerning lexicography of European lesser used languages

Introduction

The Fryske Akademy in Leeuwarden, The Netherlands does research on Frisian and Friesland. Over the years the Fryske Akademy has published a wide range of Frisian dictionaries. The Mercator Research Centre at the Fryske Akademy studies the position of lesser-used languages in education, gives information on national and regional education systems and provides the latest statistics regarding lesser-used languages in education in the European Union.

The European Association for Lexicography (Euralex) holds biennial congresses, attended by several hundred people, where refereed papers are presented on a large variety of topics relevant to its members' interests. The Fryske Akademy will host the next congress, which will be held in Leeuwarden/Ljouwert, the Netherlands, from 6 – 10 July 2010. See the congress web site for more information. One of the special features of the 2010 conference is its focus on the lexicography of lesser used non-state languages. In preparation of the conference we would like to learn more about the state of the art of the lexicography of the individual lesser used languages in Europe and also about their social and linguistic situation. To that end we have compiled the following survey that we hope you will complete. The results of the survey will be presented at the Euralex conference in 2010.

1 Contact data of the informant

1.1 last name:
1.2 first name:
1.3 address:
1.4 city:
1.5 country:
1.6 e-mail:
1.7 affiliation:
1.8 mailing-list:
1.9 website:

2 Social and linguistic situation of the lesser used language

2.1 what language are you/is your institute dealing with?
local name:
English name:

2.2 in what region(s) and country/countries is the language spoken?

………………………………………………………………………………

2.3 how many speakers does the language have?

………………………………………………………………………………

2.4 to which larger national language is the language linguistically related?

………………………………………………………………………………

2.5 has the language been recognized by the national government?
❒ no ❒ yes

– if yes, in what way?

………………………………………………………………………………

2.6 does the language receive support to survive?
❒ no
❒ yes, from

………………………………………………………………………………

2.7 does this include support for lexicographical work? ❒ no ❒ yes

2.8 do the national authorities consider lexicography as a means for language maintenance? ❒ no ❒ yes

2.9 do the local authorities consider lexicography as a means for language maintenance? ❒ no ❒ yes

2.10 does the language have an organic place in
– education ❒ no ❒ yes
– the media ❒ no ❒ yes
– other, namely

………………………………………………………………………………

2.11 does the language have an official spelling? ❒ no ❒ yes

2.12 if yes, who determines the spelling rules?

………………………………………………………………………………

2.13 are there grammars of the language? ❒ no ❒ yes

2.14 if yes, do the dictionaries (if any) follow the available grammars?
❒ no ❒ yes

2.15 do you consider the lexical distance to the official language
– considerable? ❒ no ❒ yes
– marginal? ❒ no ❒ yes

2.16 do you consider the syntactic distance to the official language
– considerable? ❒ no ❒ yes
– marginal? ❒ no ❒ yes

2.17 is there a tendency among educated speakers of the language to use words and/or constructions that most clearly show the differences with the dominant language (distancing)? ❒ no ❒ yes

3 Lexicographic situation of the language

3.1 are there monolingual dictionaries of the language? ❒ no ❒ yes

– if yes, please give the title(s) and the year(s) of publication
....

3.2 are they on-line? ❒ no ❒ yes
– if yes, please give the URL's:
....

3.3 are they on paper? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.4 are they on CD-ROM? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.5 are there bilingual dictionaries of the language? ❒ no ❒ yes
– if yes, please give the title(s) and the year(s) of publication:
....

3.6 are they on-line? ❒ no ❒ yes
– if yes, please give the URL's:
....

3.7 are they on paper? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.8 are they on CD-ROM? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.9 is the source language of the dictionaries
– the lesser used language? ❒ no ❒ yes
– the standard/dominant language? ❒ no ❒ yes

3.10 is the target/explanatory language a widely used language like English in order to reach an international audience of linguists and/or lexicographers? ❒ no ❒ yes

3.11 are there monolingual wordlists of the language? ❒ no ❒ yes
– if yes, please give the title(s) and the year(s) of publication:
....

3.12 are they on-line wordlists? ❒ no ❒ yes
– if yes, please give the URL's:
....

3.13 are they paper wordlists? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.14 are they on CD-ROM? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.15 are there bilingual wordlists of the language? ❒ no ❒ yes
– if yes, please give the title(s) and the year(s) of publication:
....

3.16 are they on-line wordlists? ❒ no ❒ yes
– if yes, please give the URL's:
....

3.17 are they paper wordlists? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.18 are they on CD-ROM? ❒ no ❒ yes
– if yes, how many copies were
– printed? ....
– sold? ....

3.19 is the source language of the wordlists
– the lesser used language? ❒ no ❒ yes
– the standard/dominant language? ❒ no ❒ yes

3.20 what is the goal/background of the dictionaries/wordlists?
❒ mainly scientific?
❒ mainly practical?
❒ mainly educational?
❒ mainly touristic
❒ other? Namely:
....

3.21 are there pre-school picture dictionaries? ❒ no ❒ yes
– if yes, please give the title(s) and the year(s) of publication:
....

3.22 are there dictionaries for elementary schools? ❒ no ❒ yes
– if yes, please give the title(s) and the year(s) of publication:
....
– if yes, how many copies were
– printed? ....
– sold? ....

3.23 are there dictionaries for secondary schools? ❒ no ❒ yes
– if yes, please give the title(s) and the year(s) of publication:
....
– if yes, how many copies were
– printed? ....
– sold? ....

3.24 are the dictionaries/wordlists a result of
– a private initiative? ❒ no ❒ yes
– an institutional initiative? ❒ no ❒ yes

3.25 are the dictionaries published
– privately? ❒ no ❒ yes
– by professional publishers? ❒ no ❒ yes
– on-line? ❒ no ❒ yes
– if yes, please give the URL's:
....

3.26 are there any subsidies to facilitate the publication of dictionaries? ❒ no ❒ yes

3.27 are there any subsidies to help setting up and maintaining a lexicographical infrastructure? ❒ no ❒ yes

3.28 have there been studies concerning the dictionary's target group and its needs? ❒ no ❒ yes

3.29 do you experience difficulties distributing lexicographical products? ❒ no ❒ yes
– if yes, please specify:
....

3.30 is the lexicography of the language embedded in other linguistic research? ❒ no ❒ yes

3.31 are there (theoretical) publications on the lexicographical practice? ❒ no ❒ yes

3.32 do you work with
– a card-index corpus? ❒ no ❒ yes
– an electronic corpus? ❒ no ❒ yes

3.33 do you work with a dictionary compilation program?
❒ no
❒ yes, a custom made program
❒ yes, a commercial program, namely:
....

3.34 do you consult fellow lexicographers of lesser used languages? ❒ no ❒ yes

3.35 would you consider it useful to have an on-line forum for lexicographers of lesser used languages? ❒ no ❒ yes

> Dictionaries and Second Language Acquisition

PAUL BOGAARDS

1 Introduction

For a number of reasons dictionaries and second or foreign language acquisition can be thought of as forming a fine pair. Language learners all over the world have dictionaries and use them regularly. Whenever they travel to the country where the other language is spoken, they tend to take a dictionary with them, not a grammar book (cf. Bogaards 1996). And this is understandable because finding your way to the railway station without knowing how such a place is called in the other language is nearly impossible whereas not knowing the grammatical structures of the correct sentences that would be needed in such a situation only makes the communication a bit harder. Besides, even those non-native speakers who have an almost perfect command of the grammar of the language continue to be at a loss for words on many occasions. Especially collocations are a constant challenge for all those who were not raised in the foreign language.

The nineteenth century philosopher Wilhelm von Humboldt (1797 – 1835) distinguished a number of levels on which language can be approached: the 'äussere Form' and the 'innere Form'. The most external aspect of language is, according to him, pronunciation, followed by vocabulary and morphology, whereas the more internal aspects are, in descending order: surface syntax, deep syntax, and semantics (cf. Muysken 2004). Now, what seems to be just a superficial feature, pronunciation, happens to be at the same time the aspect that first of all strikes the native speaker when he encounters a foreigner. Advanced learners of a second language are easily recognized as such because of their foreign accent. But in many cases it is also their choice of words that betrays them, more so than the errors they make in morphology or (surface) syntax. As is well known, people are very sensitive when it comes to variation in pronunciation (cf. Guiora 1972) and learners often feel quite embarrassed when they don't find the words they need, whereas both native speakers and language learners are more tolerant when morphological or syntactic errors are made.

In this paper I will concentrate on the place the lexicon occupies in second or foreign language acquisition (hereafter SLA) and on the place dictionaries occupy in that process. In section 2 I will try to make clear how the lexicon functions in the act of speaking. I will comment on the model that was proposed by Levelt (1989) and that has been adapted to the specific context of SLA by several scholars. In section 3 I intend to give an overview of how lexical aspects of SLA are studied in the field of applied linguistics. In section 4 the perspective will change and the field of (meta)lexicography will be examined in order to see to what extent the second language learner is taken into account. The last section will be devoted to what is known about vocabulary acquisition and about the role dictionaries can play in that context.

2 The lexicon in language use

In his groundbreaking book *Speaking. From Intention to Articulation* Pim Levelt (1989) presents the speaker as an information processor. When we speak we try to get something, a message, an emotion or anything else, through to someone else (or to ourselves for that matter). In order to understand what is happening, Levelt elaborated the model that is presented in figure 1. In the words of Levelt (1989:9)

> Talking as an intentional activity involves conceiving of an intention, selecting the relevant information to be expressed for the realization of this purpose, ordering this information for expression, keeping track of what was said before, and so on.

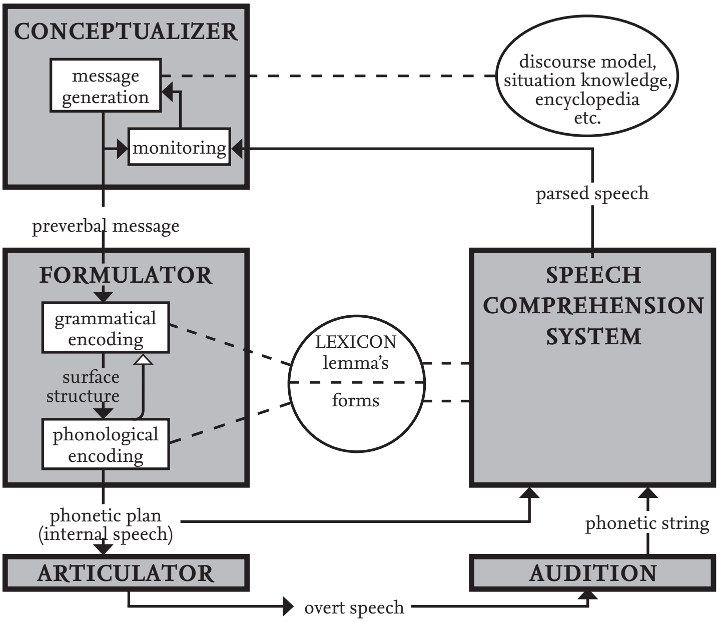


Figure 1. Levelt's model of the speaker as information processor (Levelt 1989:9)

The first stage of the act of speaking is that of the intention, the conceptual stage, where a message is generated and monitored. At this stage the discourse model plans the kind of message to be conveyed; the knowledge one has of the actual situation will influence the form in which the message will be presented, and encyclopaedic knowledge will determine the content of what will be expressed. The preverbal message that is the result of this conceptual stage is then, piecemeal, sent to the formulator. The formulator is heavily dependent on the lexicon, which is represented in the model as the central linguistic module. Once the right words have been chosen, the grammatical form can be determined and the morphological form will follow. These linguistic elements are then, again piecemeal, sent to the articulator which creates an audible form. In the whole process, feedback is crucial: not only do we monitor the content of what we intend to say, adapting our intervention to the supposed knowledge of the interlocutors as well as to their reactions, but the interaction between the conceptual level and the linguistic level, between syntax and morphology, between the linguistic level and the locomotor systems is constantly in action.

An example may make this very intricate procedure clearer. Let us take a simple event we want to talk about: an object 'BOOK' passes from one person, Albert, to another, Bernard. Depending on the state of the discourse so far, we can choose to say something about the book, about Albert, about Bernard, or about the action. This may result in sentences like 'The book was given by Albert to Bernard', 'It was Albert who gave the book to Bernard', or 'By the way, did you know that Bernard received it only yesterday?'. It is obvious that in the last case, there had already been spoken about the book; as to the sender of the book, either this aspect had been part of the conversation so far, or it is shared knowledge between the two participants, or else the speaker does not deem it necessary to mention it.

What is essential in the model is that at the conceptual stage, a virtual message like

'BOOK' – A -> B

has first of all to be put into words. So, if on the level of the preverbal message, it has been decided that the 'BOOK' is the topic of the speech act, the word that is needed will have to be retrieved in the lexicon. As Levelt (1989:11) puts it:

> A lemma will be activated when its meaning matches part of the

> preverbal message. This will make its syntax available, which in turn will call or activate certain syntactic building procedures.

In the lexicon each entry is represented as having two sides: a *lemma*, which contains the meaning and the syntactical specifications, and a *form*, which includes all the information that is needed on the levels of morphology and phonology (see figure 2). What is central to this approach, that Levelt presents as the 'lexical hypothesis', is that conceptual meanings trigger lexical elements which in turn trigger grammatical elements (Levelt 1989:181). In our example this becomes clear when we realize that the choice of the topic 'BOOK' leads to the selection of the lexical item *book* which in turn makes it necessary to begin the sentence with an article, 'a' or 'the' depending on whether the book has already been mentioned or not.

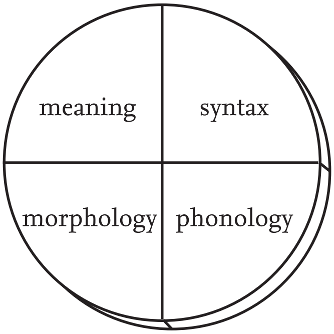


Figure 2. Internal structure of an item in the mental lexicon (Levelt 1989:182)

This procedure, that is presented in a very sketchy way here, is executed in an astonishing speed and with an incredibly high degree of correctness. It is well known that a speaker with a normal speech rate produces some 150 words per minute, but that this pace can go up to about 300 occasionally, which means that from 400 to only 200 milliseconds are needed per word. These choices are made from the total stock that constitutes the mental lexicon, which counts tenths of thousands of elements. Nevertheless, the number of wrong choices or slips of the tongue is less than about 0,1 percent (cf. Levelt 1989:199). That is what we call a skill or, in the context of language use, fluency (cf. Dörnyei 2009:286 – 293). And that is what adult native speakers of a language typically have acquired.

Although not much is known in detail about this fascinating process of lexical choices, it suggests that there must be a close link between the total knowledge that is contained in the human memory, especially all the subjects we are able to talk about, and the means we put to use during communication with others, i.e. the lexical units. The native language we are raised in offers words and expressions to name and categorize the outside world and structures to a large extent the whole of our knowledge, impressions and feelings.

Learners of a second language already have such an intimate link between their understanding of the world and the specific linguistic elements and means of their mother tongue. What they have to acquire in the other language is the skill, the more or less automatic link that permits them to pass swiftly from the one to the other. And they have to discover that the new language is not just a new set of labels for the same outside reality, but that at points the other language creates a different world (cf. Dagut 1977, Jiang 2002).

In order to understand what is going on when a second language is learnt, De Bot (1992) and Jiang (2000) have adapted Levelt's model to this particular situation. Further expanding on that Ma (2009:54-57) presents the following description. As was already said, each item in the mental lexicon consists of two parts: the *lemma* and the *form*, where the lemma consists of the semantic and syntactic information and the form (or lexeme) includes the morphophonological information (see figure 2). Now, what the second language learner is first of all confronted with, in most cases, is the form, written or spoken. At this stage the lexical item is still almost totally void (see figure 3). The only thing the learner can do is interpret that form in terms of what is available in his long term memory and which is very closely linked to his mother tongue, as we have seen. So, in a second stage the lexical item will correspond to something like the model in figure 4, where the L2 form is linked to L1 semantics and L1 syntax. Ideally the item has to develop into an element that is integrated in the second language on the levels of semantics and syntax as well as on the level of morphology (see figure 5).

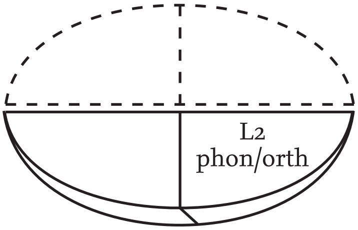


Figure 3. Lexical representation at the initial stage of lexical development in L2 (Jiang 2000:51)

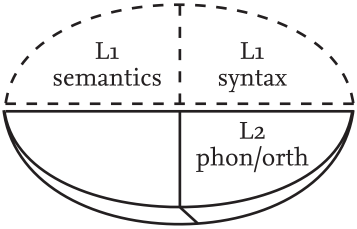


Figure 4. Lexical representation in L2 at the second stage (Jiang 2000:53)

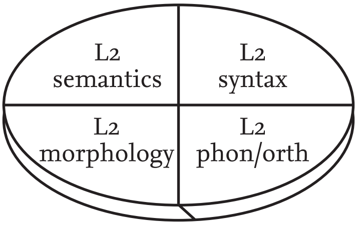


Figure 5. Lexical representation in L2 at the third stage (Jiang 2000:53)

It will be clear that this learning process will not always lead to this ideal situation and that in no way the mental lexicon of the L2 of a bilingual person will be identical to that of a monolingual person. In most cases, even for very advanced learners of a second language (also see section 3), the end state of lexical items in the mental lexicon will be much more complicated, leading to a far less orderly picture, maybe something like figure 6. In the case of the very advanced learner, morphology and syntax may be rather native like, and knowledge about the written form may well be perfect. In many cases, however, the phonetic form will be quite different from the L2 standard. The meaning will be highly influenced by the L1 and will only seldom be as rich and as easily accessible in speech as is the case for the native speaker (e.g. cognates or international items). Ma (2009:58), rightly I think, presents this description not only as a feasible model to account for L2 lexical development, but also as an insightful explanation of lexical errors and lexical stalemate or fossilization.

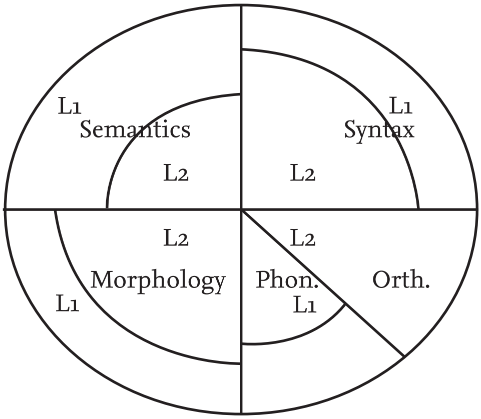


Figure 6. Lexical representation in L2 with influence of the L1.

3 The lexicon in second language research

In contrast with the central role of the lexicon that is claimed in Levelt's model and in its adaptations for SLA, most of the research done in the field of applied linguistics and second language learning is devoted not to the lexical side of the L2 but to grammatical subjects. Looking for the term 'dictionary' in more than ten handbooks and introductions in applied linguistics, bilingualism, and second language acquisition that have been published since 2000, I was struck by its almost complete absence. The term was not mentioned in the subject indexes of seven out of eleven such overviews of the field (see table 1). In three cases the term 'dictionary' was present in the subject index, but in the text itself not much was said about it. For instance, in Kaplan (2002) one is just reminded that dictionary making is one of the branches of applied linguistics, a statement that, rightly or wrongly, will certainly not please all lexicographers (cf. Wiegand 1984). In two other handbooks (Hinkel 2005 and Gass and Selinker 2008), there were only some quite obvious statements about the importance of dictionaries for SLA, and only one (Davies and Elder 2004) contained a chapter on dictionaries (by Alan Kirkness), an honest overview of the state of the art especially in pedagogical lexicography.

	'dictionary'	'lexicon'	'vocabulary acquisition'
Bhatia & Ritchie (2004)	-	+	-
Davies & Elder (2004)	+	+	-
De Bot & al. (2005)	-	+	-
Dörnyei (2009)	-	-	+
Doughty & Long (2003)	-	+	+
Gass & Selinker (2008)	+	+	+
Hinkel (2005)	+	+	+
Kaplan (2002)	+	+	-
Kroll & De Groot (2005)	-	+	+
Mitchell & Myles(2004)	-	+	-
VanPatten & Williams (2007)	-	+	-

Table 1. Presence of terms in eleven handbooks and introductions to SLA

When looking for the term 'lexicon' in these handbooks and introductions, one is better served. It is mentioned in all but one of the handbooks, the exception being Dörnyei (2009). In most cases lexical access or other aspects of the functioning of the mental lexicon are discussed in the context of psycholinguistic experiments, or else theories about the mental lexicon of balanced bilinguals are presented. In one handbook (Mitchell and Myles 2004:54), which is conceived within the Chomskyan Minimalist Program, it is stated that 'the core of human language is the lexicon (the word store)' which consists of two kinds of items:

> lexical categories, which include 'content' words such as verbs and nouns, and functional categories, which include 'grammatical' words such as determiners or auxiliaries, as well as abstract grammatical features such as Tense or Agreement, which may be realized morphologically.

The rest of the book is exclusively devoted to the grammatical side, leaving out any consideration about the acquisition of vocabulary.

The most interesting contribution is a chapter by Kroll and Sunderman (2003), who set out to describe the cognitive processes that support SLA. They discuss a number of models and theories that suggest that in proficient bilinguals

> lexical and semantic information in L1 is activated during both comprehension and production in L2 (p. 122).

One of these models is the revised hierarchical model (RHM, first proposed in Kroll and Stewart 1994), which is illustrated in figure 7 (Kroll and Sunderman 2003:114; for a discussion about this model see Brysbaert and Duyck (in press) and Kroll et al. (in press)). As can be seen, in a first stage there are stable links between the concepts and the lexical items of the L1 as well as between the lexical items of the L2 and those of the L1. In the latter case we speak of translation equivalents. Note that the lexicon of the L2 is much smaller than that of the L1. This picture reminds us of the early stage we have seen in the model presented by Jiang (2000). What have to be developed are direct links between concepts and the corresponding lexical items of the L2, now given as a dotted line. Kroll and Sunderman (2003:115) underline that

> the RHM is explicitly a developmental model. It assumes that the connections between words and concepts in bilingual memory change with increasing proficiency in the L2. ... A clear prediction of the RHM is that translation from L2 to L1 ... should be in place early in acquisition, whereas L1 to L2 translation ... will be more difficult for learners to perform.

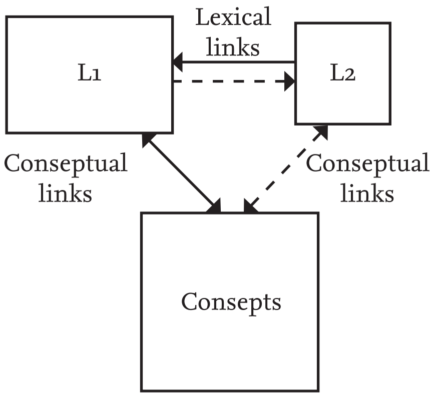


Figure 7. Revised hierarchical model (Kroll and Sunderman 2003:114)

Although this might sound like a truism, it is important to realize that now a scientific explanation is available for this well known fact. And this theoretical point permits the authors to criticize L2 teaching methods that are based on notions of inhibiting L1 activation, such as the Direct method (the Berlitz method), Total Physical Response (TPR), or the Natural Approach, as well as many modern practices in the movement of communicative second language learning (p. 122 – 123). All these methods try to exclude the L1 from the L2 learning scene, not taking into account the fact that this is impossible as even proficient bilinguals experience the influence of their mother tongue.

Returning to the handbooks and introductions in applied linguistics, one can look, on a more concrete level, for the importance of the lexicon as a subject matter in language courses. In other words: what is said about vocabulary learning? Again, one is struck by the paucity of the results: on eleven such books, only five devote some space to the concrete study of the lexical aspect of SLA (see Table 1). Fortunately, there are also a number of more specialized books that treat vocabulary learning (e.g. Schmitt 2000, Nation 2001, Bogaards and Laufer 2004, Ma 2009, to mention only the most recent ones). In the second part of this section I will try to give an outline of the main topics and findings.

Gass and Selinker (2008:449) state that

> In SLA research to date, there has been much less attention paid to the lexicon than to other parts of language, although this picture is quickly changing (see Nation, 2001; Singleton, 1999; Bogaards and Laufer, 2004). However, there are numerous

> reasons for believing that lexis is important in second language acquisition. In fact, the lexicon may be the most important language component for learners.

De Groot and Van Hell (2005:10), however, remind us that the lexical items to be learned are

> far too many to teach and learn via a method of direct teaching … it can easily be imagined that the teaching and learning of a full-fledged F[oreign] L[anguage] vocabulary is an impossible task that may discourage both teachers and learners of FL and direct their efforts to more manageable components of FL knowledge instead.

This explains why grammatical issues are more central to most language courses than lexical aspects. And as many applied linguists are linguists in the first place it is understandable that syntax is studied far more often than lexis. However this may be, it is the responsibility of the learner to come to terms with the lexicon: no course is long enough and no teacher has enough time to guide the learner through the whole vocabulary of the other language. But teachers and researchers should make clear suggestions as to how this tremendous task can be best approached and which avenues are the most successful.

One of the debates in L2 vocabulary acquisition concerns the differences between *incidental and intentional vocabulary learning*. In a very insightful paper Hulstijn (2003) first more exactly defines the terms incidental and intentional learning, which are often mistakenly used as synonyms of implicit and explicit learning. I will not go into terminological details here. What is essential is that in intentional learning

> attention is deliberately directed to committing new information to memory, whereas the involvement of attention is not deliberately geared toward an articulated learning goal in the case of incidental learning (Hulstijn 2003:361).

In incidental learning the most frequently used method is extensive reading in a situation where the learners do not explicitly have the intention to learn (new) vocabulary. They just 'pick up' elements of the language during the performance of a communicative task. Intentional learning can be done in the same way, but there are also other techniques, mainly in the form of some kind of paired-associate learning (translation pairs, illustrations and words, etc.) or of form focused activities. Laufer (2003) discusses the main assumptions underlying the 'vocabulary through reading hypothesis' and then reports three experiments in which she compares reading and several types of vocabulary focused activities (writing sentences or compositions). She arrives at the conclusion that

> if a word is practiced in a productive word focused task, its meaning has a better chance to be remembered than if a word is encountered in a text, even when it is noticed and looked up in a dictionary. (Laufer 2003:578)

Nevertheless, what is clear is that

(1) incidental learning does have an effect on the growth of the L2 lexicon;
(2) words are better learned when they occur more frequently in the input;
(3) learners with larger vocabulary sizes tend to profit more from this approach than those with small vocabularies;
(4) when learning words from context, it is not only the meaning that is learned, but collocational and grammatical aspects are taken in as well (see also Schmitt 2008:346 – 352).

On the other hand, when using some form of paired associates learning, the significance of the results may be overestimated because the learning materials and the test form are very much alike, but do not guarantee that what is learned really functions in language use. Even if many items can be learned in relatively little time in this way, they are not always easily used in language production because their grammatical and collocational properties have not been acquired ate the same time.

A point that is uncontroversial is that *vocabulary learning is an incremental process*: only seldom does one learn a lexical unit in one single moment (e.g. some cognates). This is true for vocabulary acquisition through reading as well as when using a form of paired-associate learning. In the latter case, the link between a form and a meaning can sometimes be established in a direct way, but aspects of morphological, grammatical or collocational behaviour are not taken in at the same time. In the case of reading, the growth of knowledge about particular lexical items is normally very slow. In a longitudinal small scale study of three learners of English (Schmitt 1998) it turned out that after one year none of the learners had more than a partial mastery of the meaning, the associations and the grammar of the eleven target words. Only spelling was not a problem for any of the learners. In a well designed experiment with 121 Japanese learners of English Webb (2007) shows that pseudo words in different contexts are better learned after three, seven, and ten presentations respectively. And this is true of receptive as well as productive knowledge, and of orthographic, semantic, syntactic, and other aspects of that knowledge. But even after ten encounters, the knowledge is far from complete, going from 66% (syntax) to 80% (grammatical functions) on the receptive measures used, and from 29% (meaning and form) to 77% (orthography) on the productive tests (see also Brown et al. 2008).

But not only frequency of exposure determines the learning results in L2 vocabulary acquisition. There are word type effects too. Aspects that have been fairly well researched include concreteness and cognate status. According to De Groot and Van Hell (2005:16)

> the recall scores are from 11% to 27% higher for concrete words than for abstract words

and

> the effect of cognate status varies between 15% and 19% when highly experienced FL learners were the participants in the vocabulary learning studies.

As lexicographers we know very well that these aspects are not the only ones in which words can differ. As is well known:

- nouns are different from verbs,
- adjectives behave in other ways than nouns,
- some words are highly polysemous, whereas others are strictly monosemous,
- prepositions may have a definable meaning in some of their uses, but have only a grammatical function in others (compare *The book was **on** the kitchen table*, i.e. not under it or behind it; and *The salary depends **on** the kind of job*),
- some words are complex, having clearly recognizable parts, others are opaque,
- some lexical units consist of one word, others are multi word items, and
- all lexical units have complex and unforeseeable particularities at the level of collocation.

In other words, *the lexicon is not a homogeneous mass*, as is often presupposed in vocabulary acquisition research. It is a highly complex crisscross of strictly individual elements and particular relationships in which it is difficult to see a clear structure.

All these differences and intertwining relationships that constitute the mental lexicon make me sometimes think of a wet cave with stalagmites and stalactites as in figure 8. The water is dripping at unequal paces and at different places, leaving each time some of the calcium that is in it and so building, slowly but steadily, the pillars and the 'curtains' that constitute the cave. As can be seen, sometimes solid structures are the result of a longstanding contact; in other cases the beginning structures on the floor are very far from the corresponding elements in the ceiling and it is not even very clear which drip will fall on what beginning stalagmite. This image is quite far from that of the dictionary with its typical two column pages where all words are treated like citizens in a democracy: equal, in spite of their big variation in almost all respects.

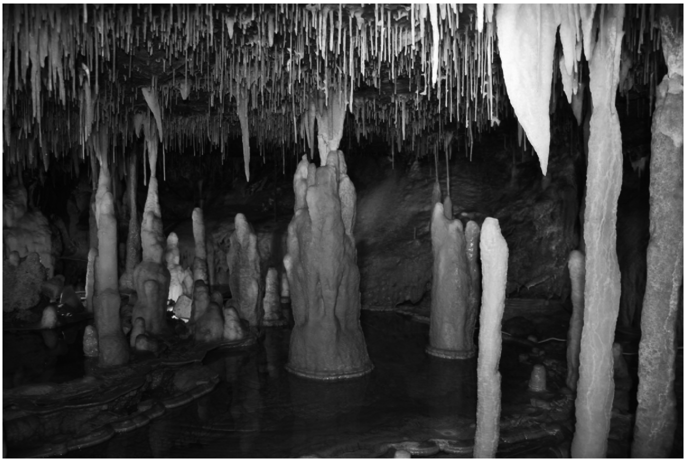


Figure 8. Artist's impression of the mental lexicon of the L2 learner.

To make the image even more complex, it is necessary to acknowledge that *vocabulary acquisition is also driven by individual aspects of the learner.* It is indeed due to the wide variety of factors involved in L2 vocabulary acquisition that the best means of mastering this crucial aspect of the second language is still unclear (cf. Schmitt 2008). Some people tend to 'pick up' new words or expressions much more easily than others. Some learners heavily rely on their L1 and on translation, others use the knowledge of their L1 or other languages and combine it with what they have acquired in the L2, inventing their own words, which sometimes happen to be correct or almost correct items of the other language. Some are very handy in deriving new forms from existing ones, whereas others do not see the structural elements of the L2.

All this has to do with what have been called *lexical skills, strategies, or techniques*. Nation (2005:589 – 593) groups them into four categories. The first is guessing or *inferring from context,* a technique everyone uses in his native language and that can be applied to the L2 as well, albeit that the ratio between what is well understood in a text and what causes problems is often so unfavourable in the L2 situation that success is far from guaranteed. The second strategy, *learning from word cards*, which is a form of paired-associate learning (for instance, words and their meanings are written down on the two sides of a card), can lead to specific learning outcomes, especially because the cards can be shuffled presenting the words in various orders. The third approach is through the *use of word parts*. Depending on the nature of the L2, it may be possible to analyse word forms into morphological elements, like '*pro*- (forward), -*gress*- (to move), -*ion* (noun)' which constitute *progression* meaning something like 'a movement forward' (Nation 2005:592). For a language like English and many others, this strategy should be applied with the utmost care, however, as is evident from words like *professor, profile* or *profit*, where the first syllable is not always a prefix and does not have a meaning like 'forward'. The last skill mentioned by Nation is *using a dictionary*. According to this author the dictionary can be used to check the guess that had been made while reading a text, or it can help in acquiring new vocabulary. Coady (1997:287), however, takes the view that dictionary use is not always positive as

> many adult L2 learners systematically misinterpret dictionary entries and take much more time on reading tasks as compared to nondictionary users.

On the other hand, Scholfield (1997:295 – 296) suggests that dictionary use may be better than just guessing, not only because the dictionary may provide more accurate information, but because it is more demanding and may therefore lead to a more elaborate mental treatment and, consequently, to better learning results. I will come back on this issue in section 5. Before that, in the next section, I want to examine the way the L2 learner is taken into account in (pedagogical) lexicography.

4 Dictionaries and the language learner

Fifty years ago, at the end of a first conference devoted to dictionaries that was held in Bloomington (Indiana), a list of recommendations was established and it was generally accepted that

> dictionaries should be designed with a special set of users in mind and for their specific needs (Householder 1962:279)

Since then a growing body of studies have been done concerning the dictionary user (see for a good overview Lew 2004). In most cases L2 learners have been the subjects and especially monolingual learner's dictionaries have been examined. So, this 'special set of users' should be fairly well known by now. However, it remains to be seen to what extent the L2 learner and especially the acquisition of L2 vocabulary play an important role in the study of dictionary use.

In overviews of the field of lexicography, the learner as learner is not really present. What most subject indexes of introductions and text books about lexicography do have are references to learner's dictionaries (e.g. Béjoint 2000, Hartmann 2001, Landau 2001, Atkins and Rundell 2008). References to the L2 learner, however, are quite rare, and this applies even more to the learning of vocabulary in a foreign language. Cowie (1999:49 – 51) speaks of the 'development of vocabulary' when presenting Hornby's ideas about the importance of relations between lexical items on the level of synonymy and antonymy, and he claims that, in order to foster L2 vocabulary acquisition, connections between senses should be reflected in the layout of an entry (Cowie 1999:148, 162). Svensén (2009) has an entry 'vocabulary learning' but the four references we find there all discuss presentation features of dictionaries: according to this author frequency indications are helpful to make clear which items are to be learned, whereas sense ordering, strict alphabetic ordering, and nesting are all claimed to have their positive or negative impact on the L2 vocabulary acquisition process.

But how do we know? What research outcomes are there to support such claims? Or to make this question more general: Is there any scientific evidence about the relationship between dictionary use and L2 vocabulary growth? Only a very limited number of studies have been done in this specific area. In order to frame the discussion it is important to make a fundamental distinction between what Galisson (1987) has called 'la lexicographie de dépannage' and 'la lexicographie d'apprentissage'. In the first case the dictionary is seen as a sort of breakdown truck which helps you out in a difficult situation. And this is the aspect that has been taken into account in most dictionary user studies. But what interests us here is the question whether the dictionary can be considered as a learning tool and to what extent dictionary use is conducive to vocabulary growth, especially on the long run.

About twenty years ago, I did a quite informal study to clear the ground in this particular respect (Bogaards 1991). I asked a group of university students of French (N = 44) to translate a Dutch text into French. The text contained 17 words that were expected to be mostly unknown to the subjects. Of the 44 students 12 had a bilingual dictionary at their disposal, 10 others could use the French learner's dictionary *Dictionnaire du français langue étrangère niveau* 2 (DFLE, Larousse 1978), 12 students worked with the monolingual *Petit Robert*, and 10 had no dictionary. The results showed that the bilingual group had looked up the most words (about 12 of the 17 target words) and had given the most correct translations (about 13.5); the DFLE group followed with about 7.6 words looked up and 7.6 correct translations; the group that used the *Petit Robert* looked up about 6 words and found about 8.0 correct translations. The group without dictionary could not look up any of the target words but nevertheless produced a mean of 5.6 correct translations. (It is clear from these figures that, contrary to the expectation, each subject already knew about five of the 17 target words.) Two weeks later a test that had not been announced was given to the students as well as to a group of 14 students who had not participated in the first part of the 'experiment'. The results of this test, that aimed at establishing the numbers of words that had been learned during the translation stage and through the use of different types of dictionaries, showed that the group that had learned the most words (about 4) was the DFLE group, followed by the two other groups that had dictionaries at their disposal (bilingual and *Petit Robert*, gain of about 3 words), whereas the students who had not used any dictionary had learned less than 2 new words. Because of the low numbers of students and the informal setting, no real conclusions can be drawn from this study, but it seems nevertheless that dictionary use can lead to more vocabulary learning than when no dictionary is used. This suggestion is confirmed by Cho and Krashen (2003) who found that two of their subjects, the ones who consistently used dictionaries while reading a book, acquired much more words than those who did not look up words in a dictionary.

Only recently some other studies have been conducted that can shed some light on the relationship between dictionary use and long term gains in L2 vocabulary knowledge. Aizawa (1999 described in Ronald 2003b:87 – 90), Laufer 2000, Ronald (2001), Yuzhen (forthcoming), and Dziemianko (in press) have studied the use of different types of dictionaries, paper or electronic, and have measured the knowledge the subjects, Japanese, Israeli, Chinese or Polish learners of English, had acquired after a delay of one to three weeks.

In Aizawa's study 308 high school students had to read a passage in English with or without a bilingual dictionary. Immediately after the comprehension test a surprise test was conducted in which the knowledge of 24 target words was tested; this test was repeated two weeks later. The results show that the dictionary group outperformed the no-dictionary group with almost 50% (15.60 as against 10.88 correct answers) on the first vocabulary test. On the delayed test this difference shrank but was still significantly higher (13.01 as against 11.42; note that the score of the no-dictionary group rose as was the case in my own study).

Laufer (2000) compared the acquisition of ten target words in a reading text where one group (N = 31) had the words glossed on the margin of the text, while the other group (N = 24) could click on the words and have access to dictionary information providing translations, definitions, and examples of usage. The two unexpected vocabulary tests, one immediately after the experimental session, the other two weeks later, both showed significantly higher scores for the 'dictionary' group than for the 'gloss' group.

Ronald (2001) asked 24 Japanese learners of English to study either a set of dictionary definitions or a set of authentic examples for 20 English adjectives. To measure vocabulary retention three weeks later, Ronald presented the subjects with the same materials in which the target words had been replaced with blanks; the subjects had to select the one correct word from four alternatives presented. Although the definition group had outperformed the example group in writing more correct sentences and in giving more correct translations during the first part of the procedure, all differences disappeared at the moment of the vocabulary retention test. The author suggests (p. 245) that the form of the test was the cause of this unexpected result, as the test did not really measure vocabulary retention but sensitivity to the contexts in which the target words had first been presented. My guess is that both (dictionary) definitions and (dictionary or authentic) examples may lead to vocabulary learning and that definitions are not necessarily superior to examples. The main point is that information that is typically found in dictionaries, definitions or examples, is favourable to L2 vocabulary acquisition.

Yuzhen (forthcoming) compares the use of palm top electronic dictionaries (PEDs) and paper dictionaries (PDs) in a task where 101

Chinese learners of English had to verify the meaning of ten target words and to write sentences including these words. After the experimental treatment the students filled in the Vocabulary Knowledge Scale (VKS, Paribakht and Wesche 1997) in which they could indicate to what extent they knew the target words. This test was repeated two weeks later without further announcement. The results show that there were no significant differences between the two groups as to immediate or long term word retention. But both groups had made some progress: they had retained about 26% of the words immediately after the treatment and about 17% two weeks later.

In a similar vein Dziemianko (in press) tries to find an answer to the question which form of a monolingual learner's dictionary (Cobuild6), the paper dictionary or its electronic format, is a better learning tool in L2 vocabulary acquisition. The subjects, 64 Polish learners of English, had to perform a receptive as well as a productive task: they were asked to explain or translate nine content words and to complete nine sentences in which prepositions had been left out. Two weeks later the same test (but with the items in a different order) was administered without any announcement and without the support of a dictionary. In this case the results show a statistically highly significant difference between the scores of the users of the electronic dictionary and those who had used the paper form of exactly the same dictionary. Whereas the former had acquired about 64% of the test items, the latter had obtained a score of only some 46%.

What can be concluded from these studies? In the first place it is clear from all of them that dictionary use can lead to gains in L2 vocabulary knowledge. In addition, both Aizawa's study and the study by Laufer confirm that the use of a dictionary leads to more vocabulary growth than no dictionary. Whereas the use of different types of bilingual dictionaries in electronic or paper form does not seem to lead to statistically significant differences in long term vocabulary growth in Yuzhen's study, in the more precise comparison of two forms of the same information in paper and electronic form in Dziemianko's research, the electronic presentation leads to far better retention of receptive as well as productive knowledge in L2.

5 The dictionary as learning tool

One of the basic aspects of knowing a language is, as we have seen, skill. This skill corresponds to one of two fundamental types of knowledge that is contained in the human brain: *procedural knowledge*. As opposed to *declarative knowledge*, which is the knowledge of facts, procedural knowledge answers the question of how things are done. We know that Paris is the capital of France, but we also know how to ride a bike. This difference between *knowing that* and *knowing how* also applies to vocabulary knowledge. We may know that a certain flower is called a *daisy* in English, a *pâquerette* in French, or a *madeliefje* in Dutch. But this knowledge is different from that about the ways these words can be properly used in a context. In the case of such concrete nouns as names of flowers the latter knowledge may be simple and easy to acquire for the L2 learner, but even then the native speaker of a language does know much more about them than most learners will ever do. Names of flowers may be associated to songs, be part of set phrases, or have symbolic values that do not exist in the same way in the native language of the L2 learner.

In a seminal paper the American psycholinguist George A. Miller (1999) tries to define what it means to know a word. He gives an overview of research that has been done by psychologists in order to explore the lexical network in native speakers and comes to the conclusion that no single theory is able to explain all the differences in verification times that are produced by subjects who are confronted with statements like 'A canary is a bird' or 'A canary is an animal'. And he underlines the importance of types of relationships between meanings other than hyponymy: i.e. synonymy, meronomy, troponymy, and various verbal entailments, which are now all contained in WordNet (see Fellbaum 1998). But even WordNet does not give a complete picture of the mental lexicon as it does not provide a topical organization: it does not give a handy overview of the words needed to discuss e.g. baseball. Another crucial feature that has not been incorporated in WordNet is a way to recognize the alternative meanings of a polysemous word. And this brings Miller to the important issue of context.

Polysemous words have been used by psycholinguists in order to find an explanation for the speed with which a particular meaning of a word can be identified and treated. From the results of this type of research Miller and his colleagues have gained the insight that,

> associated with each word meaning, there must be a contextual representation

and that

> a polysemous word must have different contextual representations [..]. A contextual representation is not in itself a linguistic context,

> but is an abstract cognitive structure that accumulates from encounters with a word in various linguistic contexts and that enables the recognition of similar contexts as they occur. (Miller 1999).

One could compare this to the 'knowledge' one has of faces which one readily recognizes in one type of social event, a family gathering or a business meeting, but that one would have difficulty in bringing home outside of a given context.

Miller (1999) then adds something that is important to lexicographers:

> Note that contextual representations are precisely what is missing from most dictionary definitions. But it is not easy to explain to lexicographers what more they should provide. Unfortunately, 'contextual representation' is not an explanation; it is merely a name for the thing to be explained.

This contextual representation limits the number of alternative meanings of polysemous words and so speeds up the process of comprehension. It is a sort of 'missing link' between declarative knowledge and procedural knowledge or skill. It could also be responsible for the fact that people are often able to finish the sentences of their interlocutor, to understand in spite of not really hearing part of what was said, and of reacting before the interlocutor has finished.

For the moment one can only speculate on the exact roles that are played by different types of context: situational context, topical context, and local or direct linguistic context. What is clear, however, is that computers are up to now fairly bad in picking the right sense of polysemous words and, what is more relevant in this context, that learners of a second or foreign language are not much better at that. Dictionaries, and especially monolingual learner's dictionaries have become much better over the years in providing information about what words mean and how they are used. As will be clear from the foregoing, however, they do not manage to tell the whole story and users will not yet find there these 'contextual representations' that native speakers seem to have and that explain their fabulous speed and correctness in handling lexical materials.

When language learners consult a dictionary, they may add something to their declarative knowledge. As the development of 'contextual representations' asks for multiple encounters of the same lexical unit in various contexts, one cannot easily overstate the importance of reading and listening in the process of L2 acquisition. Well chosen examples in dictionaries may certainly be assumed to help as well. We should, however, be unpretentious as to the role dictionaries can play in the total process of vocabulary acquisition. A dictionary cannot give as many and as varied contexts as are offered in extensive language use. And as dictionary consultation takes time, even in the electronic era, one should be reserved when advising learners about dictionary use. Although, as we have seen, dictionary consultation can be effective when L2 vocabulary acquisition is concerned, it may not always be the most efficient way. To summarize in a somewhat apodictic manner: dictionaries have their role to play when it comes to establishing declarative knowledge, but maybe not when procedural knowledge is strived for.

One other point needs to be made at the end of this paper. As we have seen in section 3, the L2 lexicon even of advanced bilinguals is influenced by the content of their L1 lexicon. So, some degree of bilingualism is always present. And the question arises if completely monolingual dictionaries are the best learning tools for L2 learners. It is not possible to give any firm answer to that question. It is in order, however, to note that the monolingual learners' dictionary came into existence in the context of direct methods that tried to avoid as much as possible the use of the native language of the learners (cf. Cowie 1999:1 – 13). Now that we know that this is impossible, we should try to take this point into account. As we have seen that bilingual dictionaries are not the ultimate answer to this situation because they do not seem to lead to better L2 vocabulary retention (see section 4), it is time to think about other, more effective and efficient lexical learning tools. At least two interesting proposals have been made in recent years. I think the avenues opened up by Laufer (1995, also see Laufer and Levitzky-Aviad 2006) as well as by Bogaards and Hannay (2004) deserve to be further explored. Both proposals try to combine the best of two worlds: the extensive knowledge that is condensed in modern monolingual learner's dictionaries and the exploitation of bilingual equivalences that are so well established in the learner's mental lexicon. In this context it is worthwhile to quote Schmitt (2008:337) who says:

> Although it is unfashionable in many quarters to use the L1 in second language learning, given the ubiquitous nature of L1 influence, it seems perfectly sensible to exploit it when it is to our advantage.

In the same vein Laufer and Girsai (2008:7) speak of the 'pervasive influence that L1 has on the learner lexis' and, after presenting an experiment the results of which give full support to a contrastive approach in L2 vocabulary acquisition, they conclude that

> there is indeed a place for contrastive analysis and translation activities in L2 teaching. .... Meaningful communication has been the goal of communicative language teaching, but the best method for achieving this goal may not be identical to the goal itself (Laufer and Girsai 2008:19).

Albert Sydney Hornby, who was one of the founding fathers of the English monolingual learner's dictionary, is described by Cowie (1999:12) as

> a man of broad sympathies and practical instincts who believed that the knowledge of the expert should be put to the service of the ordinary learner and teacher.

I am convinced that if he had known what is available as scientific evidence now, he would have been enthusiastic to adapt the dictionary to it. I think that we owe it to him to make every effort we can to better serve the ordinary learner and teacher. In order to improve the dictionary a closer collaboration between lexicographers, SLA experts and psycholinguist is more necessary than ever.

> References

Atkins, B.T. Sue and M. Rundell (2008). *The Oxford Guide to Practical Lexicography.* Oxford: Oxford University Press.

Béjoint, H. (2000). *Modern Lexicography. An Introduction.* Oxford: Oxford University Press.

Bhatia, T.K. and W.C. Ritchie (eds.). (2004). *The handbook of bilingualism.* Oxford: Blackwell.

Bogaards, P. (1991). 'Dictionnaires pédagogiques et apprentissage du vocabulaire.' In *Cahiers de Lexicologie 59*:93 – 107.

Bogaards, P. (1996). 'Lexicon and Grammar in Second Language Learning.' In P. Jordens, J. Lalleman (eds.). *Investigating Second Language Acquisition.* Berlin/New York: Mouton de Gruyter, 1996, 357 – 379.

Bogaards, P. and B. Laufer (eds.). (2004). *Vocabulary in a second language. Selection, acquisition, and testing.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Bogaards, P. and M. Hannay (2004). 'Towards a New Type of Bilingual Dictionary.' In G. Williams and S. Vessier (eds.). In *Proceedings of the 11th EURALEX International Conference.* Lorient: Université de Bretagne Sud. 463 – 474.

Brown, R., R. Waring, and S. Donkaewbua (2008). 'Incidental Vocabulary Acquisition from Reading, Reading-while-Listening, and Listening.' In *Reading in a Foreign Language* 20:136 – 163.

Brysbaert, M. and W. Duyck (in press). 'Is it Time to Leave behind the Revised Hierarchical Model of Bilingual Language Processing after 15 Years of Service?' In *Bilingualism: Language and Cognition.*

Cho, K-S. and S. Krashen (1994). 'Acquisition of Vocabulary from the Sweet Valley Kids Series: Adult ESL Acquisition.' In *Journal of Reading* 37, 662 – 667.

Coady, J. (1997). 'L2 Vocabulary Acquisition: A Synthesis of the Research.' In J. Coady and T. Huckin (eds.). *Second Language Vocabulary Acquisition*, Cambridge: Cambridge University Press, 273 – 290.

Cowie, A.P. (1999). *English Dictionaries for Foreign Learners. A History.* Oxford: Clarendon Press.

Dagut, M.B. (1977). 'Incongruencies in Lexical 'Gridding'– an Application of Contrastive Semantic Analysis to Language Teaching.' In *IRAL* 15:221 – 229.

Davies, A. and C. Elder (eds.). *(2004). The handbook of applied linguistics.* Oxford: Blackwell.

De Bot, K. (1992). 'Bilingual Production Model: Levelt's Speaking Model Adapted.' In *Applied Linguistics* 13:1 – 25.

De Bot, K., W. Lowie, and M. Verspoor (2005). *Second language acquisition. An advanced resource book.* London: Routledge.

De Groot, A.M.B. and J.G. Van Hell (2005). 'The learning of foreign language vocabulary. In Kroll and De Groot 2005:9 – 29.

Dörnyei, Z. (2009). *The Psychology of Second Language Acquisition.* Oxford: OUP.

Doughty, C. and M. Long (eds.). (2003). *The handbook of second language acquisition.* Oxford: Blackwell.

Dziemianko, A. (in press). 'Paper or Electronic? The Role of Dictionary Form in Language Reception, Production, and the Retention of Meanings and Collocations.' In *International Journal of Lexicography 23/3.*

Fellbaum, C. (ed.) (1998). *WordNet: An Electronic Lexical Database.* Cambridge, Mass. And London: MIT Press.

Galisson, R. (1987). 'De la lexicographie de dépannage à la lexicographie d'apprentissage : pour une politique de rénovation des dictionnaires monolingues de FLE à l'école.' *Cahiers de Lexicologie 51*:95 – 118.

Gass, S.M. and L. Selinker (20083). *Second language acquisition: An introductory course.* Mahwah, New Jersey: Erlbaum.

Guiora, A.Z. (1972). 'Construct Validity and Transpositional Research: Toward an Empirical Study of Psychoanalytic Concepts.' *Comprehensive Psychiatry* 13:139 – 150.

Hartmann, R.R.K. (2001). *Teaching and Researching Lexicography.* (Applied Linguistics in Action), Harlow: Longman-Pearson Education.

Hinkel, E. (ed.) (2005). Handbook *of research in second language teaching and learning.* Mahwah, New Jersey: Erlbaum.

Householder, F.W. (1972). 'Summary Report.' In F.W Householder and S. Saporta (eds.). *Problems in Lexicography: Report of the Conference on Lexicography Held at Indiana University November 11 – 12, 1960*. Bloomington, Ind.: Indiana University. 279 – 282.

Hulstijn, J.H. (2003). 'Incidental and intentional learning.' In Doughty and Long 2003:349 – 381.

Jiang, N. (2000). 'Lexical representation and development in a second language'. In *Applied Linguistics* 21:47 – 77.

Jiang, N. (2002). 'Form-Meaning Mapping in Vocabulary Acquisition in a Second Language.' In *Studies in second Language Acquisition* 24:617 – 637.

Kaplan, R.B. (ed.) (2002). *The Oxford handbook of applied linguistics***.** Oxford: OUP.

Kirkness, A. (2004). 'Lexicography'. In Davies and Elder 2004:54 – 81.

Kroll, J.F. and E. Stewart (1994). 'Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations.' In *Journal of Memory and Language* 33:149 – 174.

Kroll, J.F. and G. Sunderman (2003). 'Cognitive processes in second language learners and bilinguals: the development of lexical conceptual representations'. In Doughty and Long 2003:104 – 129.

Kroll, J.F. and A.M.B. De Groot (eds.). (2005). *Handbook of bilingualism: Psycholinguistic approaches***.** Oxford: OUP.

Kroll, J.F., J.G. van Hell, N. Tokowicz, and D.W. Green (in press). 'The Revised Hierarchical Model: A Critical Review and Assessment.' *Bilingualism: Language and Cognition.*

Landau, S. I. (2001). *Dictionaries : The Art and Craft of Lexicography.* Cambridge: Cambridge University Press.

Laufer, B. (1995). 'A Case for Semi-bilingual Dictionary for Production Purposes.' In *Kernerman Dictionary News 3.*

Laufer, B. (2000). 'Electronic Dictionaries and Incidental Vocabulary Acquisition: does Technology make a Difference?' In U. Heid et al. (eds.). *EURALEX.* Stuttgart University: 849 – 854.

Laufer, B. (2003). 'Vocabulary Acquisition in a Second Language: Do Learners Really Acquire most Vocabulary by Reading? Some Empirical Evidence.' In *Canadian Modern Language Review* 59:565 – 585.

Laufer, B. and N. Girsai (2008). 'Form-focused Instruction in Second Language Vocabulary Learning: A Case for Contrastive Analysis and Translation.' In *Applied Linguistics* 29:1 – 23.

Laufer, B. and T. Levitzky-Aviad (2006). 'Examining the Effectiveness of 'Bilingual Dictionary Plus' – A Dictionary for Production in a Foreign Language.' *International Journal of Lexicography* 19:135 – 155.

Levelt, W.J.M. (1989). *Speaking. From Intention to Articulation.* Cambridge, Mass: The MIT Press.

Lew, R. (2004. *Which Dictionary for Whom? Receptive Use of Bilingual, Monolingual and Semi-Bilingual Dictionaries by Polish Learners of English.* Poznan: Motivex.

Ma, Q. (2009). *Second Language Vocabulary Acquisition.* Bern : Peter Lang.

Miller, G.A. (1999). 'On Knowing a Word.' In *Annual Review of Psychology* 50:1 – 19.

Mitchell, R. and Myles, F. (2004^2). *Second language learning theories***.** London: Arnold.

Muysken, P. (2004). 'Two linguistic systems in contact: grammar, phonology and lexicon'. In Bhatia and Ritchie 2004:147 – 168.

Nation, I.S.P. (2001). *Learning vocabulary in another language.* Cambridge, UK: Cambridge University Press.

Nation, I.S.P. (2005). 'Teaching and learning vocabulary'. In Hinkel 2005. 581 – 595.

Qing Ma (2009). *Second language vocabulary acquisition.* Bern: Peter Lang.

Ronald, J. (2003a). 'A Review of Research into Vocabulary Acquisition through Dictionary Use. Part 1: Intentional Vocabulary Learning through Dictionary Use.' In *Studies in the Humanities and Sciences* 44/1:285 – 307.

Ronald, J. (2003b). 'A Review of Research into Vocabulary Acquisition through Dictionary Use. Part 2: Incidental Vocabulary Acquisition through Dictionary Use.' In *Studies in the Humanities and Sciences* 44/2:67 – 97.

Schmitt, N. (1998). 'Tracking the incremental acquisition of second language vocabulary: A longitudinal study'. In *Language Learning 48*:281 – 317.

Schmitt, N. (2000). *Vocabulary in language teaching.* Cambridge, UK: Cambridge University Press.

Schmitt, N. (2008). 'Instructed Second Language Vocabulary Learning.' In *Language teaching Research* 12: 329 – 363.

Scholfield, P. (1997). 'Vocabulary Reference Works in Foreign Language Learning.' In N. Schmitt and M. McCarthy (eds.). *Vocabulary. Description, Acquisition and Pedagogy.* Cambridge: Cambridge University Press. 279 – 302.

Singleton, D. (1999). *Exploring the second language mental lexicon.* Cambridge, UK: Cambridge University Press.

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making.* Cambridge: Cambridge University Press.

Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Tübingen: Niemeyer (Lexicographica Series Maior 134).

VanPatten, B. and J. Williams (eds.). (2007). *Theories in second language acquisition: An introduction***.** Mahwah, New Jersey: Erlbaum.

Webb, S. (2007). 'The effects of repetition on vocabulary knowledge.' In *Applied Linguistics 28*:46 – 65.

Wiegand, H.E. (1984). 'On the Structure and Contents of a General Theory of Lexicography.' In R.R.K. Hartmann (ed.). *LEXeter '83 Proceedings. Papers from the International Conference on Lexicography at Exeter, 9-12 September 1983.* Lexicographica. Series Maior 1, Tübingen: Max Niemeyer, 13 – 30.

Yuzhen (forthcoming). 'Dictionary Use and EFL Learning. A Contrastive Study of PEDs and PDs.' Paper submitted to the *International Journal of Lexicography.*

Telling it straight: A comparison of selected English and Polish idioms from the semantic field of speaking

ARLETA ADAMSKA

1 Introduction

This paper looks at the problem of providing idiom equivalents in bilingual dictionaries. The departure point is the common situation where one L1 idiom is translated by several L2 idioms (and vice versa), often within the same dictionary. We believe that greater precision in the provision of idiomatic equivalents can be achieved through implementing the methodological instrument devised by Dobrovol'skij and Piirainen (2005).[1] D&P (2005) are proponents of the so-called Conventional Figurative Language Theory, an approach which differs from the better-known Cognitive Metaphor Theory (CMT) in terms of its goals:

> For the CMT, it is important to discover quasi-universal conceptual metaphors that underlie each single metaphorical expression (...) For the Conventional Figurative Language Theory, however, the level of the very general metaphor is mostly of no interest. The Conventional Figurative Language Theory has to explain how the characteristics of figurativeness (above all, the image component) influence semantic and pragmatic specifics of CFUs [conventional figurative units] (D&P 2005:130).

In the following, Conventional Figurative Language Theory will be combined with those basic elements of CMT to which D&P themselves make frequent reference.

2 Sources and types of idiom equivalence

The idiomatic content plane is built of two elements: the actual meaning and the mental image. The fact that these are relatively independent of each other results, as D&P (2005:68) put it, in 'the existence of idioms which have (nearly) the same image, but differ with regard to their actual meanings, as well as the existence of idioms which have (nearly) the same actual meaning, but differ with regard to their images. Hence, these two major types of non-equivalence and their different combinations can be distinguished.' Establishing whether two idioms are equivalent (either within one language or cross-linguistically) requires a systematic

[1] Henceforth D&P (2005).

comparison between their respective actual (i.e. figurative) meanings and underlying images (literal readings). The possible configurations of differences constitute the semantic parameter of idiom equivalence, the other two dimensions being the syntactic and the pragmatic parameter.

For reasons of space, we shall only consider four candidates for equivalence: two idioms each for English and Polish.

3 Analysis

3.1 – not mince (one's) words (variant: *without mincing words*)

3.1.1 – Semantics

actual meaning:[2] to use direct, forceful words when speaking your mind, to say what you mean without trying to be polite
The positive form, *mince (one's) words* is not an exact antonym, as it includes a pejorative component:

> CIDE: *She found herself irritated by the interviewer's mincing* (=too delicate and not direct enough) *way of asking questions.*
> BYU-BNC:[3] In a very real sense, then, the Big Bang Universe has existed forever. Some scientists --; and we should add hastily that they are in a small minority --; feel that this is mincing words: they feel that by 'forever' we should mean an infinite number of billions of years rather than a period whose duration can be estimated.

mental image
Holt (1961) relates *mince* to Latin *minutia* 'smallness, fineness', offering the analogy with the mincing of meat as a motivating link. The idea is repeated in Brewer: '[f]rom the mincing of meat to make it more digestible or pleasing.' Apparently, the image is that of words (which

[2] The actual meanings we postulate for all analysed idioms have been synthesised from the dictionaries cited in the References.

[3] http://corpus.byu.edu/bnc

contain ideas)[4] being cut into small pieces so that they can be received without discomfort. 'Unminced' words are gross and indigestible. The secondary meaning of the adjective *indigestible* ('not easy to understand') constitutes part of the linguistic evidence for the metaphor IDEAS ARE FOOD, with its submappings: ACCEPTING IS SWALLOWING and UNDERSTANDING IS DIGESTING (Kövecses 2002:72-74). The mental image highlights the manner of serving the 'food' (words/ideas).

3.1.2 – *Syntax*

The idiom adopts two major patterns: [Neg (*not*) +V (*mince*) + (det) (*one's*) N (*words*)] or [prep (*without*) + gerund (*mincing*) + N (*words*)]. It takes both human and, less frequently, non-human subjects (2 out of the 37 hits in the BNC):

> Auto Express did not mince its words: 'It's difficult to see how much more Ford could have done to improve a car that was already very good.'
> The introductory leaflet did not mince its words: 'It was a period when words contradicted deeds, propaganda realities, and when everyday life was full of fear, hypocrisy, and people felt helpless, having been at the mercy of those in power.'

3.1.3 – *Pragmatics*

<u>degree of familiarity and/or textual frequency</u>

The expression appears to be commonly known; it is marked as a key idiom in CCID2.

<u>illocutionary function</u>

First, the idiom acts metalinguistically, preparing the addressee for the harsh words they are about to hear (occasionally with an implied negative assessment of whoever utters them):

[4] In accordance with Reddy's (1979) Conduit Metaphor, (i) THE MIND IS A CONTAINER (FOR IDEAS), (ii) IDEAS (OR MEANINGS) ARE OBJECTS, (iii) COMMUNICATION IS SENDING, (iv) LINGUISTIC EXPRESSIONS ARE CONTAINERS (FOR IDEAS-OBJECTS) (Krzeszowski 1997:170). As further noted by Krzeszowski (1997:174), component (iv) of this conceptual metaphor 'can also be instantiated by more specific source domains since (...) containers may be of various kinds. There are numerous linguistic expressions coherent with such instantiations. Thus, words can be 'soft' or 'hard', 'light' or 'heavy', 'delicate' or 'rough', 'sharp' or 'blunt, 'heated/hot' or 'cold/cool'. They can be 'cracked' (like nuts), 'coined' (like medals), distorted (like practically anything), 'minced' (like meat), 'borrowed', though practically never returned, 'played on' (like musical instruments), 'broken' (like fragile objects), and 'weighed'.'

> COBUILD2: *Never one to mince words, Carlie told her daughter that her looks were fading.*
> LDOCE3: *He's a brash New Yorker who doesn't mince his words.*
> BYU-BNC: They did not mince their words. One developer said his speech was 'as welcome as a bad smell in a space capsule'.

Secondly, mincing one's words is generally regarded as undesirable:

> BYU-BNC: You don't have to mince your words for my benefit, Harry
> BYU-BNC: Can you promise you won't be mincing your words? No way --; we'll tell the truth when it's needed to be told.
> BYU-BNC: It is unkind --; perhaps I should not mince words --; it is cruel to keep albino fish under bright lights; with a fish like the Oscar, often kept in a tank with no shade to escape into, this detail is particularly important.

3.2 – not beat about/around the bush (variant: *without beating about/around the bush)*

3.2.1 – Semantics

<u>actual meaning</u>

It is convenient to start the analysis with the definition of the opposite: *beat about the bush* 'approach a matter cautiously, indirectly, even over-cautiously or circuitously, because it is unpleasant, embarrassing, or delicate' (Brewer, CIDE, LDOCE3). *Not beat about the bush* is the exact antonym: 'approach a matter directly and immediately, without unnecessary delay, get to the point quickly.'

<u>mental image</u>

The expression is believed to have originated in hunting: 'one goes carefully when beating a bush to find if any game is lurking within' (Brewer). The image of carefully circling a bush in search of game corresponds to the cautious manner of approaching the subject as if it were a dangerous or skittish animal. The opposite, *not beating about the bush*, evokes the image of not bothering to be very careful in approaching the game (or the subject of conversation).

3.2.2 – Syntax

The idiom is sometimes preceded by *(there is) no point in* … (5 times out of 35 in the BNC). Two major patterns are attested: [Neg (*not*) +V (*beat*) + prep (*about*) NP (*the bush*)] or [prep (*without*) + gerund (*beating*) + prep (*about*) NP (*the bush*)]. The expression typically takes a human subject:

> BYU-BNC: I mean, let's not beat about the bush here…
> BYU-BNC: I won't beat about the bush about creeping privatisation because…
> In the version without *not*, it is often preceded by other elements of negation:
> BYU-BNC: He never beat about the bush when something was annoying him.
> BYU-BNC: Seeing no point in beating about the bush, she spoke directly.

The positive form, although less frequent, is by no means rare:
BYU-BNC: ...well we could beat about the bush but...
BYU-BNC: She winced at their infelicities, at the clumsy way they beat about the bush.

3.2.3 – Pragmatics
stylistic properties
The idiom can be encountered in informal as well as formal contexts. Examples of the former include:

> BYU-BNC: ..., and let's not beat about the bush, 'drop the dead f*****g donkey' and play Wallace and Forrester up front.
> BYU-BNC: You know, the one that gets on your nerves! Not very nice! Well Well she was! Not worth beating around the bush is there? Well are there single rooms there or Yeah! they're single, the accommodation is single then is it?

In formal contexts, the expression seems to stand out and, as such, is sometimes marked graphically by inverted commas. The use of the familiar idiom presumably serves to make a difficult message more accessible:

> BYU-BNC: The reason that Jesus talked with this woman was that he wanted to save her. That is, to say, he wanted to reveal to her, her sinful condition and need, and this he did when he speaks to her, about her sinful life. He doesn't beat about the bush. He doesn't come soft with it.
> BYU-BNC: That is to say, the important international conference to take place at Darlington Hall was by then looming ahead of us, leaving little room for indulgence or 'beating about the bush'.

degree of familiarity and /or textual frequency

The BNC total frequency is 35 (per 100 million words), that is, similar to that of *not mince (one's) words.*

cultural component

The background knowledge pertains to hunting customs, in particular to the fact that the hunter was required to walk carefully around any vegetation suspected of hiding a fox or a game bird.

illocutionary function

The idiom functions as a veiled comment on people's verbal behaviour. While *beating about the bush* tends to be condemned, the opposite is praised as a token of the speaker's openness and directness:

> BYU-BNC: She winced at their infelicities, at the clumsy way they beat about the bush. She saw that it had been a mistake --; an evasion perhaps? --; to hamper herself with the abstractions of that cryptic poem
>
> BYU-BNC: He was often charming, sometimes rude, but always straight. He never beat about the bush when something was annoying him, he was never afraid to give it to you straight, to say exactly what he thought.

3.3 – nie przebierać w słowach

3.3.1 – Semantics

actual meaning: to use direct, forceful, often crude or vulgar words when speaking your mind, to say what you mean without trying to be polite

The use of crude words is foregrounded in all dictionary definitions; in the examples found in the National Corpus of Polish,[5] words of this kind often feature in the co-text:

> Do du... z taką demokracją – nie przebiera w słowach 46-letni Kazimierz Świdroń. [Sod such democracy – 46-year-old Kazimierz Świdroń doesn't pick and choose words]
>
> ...sąsiadka przystępuje do ataku. Nie przebiera w słowach. Wyzywa mnie od najgorszych i wciąż grozi... [the neighbour launches an attack. She doesn't pick and choose words. She calls me names and keeps threatening me...]

mental image

The verb *przebierać* is defined in SJP online as follows:

[5] http://www.nkjp.pl

1 oczyścić coś, wybierając to, co właściwe, a odrzucając to, co uszkodzone, zepsute [clean something, choosing what is appropriate, and rejecting what is damaged, rotten]

2 nie móc się zdecydować na coś, rzadziej na kogoś [be unable to decide on sth, rarely on sb]

SSJP further specifies the latter meaning as: wybredzać, grymasić 'be choosy, fussy'. The verb *przebierać* is also encountered in expressions such as *przebierać jak w ulęgałkach* lit. 'sort sth like wild pears', meaning 'pick and choose'; *nie przebierać w środkach* 'do sth by fair means or foul'. Accordingly, the emerging image of the idiom under analysis is that of someone not caring to separate the good objects (words) from the bad. If one is not 'choosy' about words, one does not care whether they are appropriate or polite.

3.3.2 – Syntax

Nie przebierać w słowach [Neg (*not*) + V (*pick and choose*) + prep (*in*) +Nloc (*words*)] typically takes a human subject, which is sometimes expressed by a collective noun, such as *załoga* 'crew' or *opozycja* 'the opposition'. The verb component can take the present participle form, e.g.:

> NKJP: …choć pismo nie przebierając w słowach atakuje i Cimoszewicza, i także jego samego. [...although the magazine, not picking and choosing words, criticizes both Cimoszewicz and himself.]

There are also examples of the adjectivised expression [Neg. (*not*) + Adj (*picking and choosing*) + prep (*in*) +Nloc (*words*)] modifying a nonhuman subject such as *polemika* 'polemics', *debata* 'debate', *atak* 'attack' or *krytyka* 'criticism', e.g.:

> NKJP:…obok dawnej, często totalnej i nie przebierającej w słowach krytyki, dostrzeżemy również elementy….[side by side with the old, often total criticism which did not pick and choose words, we can also discern elements of...]

The idiom is only occasionally used without the negative:

> NKJP: ...sympatyczny, przyjemny, łagodny. Mówi najwyraźniej przebierając w słowach, jest cukierkowaty. […likeable, pleasant, gentle. He evidently speaks picking and choosing his words, he is sugary…']

Przebierać w słowach is not a simple antonym of the canonical form. It means: ‘to use excessively mild terms to express your opinion’, rather than: ‘not to use rude words when expressing your opinion’.

3.3.3 – *Pragmatics*

stylistic properties
The dictionaries are silent on this point. In the corpus, the idiom is found in both informal and formal contexts.
degree of familiarity and /or textual frequency
The idiom is commonly known. It appears 146 times per 350 million words.
illocutionary function
The expression is used to prepare the addressee for the unpleasant words which are about to follow. It can also act as a (negative) metalinguistic comment on someone’s way of speaking.

3.4 – *nie owijać w bawełnę (variant: bez owijania w bawełnę)*

3.4.1 – *Semantics*

actual meaning: to speak or write about something openly, without using euphemisms, e.g.:

> NKJP: Mówi prosto, dosadnie, nie owija w bawełnę…[He speaks simply, bluntly, doesn’t ‘wrap it in cotton’...]
> NKJP: Jak coś mi leży na wątrobie, to mówię. Na ogół nie owijam w bawełnę. W związku z tym mam dużo wrogów. [When something bothers me, I speak out. I do not normally ‘wrap it in cotton’. As a result I have many enemies.]
> NKJP: Nie ma co owijać w bawełnę – chciałem uciekać. [No use ‘wrapping it in cotton’ – I wanted to run.]

The truth being told is usually difficult and/or unpleasant, e.g.:

> NKJP: Dąbrowski patrzy smutnej prawdzie prosto w oczy i nie owijając w bawełnę mówi: – Nie myślałem, że będzie aż tak źle. [Dąbrowski faces the sad truth and says without ‘wrapping it in cotton’ – I did not expect it to be that bad.]
> NKJP: Nie ma co owijać w bawełnę. Spisujemy się słabo… [No point in ‘wrapping it in cotton’. We are doing poorly…]

mental image

As Professor Długosz-Kurczabowa explains,[6] 'the image of the burgeoning white fluff coming out of the cotton seed forms the basis for metaphorical meanings of this word [*cotton*]'. Thus, the idiom's underlying image is that of wrapping something in the threads of the cotton plant. Even if the surface of the object is rough, the layer of white cotton makes it look soft, deceitfully safe and innocent: cotton hides the real nature of the thing it envelops. Here it is, of course, the objectified hard words that are (not supposed to be) wrapped in cotton. The motivating metaphor seems to be KNOWING IS SEEING (Kövecses 2002:59).

3.4.2 – Syntax

The idiom is frequently accompanied by *nie będę* 'I'm not going to…', *nie ma co…* 'There is no point in…'. It often combines with *mówić* 'speak'. Adverbials of manner such as *szczerze* 'sincerely', *otwarcie* 'openly', *uczciwie* 'honestly' are frequently found in the immediate co-text, echoing and reinforcing the meaning of the phrase.

The idiom occurs both in the negative and positive form. The latter, less frequent, has the meaning 'express yourself indirectly, use euphemisms'. The Agent is always human. The following patterns are attested:

nie owijać coś/nie owijać czegoś w bawełnę [Neg *(not)* + V (*wrap*) (*sth*)+ prep (*in*) + Nacc (cotton)];

(*mówić*) *nie owijając w bawełnę* [(V)(*speak*) +Neg (*not*) + PresParticiple (*wrapping* + prep (*in*) + Nacc (*cotton*)];

(mówić) bez owijania w bawełnę [(V) (*speak*) + prep (*without*) + Ngen (*wrapping*) + prep (*in*) + Nacc (*cotton*)].

In the corpus there are only two examples of the idiom without negation:

> NKJP: Pisałem, naturalnie, bardzo ostrożnie, owijając w bawełnę.
> [I wrote very carefully, of course, 'wrapping it in cotton']
> NKJP: Przez długi czas wszelkie nasze zabiegi kampanijno-prewencyjne odwoływały się do aluzji i 'owijania w bawełnę'. Ich przesłanie było słabo czytelne, zwłaszcza dla młodzieży. [For a long time all our campaigning-preventive efforts relied on allusion and 'wrapping it in cotton'. Their message was unclear, especially to young people.]

There is also a maximally reduced nominalised form, the relatively new (and infrequent) creation:

bez bawełny [prep (*without*) + Ngen (*cotton*)]

[6] PWN online information service: http://poradnia.pwn.pl/lista.php?szukaj=bez+ogr%F3dekandkat=18.

3.4.3 – Pragmatics

stylistic properties

The expression is definitely informal, although none of the consulted dictionaries labels it as such.

degree of familiarity and /or textual frequency

The idiom is commonly known. The corpus search yields 182 hits (per 350 million words).

cultural component

The knowledge involved is that of cotton growing on plantations, especially the appearance of cotton seeds surrounded by soft white threads, with their connotations of delicacy and innocence.

illocutionary function

'Wrapping (sth) in cotton' is synonymous with being devious, and the behaviour in question is generally condemned, as evidenced by the frequent occurrence of the negative imperative ('Don't wrap (it) in cotton'). Speaking 'without cotton', on the other hand, is assumed to characterise a straightforward, honest person:

> NKJP: ...jest człowiekiem prostolinijnym, nie kluczy, nie owija w bawełnę tego, co ma do powiedzenia. [...he is a straightforward man, he does not hedge, does not 'wrap in cotton' what he has to say]
> NKJP: …nie boi się mówić wprost, gdy coś mu się nie podoba, nie owija niczego w bawełnę, tylko przedstawia swój punkt widzenia. [...he is not afraid to speak out when he doesn't like something, he does not 'wrap anything in cotton' but presents his point of view]

As a rule, this way of telling the truth is appreciated:

> NKJP: Potrzebujemy kogoś, kto powiedziałby nam brutalnie, bez owijania w bawełnę kilka ważnych prawd o naszym stanie zdrowia [We need someone who would tell us with brutal honesty, 'without wrapping in cotton', a few truths about our state of health.]
> NKJP: Jego siła polega na tym, że nie owijając w bawełnę, mówi ludziom o tym, co ich naprawdę interesuje i boli…[His strength lies in his 'not wrapping it in cotton'; he talks to people about what they are really interested in and concerned about...]

Still, isolated contexts can be found when such an open way of speaking is evaluated negatively:

> NKJP: …zawsze rąbał swoją prawdę nietaktownie, bez owijania w bawełnę. [...he would always give it straight, tactlessly, without 'wrapping it in cotton']

4 Discussion

4.1 – not mince (one's) words vs. *nie przebierać w słowach*

The actual meanings of both idioms involve the component of verbal aggression. The difference lies in the level of that aggression and in the resulting degree of perceived rudeness. The English idiom implies that the speaker does not bother to use euphemisms ('minced' words), so that neutral words are used where more delicate ones might be appropriate; the Polish expression stresses that, instead of neutral words, rude ones are used. The ultimate effect is, of course, the same: perceived tactlessness and, possibly, offence.

Both idioms include the lexical component *words*. Words are objectified, which makes it possible to handle them manually. Though in one case the mental image is that of the mincing of objects and in the other of choosing them carefully, both idioms hinge on not doing something because of disregard for another person's feelings, and both are motivated by the combination of the ontological metaphor IDEAS ARE THINGS and the structural conduit metaphor LINGUISTIC EXPRESSIONS ARE CONTAINERS FOR IDEAS. There are also syntactic analogies: the subject can be human or not; the patterns have a common element [Neg+V+(…)+N]. Neither idiom is subject to any special stylistic restrictions, and both can serve the purpose of pejorative evaluation in discourse. The implications of their opposite (but not exactly antonymous) forms are also similar: *przebierać w słowach* 'to use excessively mild terms to express your opinion' corresponds to the sense present in *mincing manner* 'a prim manner, one of affected delicacy' (Brewer).

4.2 – not beat about the bush vs. *nie owijać w bawełnę.*

Despite their divergent images, the idioms are semantically close. Superficially very different, the actions of circling a bush and of wrapping something in cotton share the circular movement. The two images evoke slightly different connotations: we circle the bush because we are afraid of scaring away whatever it hides; we wrap an object in cotton in order to hide it or to make it appear more acceptable on the outside. *Beating about the bush* entails more hesitation on the part of the speaker, caused by apprehension in approaching the subject, whereas ‘wrapping (sth) in cotton’ seems to require more control and premeditation. Nonetheless, these are nuances of emotive meaning, so the actual meanings can be regarded as equivalent. Additionally, the syntactic patterns and combinatorial properties are the same. Although the Polish idiom may be slightly more informal than the English one, both carry a positive evaluation in discourse.

5 Conclusions

The important question of bilingual lexicography: whether to give a verbatim translation of a SL phraseological unit or always aim at a TL phraseological unit of the same kind as the SL item, does not have an agreed upon answer. Still, the general preference for idiomatic equivalents seems clear. Thus, e.g., Svensén (1987 [1993]:156) claims that ‘[i]dioms in the source language must as far as possible be paralled in the target language by idioms with the same content.’ Roberts (1996:193) recommends idiomatic translations whenever possible; only in cases where an idiomatic equivalent is clearly lacking is she willing to accept a literal translation.

We are inclined to agree, with the important proviso that, to start with, a microanalysis along the lines presented here should be conducted in order to ascertain the degree of similarity between the candidates for equivalents. Admittedly, this kind of fine-grained comparison may not be a realistic requirement for general-purpose dictionary projects, but it should definitely feature as a crucial step in the compilation of bilingual dictionaries of idioms. Ideally, such dictionaries ought to include in their entries both TL paraphrases and idiomatic equivalents – provided, of course, that the latter can be identified, i.e. in cases such as those presented above.

> References

Dictionaries

Brewer = Evans, I. H. (1981). *Brewer’s Dictionary of Phrase and Fable.* (Revised ed.) London: Cassell.

CIDE = Procter, P. (ed.). (1995). *Cambridge International Dictionary of English.* Cambridge: CUP.

CCID2 = Sinclair, J. (ed.). (2002). *Collins COBUILD Idioms Dictionary.* (2nd ed.) Glasgow: HarperCollins.

COBUILD2 = Sinclair, J. (ed.). (1995). *Collins COBUILD English Dictionary.* (2nd ed.) London: HarperCollins.

Holt, A. H. (1961). *Phrase and Word Origins.* (Revised ed.) New York: Dover Publications.

LDOCE3 = Summers, D. (ed.). 1995. *Longman Dictionary of Contemporary English.* (3rd ed.) Harlow: Pearson Education Limited.

PSF = Głowińska, K. (2000). *Popularny słownik frazeologiczny*. Warszawa: Wilga.

SSJP = Szymczak, M. (ed.). (1978). *Słownik języka polskiego.* Vols. 1-3. Warszawa: PWN.

WSFJP = Müldner-Nieckowski, P. (2003). *Wielki słownik frazeologiczny języka polskiego.* Warszawa: Świat Książki.

Other works

Dobrovol'skij, D. and E. Piirainen. (2005). *Figurative Language: Cross-cultural and Cross-linguistic Perspectives.* Amsterdam: Elsevier.

Kövecses, Z. (2002). *Metaphor: A Practical Introduction.* Oxford: OUP.

Krzeszowski, Tomasz P. (1997). *Angels and Devils in Hell: Eleme.nts of Axiology in Semantics.* Warszawa: Wydawnictwo Energeia.

Reddy, M. J. (1979). 'The Conduit Metaphor: A Case of Frame Conflict in our Language about Language.' In Ortony, A. (ed.). *Metaphor and Thought.* Cambridge: CUP, 284-297.

Roberts, R. (1996). 'Le traitement des collocations et des expressions idiomatiques dans les dictionnaires bilingues'. In Béjoint, H. and P. Thoiron (eds.). *Les dictionnaires bilingues.* Louvain-la-Neuve: Aupelf-Uref-Éditions Duculot. 181-197.

Svensén, B. (1987) [1993]. *Practical Lexicography: Principles and Methods of Dictionary Making.* (Transl. by J. Sykes and K. Schofield.) Oxford: OUP.

··· *Papers*

> Reuse of Lexicographic Data for a Multipurpose Pronunciation Database and Phonetic Transcription Generator for Regional Variants of Portuguese

ASHBY, SIMONE AND FERREIRA, JOSÉ PEDRO

1 – Computational Lexicography and Lexicology

Among the benefits of a flexible and modular lexical database are: the facility of building new modules from existing ones, the reuse of lexicographic data to both enhance the user experience and achieve NLP aims, the time saved in accomplishing these objectives, and the economy that comes from minimizing redundancy (van der Eijk, Bloksma, and van der Kraan, 1992). LUPo, or the Portuguese Unisyn Lexicon, is one of the first *speech*-dedicated applications to take full advantage of a collection of lexical resources as the basis for a text-to-speech system. Consisting of a pronunciation lexicon and rule system for generating accent-specific phonetic transcriptions for Portuguese, LUPo circumvents the cost of producing high-quality phonetic transcriptions by hand, while attracting a wider pan Lusophone audience to the lexical database in which it resides, and providing the research community with a vast resource of Portuguese accent data for evaluating speech applications and testing theories.

> From the Definitions of the *Trésor de la Langue Française* to a Semantic Database of the French Language

BARQUE, LUCIE; NASR, ALEXIS AND POLGUÈRE, ALAIN

1 – Computational Lexicography and Lexicology

The *Definiens* project aims at building a database of French lexical semantics that is formal and structured enough to allow for a fine-grained semantic access to the French lexicon—for such tasks as automatic extraction and computation. To achieve this in a relatively short time, we process the definitions of the *Trésor de la Langue Française informatisé* (TLFi), enriching them with an XML tagging that makes explicit their internal organization (roughly, *genus* and *differentiae*) and enhancing the components with semantic labels that explicit their role in the definition. There is, to our knowledge, no existing broad coverage database for the French lexicon that offers to researchers and NLP developers a structured decomposition of the meaning of lexical units. Definiens is an ongoing research that will hopefully fill this gap in the near future.

> Morphosyntactic Lexica in the OAL Framework: Towards a Formalism to Handle Spelling Variants, Compounds and Multi-word Units

BLANCAFORT, HELENA; COUTO, JAVIER AND SENG, SOMARA

1 – Computational Lexicography and Lexicology

The creation and maintenance of lexicographic resources are labour-intensive tasks. In this paper we present SylLex, a formalism to encode morphosyntactic lexica and how it is used in the OAL framework, a tool to aid the linguist to create and maintain such resources in an industrial context. The aim is to have an intuitive and easy-to-use formalism, SylLex, implemented in a user-friendly and ready-to-use tool so that a linguist without previous experience is able to work effectively after a short training of one or two hours.

The paper is organised as follows. First, we present the SylLex formalism. SylLex organises the lexical and inflectional information by dividing the lexicon in three components: lemma, inflection paradigms and patterns. Thus, the linguist can better manipulate the data, assure consistency and correctness, and maintain the lexicon by modifying inflection paradigms instead of dealing with scripts to directly modify lexicon entries in text files. In addition to this, the system takes advantage of the notion of inheritance between different inflection paradigms of a same word category. Furthermore, we present how different kinds of variants of forms like geographic and spelling variants are encoded in SylLex. Moreover, we discuss the problem of defining and delimiting compound and multi-word units, and explain how they are stored in the lexicon. Finally, we draw conclusions on the advantages and disadvantages of the formalism and OAL and mention further work.

> TTC: Terminology Extraction, Translation Tools and Comparable Corpora

BLANCAFORT, HELENA; DAILLE, BÉATRICE; GORNOSTAY, TATIANA; HEID, ULRICH; MECHOULAM, CLAUDE AND SHAROFF, SERGE

1 – Computational Lexicography and Lexicology

The need for linguistic resources in any natural language application is undeniable. Lexicons and terminologies play indeed a central role in any machine translation tool, regardless of the theoretical foundations upon which the machine translation tool is based (e.g. statistical machine translation or rule-based machine translation). The EU project TTC ('Terminology Extraction, Translation Tools and Comparable Corpora') aims at leveraging machine translation tools, computer-assisted translation tools, and terminology management tools by automatically generating bilingual terminologies from comparable corpora in several European Union languages (English, French, German, Latvian and Spanish), as well as in Chinese and Russian. The TTC project will integrate developed and existing tools in an online platform including a tool to compile and handle comparable corpora, as well as a terminology management tool. The platform will be based on Web Services and will use reputable open solutions such as UIMA (Unstructured Information Management Architecture) and EuroTermBank.

> The Past Meets the Present in Swedish FrameNet++

BORIN, LARS; DANÉLLS, DANA; FORSBERG, MARKUS; KOKKINAKIS, DIMITRIOS AND TOPOROWSKA GRONOSTAJ, MARIA

1 – Computational Lexicography and Lexicology

The paper is about a recently initiated pilot project which aims at the development of a Swedish framenet as an integral part of a larger lexical resource, hence the name 'Swedish FrameNet++' (SweFN++). The SweFN++ project has four main goals: (1) to 'revitalize' a number of existing lexical resources and integrate them into a multi-faceted lexical resource for language technology (LT) applications, in the process enriching the individual resources using semi-automatic methods; (2) to construct a Swedish framenet (SweFN) and make it part of the integrated resource; (3) to develop a methodology and workflow which makes maximal use of LT and other tools in order to minimize the human effort needed to build the resource; and (4) to release the resource under an open content license. The above goals are also of great significance for lexicological research and computational lexicography, as a SweFN will lend relevant support in bringing to light semantic relations implicit in word meanings. The theoretical assumptions elaborated by the Berkeley FrameNet make up the backbone of the SweFN resource, which will pay special attention to compounds and multi-words expressions when used as target lexical units or frame elements. In this article, we present an inventory of free electronic resources with a focus on their role in the semi-automatic acquisition and population of Swedish frames. After a brief overview of Swedish resources, we reflect on attempts to recycling and linking lexical data in a semi-automatic manner and report on our work in progress, which can be followed at <http://spraakbanken.gu.se/swefn/eng/>.

> Encoding Attitude and Connotation in wordnets

BRAASCH, ANNA AND PEDERSEN, BOLETTE S.

1 – Computational Lexicography and Lexicology

The Danish wordnet, DanNet, though part of the global WordNet family, contains some information types that are not generally provided in wordnets such as qualia roles and *connotation* of words. Connotation is seen as the set of associations implied by a lexeme in addition to its primary, literary meaning; it is evoked by one (or more) particular feature of the entity referred to and suggests attitudes, emotions and opinions like admiration or disapproval. Lexemes with a connotation have an observable pragmatic effect in texts making them *subjective* or *opinionated.*

In the paper, we discuss the relevance of connotation information in lexicons for computational applications in general and present the set of encoded semantic information exemplified by empirical data. We focus on a particular ontological type of entities, namely *human*s with the focus on selected hyponyms of *person* that are encoded with a connotation value and discuss the prototypical properties evoking *positive* or *negative* connotations. The qualia structure based approach enables to encode both the prevalent, connotation evoking features and prototypical activities of the person.

The material encoded with connotation so far consist of 650 nouns and comprises a male, a female and a gender-neutral group, thus it lends itself to comparative examinations concerning the distribution of connotation evoking features and polarity distribution within each individual group and between the groups as well. One of the most striking observations says that (in our material) the negative connotation polarity is predominant; the most important feature of female persons seems to be their positive appearance, and a general disparaging attitude dominates as regards the conduct and manners of male persons.

> The DANTE Database (Database of ANalysed Texts of English)

CONVERY, CATHAL; Ó MIANÁIN, PÁDRAIG; Ó RAGHALLAIGH, MUIRIS; ATKINS, SUE; KILGARRIFF, ADAM AND RUNDELL, MICHAEL

1 – Computational Lexicography and Lexicology

This database (www.webDante.com) was designed and created for Foras na Gaeilge by the Lexicography MasterClass and their 15-strong team led by Valerie Grundy (Managing Editor); textflow is managed by Diana Rawlinson (Project Administrator). The corpus of 1.7 bn words of current English, custom-built in 2007, was queried using the Sketch Engine (www.sketchengine.co.uk/), and the database was compiled in IDM's Dictionary Production System (DPS: www.idm.fr). The present volume contains a fuller description of this project (Atkins, Kilgarriff and Rundell *Database of ANalysed Texts of English (DANTE): the NEID database project*) and of its use in a bilingual dictionary (Convery, Ó Mianáin and Ó Raghallaigh *Covering all bases: Regional Marking of material in the New English-Irish Dictionary).*

The 95,000 or so DANTE entries cover approximately 50,000 headwords and 45,000 compounds, idioms and phrasal verbs, using over 40 datatypes. The lexical entry is subdivided into lexical units, each a sense of a single- or multi-word lemma. Almost every linguistic fact recorded is accompanied by full corpus sentences illustrating its use in text. Apart from the definitions and the corpus-derived example sentences, all the significant information is machine-retrievable. Functionality demonstrated here includes simple and complex searches over various combinations of datatypes and the automatic insertion of empty translation fields for use in dictionary building. .

DANTE was created as the initial stage of compilation of the *New English-Irish Dictionary*. Its long-term potential is much more far-reaching: it offers publishers world-wide a comprehensive launchpad for bilingual dictionaries with English as the source language or the draft stage of a learners' dictionary of English; a source of updating material for an existing dictionary, etc. It offers software developers, universities and other research institutions a resource for improved word sense disambiguation, the creation or enhancement of online lexicons, and other uses in software applications such as machine-assisted translation, information retrieval systems, etc. More details from info@webDante.com.

> From a Bilingual Transdisciplinary Scientific Lexicon to Bilingual Transdisciplinary Scientific Collocations

DROUIN, PATRICK

1 – Computational Lexicography and Lexicology

Most linguistic studies dealing with the lexicon of scientific corpora are interested in subject area lexicon or terminology which leads to a general lack of description of the other types of lexical items contained in these corpora. The main exception to the previous statement is the work being done in the area of specialized language teaching like the studies of Coxhead (1998, 2000). In most cases, as pointed out by Tutin (2007), the lexicon itself is not what is being studied.

We consider that the lexicon used in scientific writings can be divided into three categories. The first one is the common basic lexicon, which includes function words such as determiners, auxiliary verbs and conjunctions, and content words of the general language. The second category is the transdisciplinary lexicon that includes abstract verbs such as *to think* or *to consider* and abstract nouns such as *idea*, *factor* and *relation*. It also includes a methodological lexicon that refers to the abstract lexicon used for the description of scientific activities and scientific reasoning. Examples of lexical items one would find in this category are *hypothesis*, *data* and *approach.* The last category of lexicon found in scientific writings is subject specific terminology, which refers to all concepts used in a particular domain.

Our study here is focused on the second category, which we called Transdisciplinary Scientific Lexicon (TSL), and its behavior in scientific writings. The main goal of this paper is to test the idea that we can start from a raw bilingual scientific corpus and automatically build a list of bilingual transdisciplinary scientific collocations around the lexical items from the second category described above (TSL).

> iLEX, a general system for traditional dictionaries on paper and adaptive electronic lexical resources

ERLANDSEN, JENS

1 – Computational Lexicography and Lexicology

Dictionaries are different: the purpose, the language(s) covered, authors co-operation during production, and the traditions and styles have asked for many solutions and editorial rule sets. The needs of individual projects can be covered in two ways: The specific system is customized at the programming level (maybe with reuse of existing modules); the general system is customized at a higher level that does not require programming skills.

iLEX is a general system integrating spell checking, structural and lexical help, lists, workflow, smartEditing, advanced graphical statistics, separated metadata, change tracking, fast powerful searching, alphabetization, tilde functions, element sorting and uniqueness control, and more for Windows, Linux and Mac. For more details, visit www.emp.dk. Based on full Unicode, XML schemas, Namespace, ISO Schematron, Xquery, Xpath and XSL 2.0 iLEX is a secure choice for the future.

Being 100 percent XML conforming, iLEX may be used for a broader range of texts: not only TEI and related standards as MENOTA, but any of the emerging application standards, e.g. DITA.

Single sourcing an indispensable aspect of XML, is a strong feature of iLEX: XSL publication on any platform is supported as well as ready-to-publish ilEX Comunico applications for internet and mobile phones using Symbian60 and Android.

Lately, online publishing has changed towards more adaptive approaches for two aspects. Adaptation of user interface layout and functionality to ensure higher user satisfaction is a must when financed by ads. Adaptation of content based on composition of both existing lexical resources and resources dynamically generated from computational analysis of language use on the internet is a way to keep costs within control.

This asks for new tools and methods. Lexical Information Mapping Architecture (LIMA) which will be sketched out as a part of this presentation may form the ground for a new standard for adaptive lexical work. It will relate to other standards as XML, TEI, DITA, SCORM, and ISO 1951 and more.
For more information about iLEX and LIMA visit www.emp.dk

> Corpus Exploitation Strategies for the Lexicographic Definition Task

FELIU, JUDIT; GIL, ÀNGEL; PEDEMONTE, BERTA AND GUIRADO, CRISTINA

1 – Computational Lexicography and Lexicology

The main goal of this paper is to formalize and to present some guidelines helping the lexicographic definition task. The research is applied to the corpus query procedures in order to retrieve some refined results that benefit the definition writing up process as much as possible. The paper will briefly introduce the three main language resources (LR) involved in the authors' daily linguistic job, that is, the Catalan descriptive dictionary built on the basis of a Catalan corpus, the corpus itself and the Catalan main dictionaries repository normally looked up. The focus of the paper will be put on the improvement of the strategies followed so far in order to use the corpus query system in an efficiently oriented manner as far as the descriptor selection and the extrinsic part of the definition fulfilment is concerned. The use of set of patterns will allow retrieving certain kind of information for each type of unit defined. Data retained will be more precise and the results are proved to be useful for maintaining coherence among different team members and also, and probably most important, among different but semantically related types of words defined in the descriptive dictionary.

> Mit einem Klick zu vielen Möglichkeiten: *www.deutsches-rechtswoerterbuch.de*

FRIELING, STEFANIE

1 – Computational Lexicography and Lexicology

The *Deutsches Rechtswörterbuch (DRW)* is a historical dictionary (based at the *Heidelberger Akademie der Wissenschaften*) which documents and describes the vocabulary of the historical legal language of 7th to 19th century German. *German* here refers to the West Germanic language family which includes amongst others Old Frisian, Old Anglo Saxon, Old Saxon, Langobardic, Low German und High German. Besides the words directly referring to juridical expressions (such as *Gericht* or *Prozess*) the dictionary also includes many words of the common language if they

occur in a legal context (e.g. in edicts or contracts). This conceptual design makes the DRW an important tool for a diverse range of historic research, e.g. cultural and social history, history of law, history of economy or linguistics.

Until now, 11 of the 16 planned volumes have been published. The majority of the nearly 100.000 dictionary articles is freely available online, too. The web version of the DRW, however, is much more than just the digital pendant of the printed reference book: Via the website www.deutsches-rechtswoerterbuch.de the dictionary user has access to a wide range of options to use the dictionary as a large information system.

Based on the software FAUST – also used by the lexicographers for their daily work – four databases (the dictionary itself, the sources, the digitised sources, the full text archive) can be searched online.

During the software presentation the use of the DRW web version as a powerful research tool will be demonstrated. By means of a variety of possible enquiries it will be shown how the user benefits from the interconnectedness of the FAUST databases and the integration of external information resources (e.g. other important historical dictionaries or digitised primary sources).

[Note: the author did not submit a full text version of this software demonstration]

> The *Louvain EAP Dictionary (LEAD)*

GRANGER, SYLVIANE AND PAQUOT, MAGALI

1 – Computational Lexicography and Lexicology

In our software demonstration, we describe a web-based English for Academic Purposes dictionary-cum-writing aid tool, the *Louvain EAP Dictionary (LEAD)*. The dictionary is based on the analysis of c. 900 academic words and phrases in a large corpus of academic texts and EFL learner corpora representing a wide range of L1 populations. The dictionary contains a rich description of non-technical academic words, with particular focus on their phraseology (collocations and recurrent phrases). Its main originality is its customisability: the content is automatically adapted to users' needs in terms of discipline and mother tongue background. Another key feature of the *LEAD* is that is makes full use of the capabilities afforded by the electronic medium in terms of multiplicity of access modes (Tarp 2009). The dictionary can be used as both a semasiological dictionary (from lexeme to meaning) and an onomasiological dictionary (from meaning/concept to lexeme) via a list of typical rhetorical or organisational functions in academic discourse (cf. Pecman 2008). It is also a semi-bilingual dictionary (cf. Laufer &

Levitzky-Aviad 2006) as users who have selected a particular mother tongue background can search lexical entries via their translations into that language.

The *LEAD* is designed as an integrated tool where the actual dictionary part is linked up to other language resources and learning tools. It is a hybrid dictionary (cf. Hartman, 2005) that includes both a dictionary-cum-corpus and a dictionary-cum-CALL component. As regards direct corpus access, the *LEAD* innovates by giving access to discipline-specific corpora rather than generic corpora.

While the current version of the tool is restricted to some disciplines and mother tongue backgrounds, its flexible architecture allows for further customisation (other L1 background populations, other disciplines, other languages).

> One Structure for Both Monolingual and Bilingual Dictionaries Converting a Large Number of Different Dictionaries to a Single XML Format

GROOT, HANS DE AND MASEREEUW, PIETER

1 – *Computational Lexicography and Lexicology*

Van Dale converted the source databases of all its dictionaries to a type of XML mainly designed to capture the function of its elements, rather than their formatting. We found that we could apply the same principles consistently to various types of dictionaries (monolingual, bilingual) and capture all content within a single XML structure. The new structure reduces the time needed for our production processes and for database maintenance. This article reports on our findings during the conversion and the principles we applied.

> Corpus-derived data on German multiword expressions for lexicography

HEID, ULRICH AND WELLER, MARION

1 – *Computational Lexicography and Lexicology*

We show a parsing-based architecture for the extraction of German verbal multiword expressions. It uses dependency parsing as a preprocessing step, allows us to extract syntactic patterns of arbitrary form from the parsed data, and comprises a relational database where each extracted multiword occurrence is stored along with the sentence it is extracted from, and with a number of morphosyntactic and syntactic features. These features serve (i) for an automatic decision about the likely idiomatization of the candidate under review, and (ii) in later lexicographic work to get a clear picture of lexicographically relevant linguistic properties of the selected candidates.

We use dependency-parsed text, because this allows us to find non-adjacent multiwords and to use subcategorization knowledge to identify e.g. verb + object pairs more reliably than on the basis of ourface patterns.

The extraction results illustrate the potential of the tools; we can identify morphosyntactic preferences in collocations (these often indicate idiomatization), longer collocational or idiomatic structures (where e.g. the core elements and possible modifies can be clearly distinguished), lexical variation in idioms, as well as certain specific features of collocations or idioms (e.g. preferences for negation).

As all data are stored in a database, which supports a variety of generalization steps, it is in principle possible to prepare different layouts (i.e. presentations and selections) of dictionary entries, for different user groups and user needs.

> Dictionary Building based on Parallel Corpora and Word Alignment

Héja, Enikő

1 – Computational Lexicography and Lexicology

The paper describes an approach based on word alignment on parallel corpora, which aims at facilitating the lexicographic work during dictionary building. This corpus-driven technique, in particular the exploitation of parallel corpora, proved to be helpful in the creation of bilingual dictionaries for several reasons. Most importantly, a parallel corpus of appropriate size guarantees that the most relevant translations are included in the dictionary. Moreover, based on their translational probabilities it is possible to rank translation candidates, which ensures that the most likely translation candidates are ranked higher. A further advantage is that all the relevant example sentences from the parallel corpora are easily accessible, thus alleviating the selection of the most appropriate translations from possible translation candidates. Due to these properties the method is particularly apt to enable the production of active or encoding dictionaries

> Using a Dictionary Production System to impose a WordNet on a Dictionary. A Software Presentation

Hvelplund, Holger and Ørsnes, Allan

1 – Computational Lexicography and Lexicology

In this demonstration, we will make a practical presentation of how a semantic web – a WordNet – can be imposed on a dictionary using a Dictionary Production System. Using a WordNet structure makes a lot of sense as there are already several WordNets for different languages which can be used freely.

We will demonstrate how we first impose the structure in a rather crude way allowing for lots of ambiguities and then use the various tools in the Dictionary Production System to target the dubious areas and refine them through manual intervention, either as part of a usual editing of a new edition or in a separate run.

Finally we will demonstrate how the data can be extracted and made available to a Dictionary Publishing System.

> Working with the web as a source for dictionaries of informal vocabulary

JANSSON, HÅKAN

1 – Computational Lexicography and Lexicology

Informal vocabulary, e.g. slang, jargon and other forms of expression that are particular to different types of small or closed groups, is usually suppressed in writing that has passed an editorial process. This means that most of the corpora used for lexicographical are lacking in this area. Today, however, the Internet has given us new possibilities to tap into to the flow of colloquial and informal language. The aim of this presentation is foremost to give a brief account of how the Internet could be 'harvested' for the purpose of creating corpora which include substantial amounts of informal language, and secondly, how to use these (in this case Swedish and Icelandic) corpora to gather candidates for headwords with informal markings such as ***coll.***, ***slang***, and the like. The topic of evaluation of results of this kind of work will also be touched upon. The work here presented has been done utilizing Sketch Engine, and strategies employed in using that tool are thence also accounted for.

> Building Russian Word Sketches as Models of Phrases

KHOKHLOVA, MARIA

1 – Computational Lexicography and Lexicology

Without any doubt corpora are vital tools for linguistic studies and solution for applied tasks. Although corpora opportunities are very useful, there is a need of another kind of software for further improvement of linguistic research as it is impossible to process huge amount of linguistic data manually. The Sketch Engine representing itself a corpus tool which takes as input a corpus of any language and corresponding grammar patterns. The paper describes the writing of Sketch grammar for the Russian language as a part of the Sketch Engine system. The system gives information about a word's collocability on concrete dependency models, and generates lists of the most frequent phrases for a given word based on appropriate models. The paper deals with two different approaches to writing rules for the grammar, based on morphological information, and also with applying word sketches to the Russian language. The data evidences that such results may find an extensive use in various fields of linguistics, such as dictionary compiling, language learning and teaching, translation (including machine translation), phraseology, information retrieval etc.

> A Quantitative Evaluation of Word Sketches

KILGARRIFF, ADAM; KOVÁŘ, VOJTĚCH; KREK , SIMON; SRDANOVIC, IRENA AND TIBERIUS, CAROLE

1 – Computational Lexicography and Lexicology

A word sketch is an automatic corpus-derived summary of a word's grammatical and collocational behaviour. Word sketches were first prepared in 1999 for the compilation of the Macmillan English Dictionary for Advanced Learners (Rundell 2002). They have since been integrated into the Sketch Engine corpus query tool (Kilgarriff et al 2004), prepared for fifteen languages, and used on a large scale for lexicography by a number of publishers. We are frequently told how impressive they are and how little they miss – but we would like a more rigorous assessment.
We describe a formal, quantitative evaluation of word sketches, from a user perspective, for four languages (Dutch, English, Slovene, Japanese), with the critical question being 'is the collocation suitable for inclusion in a published collocation dictionary'. For each language, we inspected twenty collocates for each of forty-two headwords. In each case two thirds or more of the collocations were of publishable quality.

> Semantic Relations in Cognitive eLexicography

KREMER, GERHARD AND ABEL, ANDREA

1 – Computational Lexicography and Lexicology

Whereas dictionary design has traditionally been guided by the results of dictionary use research, recent approaches in lexicographic research are strictly user-centred. We support the idea of integrating empirical cognitive evidence into this type of research, thus fruitfully exploiting it for both, the selection (and subsequently presentation) of lexical data and the acquisition of such data from corpora. Focusing on the extraction of semantic relations to be illustrated in electronic learners' dictionaries, we analyse the results of two behavioural experiments on the production as well as the perception of semantic relations. The main goal of the experiments was to determine which relations are cognitively salient in speakers' minds. With the objective of developing a method to automatically extract cognitively salient semantic relations from corpora, we describe and discuss findings of the first analyses conducted on composite part relations. In future this might serve as a basis for the elaboration of new strategies aimed at enriching lexical databases and dictionaries.

> The IDM Free Online Platform for Dictionary Publishers

LANNOY, VINCENT

1 – Computational Lexicography and Lexicology

Printed dictionaries have built a genuine identity over the years. Lexicographers work for renowned publishers according to specific rules and processes; distribution channels are well-organized and efficient at delivering to educational or public markets. The emergence of new actors, exclusively focused on the Web, is a major upheaval as they deliver large corpora to a worldwide audience. Those Pure Players are now dominating the online dictionary market not only in terms of audience but also by establishing their own brands, independent of existing print brands.

These new actors bring their own vision of what an online dictionary should be. This presents a great opportunity for the industry to rethink the way dictionaries are written and published, inspired by the distinctive strengths of the Internet as a medium which call for clarity of the information, easiness of the service, and above all, intrinsic value of linguistic, i.e. lexicographic data.

Our experience, built through day-to-day management of several major free online dictionary websites, demonstrates the strong draw of dictionary content. Since dictionary websites encompass a very broad spectrum of the language and make it available for free on the Internet, users discover online dictionaries by very diverse means. Their distinct paths to a dictionary reflect their different interests in the content, and also their different expectations for the content delivered.

Making dictionary data amenable to favourable placement in search engines, for searches made in many languages, requires close involvement of lexicographers. These lexicographers must adapt to a process of creating entries for dynamic display on screen in addition to static display in print; understanding the impact of Search Engine Optimization (SEO) on entry structure; integrating a rich network of hyperlinks and making use of non-textual media to enrich their lexical content. Lexicographers are in the spotlight of the digital paradigm!

Quality of the content and publishers' care over data play a key role in building user loyalty and depth of visit on the Website. On average, in a language learning context, we observe that visits last between 5 and 7 pages, providing the publisher with the opportunity to be in contact with its users for several pages. The question is *to do what?* For the moment, most of the dictionary websites are dead ends: a user enters for one or several definitions and leaves though his needs or interests can be much deeper. He may require course books, vocabulary lists, exercises for learners, novels, reference content, etc. Affiliation models help propose not only the publisher's own content but complementary contents, products or services coming from partners. We are currently successfully experiencing with a partner the efficiency of an up-sales model based on dictionary free entries. Dictionary content is not only an efficient attraction point but plays also the role of a *user qualification filter for targeted up-sales.* Dictionary is an intermediary between a query and a targeted product.

> Constructing a Constructional MWE Lexicon for Psycho-Conceptual Annotation: An Evaluation of CPA and DUELME for Lexicographic Description

LUDER, MARC AND CLEMATIDE, SIMON

1 – Computational Lexicography and Lexicology

The German JAKOB lexicon provides a basis for the coding of patient narratives and is currently extended in the direction of a phraseological and construction-grammar resource. For this purpose, we will compare two formalisms for the representation of multiword expressions (MWE): The Dutch Electronic Lexicon of Multiword Expressions (DuELME, Grégoire 2009) and the verb patterns from Corpus Pattern Analysis (CPA, Hanks 2008). We are looking for a representation format which is human-readable, and equally adapted for natural language processing (NLP). The JAKOB lexicon is implemented in the OLIF format and currently contains 7000 entries. The MWEs investigated are verbal phraseologisms and originate from the corpora of three different clients, consisting of a total of more than 400 transcribed sessions.

The narrative analysis method JAKOB is a tool for investigating everyday stories from psychotherapy transcripts (Boothe, 2004). Stories are annotated on the basis of our predefined psycho-conceptual coding system represented in the lexicon. JAKOB allows formulating hypotheses about the client's conflicts, the analysis of the discourse being one component thereof.

DuELME is an NLP lexicon project which encodes MWE descriptions in a theory- and implementation-independent way. Every MWE is an instance of a construction class with elements including morphosyntactic parameters. CPA patterns represent semantic properties for the elements of a (verbal) construction, whereas syntactic properties are represented in the JAKOB lexicon by the subcategorization frames (Satzmuster) of Wahrig (2007). We are implementing an additional lexicon property 'bauplan' which is formally constructed as a combination of the DuELME component list, the Wahrig subcategorization frame and semantic information out of the CPA-pattern. Because this structure is difficult to read for the lexicographer, it is generated automatically and can be hidden from the user, but is available for NLP tasks.

> Une nouvelle ressource lexicographique en ligne: le Petit Larousse Illustré de 1905

MANUÉLIAN, HÉLÈNE

1 – Computational Lexicography and Lexicology

Cet article présente une nouvelle ressource lexicographique ancienne mise à disposition sur Internet : le Petit Larousse Illustré de 1905. Faisant suite à des œuvres de plus grande ampleur et de plus grande renommée (le dictionnaire critique de Féraud, le dictionnaire de Nicot, les différentes éditions de celui de l'Académie, etc.), le Petit Larousse Illustré de 1905, bien plus modeste que ses prédécesseurs – en volume tout au moins – a été numérisé et sera mis en ligne prochainement.

L'intérêt de la mise en ligne d'une telle ressource réside dans sa nature. Il s'agit d'un petit dictionnaire illustré, et la présence d'images est importante. Par ailleurs, il est le premier d'une série de dictionnaires grand public, ce qui le rend fondamental dans l'histoire de la lexicographie.

L'informatisation s'est déroulée en plusieurs phases, de façon à permettre une interrogation fine du dictionnaire. Les différents éléments des articles du dictionnaire ont été décrits et listés, puis balisés en XML selon les standards décrits dans la proposition 5 de la TEI. Le texte a ensuite été balisé automatiquement grâce à des programmes écrits en langage Python contenant des expressions régulières. Le balisage s'est déroulé en trois passes, chacune exploitant le résultat de la précédente.

Le résultat de l'informatisation est une base de données lexicales riche qui permet à l'utilisateur deux sortes de consultations : il peut choisir de faire une interrogation plein texte. Dans ce cas, le résultat apparaîtra avec les images associées aux articles répondant à sa requête. L'utilisateur peut aussi faire une recherche avancée, c'est-à-dire n'interroger qu'un seul champ de l'article du dictionnaire (vedette, prononciation, information grammaticale, étymologie, définitions, définitions encyclopédiques, renvois, proverbes, exemples, expressions figées). Seules les requêtes sur la vedette permettent l'affichage des images.

> Getting Synonym Candidates from Raw Data in the English Lexical Substitution Task

McCarthy, Diana; Keller, Bill and Navigli, Roberto

1 – Computational Lexicography and Lexicology

Distributional similarity provides a technique for obtaining semantically related words from corpus data using automated methods that compare the contexts in which the words appear. Such methods can be useful for producing thesauruses, with application to work in lexicography and computational linguistics. However, the most similar words produced using these methods are not always near synonyms, but may be words in other semantic relationships: antonyms, hyponyms or even looser 'topical' relations. This means that manual post-processing of such automatically produced resources to filter out unwanted words may be necessary before they can be used. This paper evaluates the performance of distributional methods for finding synonyms on the English Lexical Substitution Task, a lexical paraphrasing task where it is necessary to generate candidate synonyms for a target word and then select a suitable substitute on the basis of contextual information. We examine the performance of distributional methods for the first step of generating candidate synonyms and leave the second step of choosing a candidate on the basis of context for future work. A number of automated distributional methods are compared to techniques that make use of manually produced thesauruses. We demonstrate that while the performance of such automatic thesaurus acquisition methods is often below manually produced resources, precision can be greatly increased by using two automatic methods in combination. This approach gives precision results that surpass methods that exploit manually constructed resources for the same task, albeit at the expense of coverage. We conclude that such an approach to increase the precision of automatic methods to find near synonyms could improve the use of distributional methods in lexicography.

> What WordNet does not know about selectional preferences

Měchura, Michal Boleslav

1 – Computational Lexicography and Lexicology

Selectional preferences are the tendencies of words to co-occur with other words that belong to certain semantic types. In this paper, I will investigate how closely these corpus-attested preferences correspond to WordNet. For example, for all possible direct objects of *cancel*, is there a single category (or a union of several categories) in WordNet that subsumes them, and only them? Selectional preferences manifest themselves in authentic texts and can be revealed through corpus analysis. I will introduce an experimental tool I have built which attempts to do this automatically by aligning corpus-extracted lists of collocates (for example a list of the direct objects of *cancel*) with WordNet. The strength of this method is that it can discover and name selectional preferences automatically, but its weakness is that it can only do so when WordNet contains a suitable category. We will see that WordNet often lacks a category (or even a union of several categories) that fully corresponds to an attested selectional preference – for example, there is no category in WordNet that includes all the kinds of events that can be direct objects of *cancel* (*meeting, wedding, concert* etc.) but excludes those that cannot (*accident, sunset, invention* etc.).

> OWID – A dictionary net for corpus-based lexicography of contemporary German

MÜLLER-SPITZER, CAROLIN

1 – Computational Lexicography and Lexicology

The *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online German Lexical Information System) is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the Institute for the German Language (Institut für Deutsche Sprache, IDS). The main emphasis of OWID is on academic lexicographic resources of contemporary German. Presently, the following dictionaries are included in OWID: a dictionary of contemporary German called *elexiko*, a dictionary of neologisms, a small dictionary of collocations, and a discourse dictionary covering the lexemes that establish the discourse about 'guilt' in the early post-war era 1945-1955. In the near future (2010/2011), several additional dictionaries will be published in OWID: a Textbook of German Communication Verbs, a Valency Dictionary of German Verbs, two further discourse dictionaries – one about the 'democracy' discourse around 1968, the other covering the keywords of the German reunification 1989/1990. Moreover, 300 entries from a corpus-based project on proverbs will be integrated into OWID. Thereby, OWID is a constantly growing resource for academic lexicographic work of the German language.

Altogether, OWID is a special kind of dictionary portal owing to its content and its design, namely the integration of the various dictionaries, the access possibilities and the presentation features. With OWID, we try to establish a dictionary net where the different resources are jointly accessible not only by headwords, but also on the microstructural level. Prerequisite for these common access- and navigation-possibilities across the various dictionaries is the same concept for the lexicographic data model which we put into practice in OWID. Data from all dictionaries

in OWID are structured according to a tailor-made, fine-granular, XML-based data model. In this data model, similar content is modelled similarly, dictionary related differences are preserved.
The main tasks for the future are to enhance OWID with further dictionary resources, to improve the inner access structures so that they exhaust the possibilities of the data model, and to customize the layout of the dictionaries as well as the search options according to the user's needs.

> OBELEX – the 'Online Bibliography of Electronic Lexicography'

MÜLLER-SPITZER, CAROLIN AND MÖHRS, CHRISTINE

1 – Computational Lexicography and Lexicology

Digital or electronic lexicography has gained in importance in the last few years. This can be seen in the growing list of publications focusing on this field. In the OBELEX bibliography (http://www.owid.de/obelex/engl), the research contributions in this field are consolidated and are searchable by different criteria. The idea for OBELEX originated in the context of the dictionary portal OWID, which incorporates several dictionaries from the Institute for German Language (www.owid.de). OBELEX has been available online free of charge since December 2008.
OBELEX includes articles, monographs, anthologies and reviews published since 2000 which relate to electronic lexicography, as well as some relevant older works. Our particular focus is on works about online lexicography. Systematically evaluated sources are relevant journals like *International Journal of Lexicography*, *Lexicographica*, *Dictionaries*, *Lexikos*; furthermore *Euralex-Proceedings*, proceedings of the *International Symposium on Lexicography* in Copenhagen as well as relevant monographs and anthologies. Information on dictionaries is currently not included in OBELEX; the main focus is on metalexicography. However, we are working on a database with information on online dictionaries as a supplement to OBELEX.
All entries of OBELEX are stored in a database. Thus, all parts of the bibliographic entry (such as person, title, publication or year) are searchable. Furthermore, all publications are associated with our keyword list; therefore, a thematic search is possible. The subject language is also noted.
With this type of content, the OBELEX bibliography supplements in a useful way other bibliographic projects such as the printed 'Internationale Bibliographie zur germanistischen Lexikographie und Wörterbuchforschung' by H. E. Wiegand (Wiegand 2006/2007), the 'Bibliography of Lexicography' by R. R. K. Hartmann (Hartmann 2007), and the 'International Bibliography of Lexicography' of Euralex (cf. also DeCesaris and Bernal 2006). OBELEX differs from all these bibliographic

projects by its strong focus on electronic lexicography and its ability to retrieve bibliographic information.

> Towards Semi-Automatic Dictionary Making Creating the Frequency Dictionary of Hungarian Verb Phrase Constructions

PAJZS, JÚLIA AND SASS, BÁLINT

1 – Computational Lexicography and Lexicology

The paper describes the lexicographical aspects of creating a frequency dictionary by a semi-automatic process. The bulk of the work is made by task specific software. The output of the program is then manually checked, corrected and filtered. The result is a collection of the most frequent Hungarian verb phrase constructions (VPCs), illustrated by corpus examples. This is a corpus driven dictionary, based on the 187,6 million word synchronic Hungarian National Corpus (http://corpus.nytud.hu/mnsz) which was analyzed by a series of programs. Its output is a set of XML format draft entries, which were then hand validated and edited by lexicographers. The dictionary contains the most frequent Hungarian verbs along with their most typical syntactic constructions. At the current phase of the project we decided to collect the most frequent constructions only: their absolute frequency had to be more than 250. The dictionary contains roughly 2300 entries and 6500 VPCs. Each construction is illustrated by a corpus example. The verbal entries are presented in alphabetical order primarily. Different kinds of indices are also included in the printed version. The users of this dictionary envisaged to be mainly linguists, working on Hungarian grammars, lexicographers working on bilingual dictionaries and last but not least: advanced level learners of Hungarian, who want to expand their knowledge on the Hungarian nominal verbal collocation relationships. The dictionary is planned to be published both in printed and electronic format.

Parts of the algorithm used for this project could be applied to produce other dictionaries, all the more so, as some of them are actually language independent. It is also highly cost effective: both the programming and the lexicographic work required one person year each.

> Developing GiGaNT, a lexical infrastructure covering 16 centuries

RUITENBERG†, TILLY; DOES, JESSE DE AND DEPUYDT, KATRIEN

1 – Computational Lexicography and Lexicology

GiGaNT is a new INL initiative which sets out to develop a computational lexicon (lexical database) covering 16 centuries of Dutch language. This means that all lexical data of the dictionaries, corpora and computational lexica of the Institute for Dutch Lexicology (INL) will be stored into a central database, functioning both as computational lexicon and central infrastructure for the maintenance of lexical data. Dictionaries, corpora and this computational lexicon are all part of the Dutch Language Bank (DLB).

The immediate incentive to develop GiGaNT was the need for a diachronic computational lexicon, to serve both as a link between texts and dictionaries in the DLB and as a solid infrastructure for other, similar lexical data at the INL. The GiGaNT lexicon will be used for text or corpus annotation, facilitating the retrieval and investigation of the annotated texts.

Integration of existing material into GiGaNT and its subsequent adaptation to enable it to function within computational applications will be a huge step towards another aim: the systematic screening of the complete Dutch word stock for 'gaps' in lexicographic description. This applies to both neologisms and hitherto undescribed historical words.

Users will benefit from the possibility to link from word forms in running text to lexicographical definitions in the INL dictionaries. Researchers, who now only have access to separate collections, will benefit as well: in the future they will have one single starting point for their searches and one single basis from which to develop new lexical material. GiGaNT will also give expert users better access to the lexical data maintained by the INL. The infrastructure will function as a database which will be accessible to API's and as a 'service' that enables researchers to compare their data with GiGaNT and eventually to contribute their own material to GiGaNT.

> Electronic Dictionary and Dictionary Writing System: how this duo works for dictionary user's needs (ABBYY Lingvo and ABBYY Lingvo Content case)

RYLOVA, ANNA

1 – Computational Lexicography and Lexicology

The main idea we present in this paper is that using special markup in dictionary writing system and having appropriate functionality in electronic dictionary software we can achieve new results in satisfying most important needs of dictionary user. We describe the core functionality of the ABBYY Lingvo Content dictionary writing system and some features of ABBYY Lingvo electronic dictionary software that presents the dictionaries made in DWS. Then we show how the dictionary data can be used in text translation scenario and how DWS and electronic dictionary work together to meet the user's needs in translation and text analyzing.

Besides the core functionality ABBYY Lingvo Content DWS includes

- embedded in DWS interface user-friendly entry filtration system. Lexicographer doesn't need to know any special query language – just tick boxes in filtration window tabs;
- also embedded in interface tool for dictionary comparison and merge;
- visual markup of changes – you can always compare any two versions of dictionary entry and see what was added, deleted, changed or restored to their earlier versions.;
- possibility of working with many dictionaries (2 and more) in one window, editing their entries simultaneously.

ABBYY Lingvo *electronic dictionary* has been developed since 1989 and nowadays it is used by 7 million users worldwide.

One of the basic dictionary user's need is to find the appropriate translation for the word he met in a text (text reception) or translate a word from their mother tongue to a foreign language. Here we will describe how the electronic dictionary software works to satisfy the text reception need. One of the most challenging task for a dictionary producer – to help the dictionary reader find a good translation and all the relevant information about the word. This task can be done well if a lexicographer puts relevant markup for a dictionary entry in DWS and electronic dictionary has a proper functionality to process this markup and a good interface to show the result of this processing to dictionary user.

> The Cornetto database: Semantic issues in linking lexical units and synsets

VLIET, HENNIE VAN DER; MAKS, ISA; VOSSEN, PIEK AND SEGERS, ROXANE

1 – *Computational Lexicography and Lexicology*

Cornetto is a lexical semantic database that combines the Dutch Wordnet (Vossen (1998)) and the Referentie Bestand Nederlands (Van der Vliet (2007)). The Dutch Wordnet (DWN) is similar to the Princeton Wordnet for English (Fellbaum (1998)), and the Referentie Bestand Nederlands (RBN) includes frame-like information as in FrameNet (Fillmore, Baker, Sato (2004)) as well as information on the combinatorical behaviour of word meanings. The combination of the lexical resources has resulted in a rich relational database that may improve natural language processing technologies.

An important aspect of combining the resources is the alignment of the lexical units (LU's) and the synsets. Automatic alignment of RBN and DWN resulted in an initial version of the Cornetto database. This version has been further extended both automatically and manually. The resulting data structure is stored in a database that keeps separate collections for LU's (mainly derived from RBN), synsets (derived from DWN) and, in addition, a formal ontology (SUMO/MILO, see Niles and Pease (2001)). These 3 semantic resources represent different viewpoints and layers of linguistic and conceptual information. The resulting resource is freely available for research in the form of an XML database.

In this contribution, we will concentrate on the semantic information in Cornetto. We will discuss the differences in the perspective on semantics in the LU's and synsets and we will give a brief overview of the differences with regard to semantic information. The merging of the two resources resulted in very rich semantic database. However, combining lexica with different perspectives on semantics causes specific problems in the alignment of LU's and synsets and leads to findings that shed light on the organization of meaning in the lexicon.

> ¿Lo que necesitan es lo que encuentran? Reflexiones a propósito de la representación de los verbos en los diccionarios de aprendizaje del español

BERNAL, ELISENDA AND RENAU, IRENE

2 – *The Dictionary-Making Process*

The verb is one of the most analysed parts of speech in lexicographical research and, as a result, the tendency of establishing and putting into practise microstructure models that include more and more grammar appears to be consolidated. With respect to Spanish foreign learners (ELE) dictionaries, this tendency is still in an initial stage.

We believe it is necessary to pay more attention to users, in order to provide them with this grammatical information, which is required for production. In this sense, we present the results of an experiment with intermediate-advanced students of ELE, to determine if the *Diccionario de español como lengua extranjera* (DAELE, *Spanish Dictionary for Foreign Learners*) that we are developing, in fact satisfies users' needs in this respect. DAELE is an online dictionary that is fully based on corpus analysis is aimed at advanced learners.

We tested two groups of ELE from the Universitat Pompeu Fabra (Barcelona, Spain). In both cases, a control group was used. The test consisted of two exercises, a composition task and a questionnaire in which participants were asked to give their opinions about the use of the dictionary that they consulted.

Results of the experiment show that there are no significant differences related to the number of correct answers of the DAELE groups with respect to the control groups. We find, however, qualitative differences with regard to what students miss or value in every dictionary. The test confirms that the current approach taken in the preparation of DAELE, in which we aim to offer users the possibility of expanding or reducing the amount of information they see in response to a search, and to give them grammatical indications that are easier to understand and better suited to their needs.

> An innovative medical learner’s dictionary translated by means of speech recognition

EERENBEEMT, ARNOUD VAN DEN

2 – *The Dictionary-Making Process*

I will discuss a medical dictionary based on the Keyword in context (KWiC) concept and speech recognition as a valuable tool. My daily work consists of compiling medical dictionaries for students and professionals (creating and updating complex and dynamic data) and creating medical spellcheckers.

Non-Anglophone medical students and health care professionals around the globe need an active command of professional English for their career. Yet the lexical tools available for acquiring these skills are few and insufficient: American and British explanatory dictionaries expect readers to be native speakers, while bilingual medical dictionaries are basically glossaries and provide unlabelled translations.

The only medical learner’s dictionary in the world to date is the excellent Fachwortschatz Medizin by Michael and Ingrid Friedbichler, teaching English for medical purposes (EMP) in Austria. Their opus magnum, which helps non-native speakers to acquire language skills step by step, is structured using modular medical concepts and combines various lexical features:

- a monolingual dictionary: 100,000 medical terms grouped into 1400 sections with key headwords defined in simple English; contextualized with collocations and sample sentences demonstrating correct use, extracted from a 20-million-word corpus of medically authoritative texts;
- a semi-bilingual dictionary: support in the user’s native language (German, Dutch) in the form of 42,000 translated keywords;
- a thesaurus: synonyms, antonyms and related terms;
- a domain-specific glossary: readers from all medical fields can focus on content relevant to their specialization;
- Windows edition: full-context search, customizable display (pronunciation, definition, translations, collocations), cross-references etc.

After acquiring the Dutch rights I realised that farming out the translation work would require me to extensively monitor the translators, who were discouraged by the highly condensed lexical content. I decided to translate the 42,000 indexed medical terms myself instead, using speech recognition and a 24" HD monitor to display my database content, a web browser, a word processor and two medical dictionaries. I developed voice-driven macros for automating 600,000 Google searches, creating 2000 records, searching dictionaries and my 52,000-record medical database etc. This allowed me to translate up to 400 terms per day.

www.pinkhof.nl/medisch-engels: full Windows edition 10-day trial period, free download of 60-page sample PDF

> Thinking out of the box – perspectives on the use of lexicographic text boxes

GOUWS, RUFUS H. AND PRINSLOO, DANIE J.

2 – *The Dictionary-Making Process*

Although text boxes have become a common phenomenon in dictionaries relatively little attention has been paid to their presentation and to the motivation for their use and the type of data to be included in a dictionary in this specific way. Text boxes are salient dictionary entries and as such they are used to place more than the default focus on a specific data item. Dictionaries offer a variety of data types in text boxes such as guidance in terms of sense, contrasting related words, restrictions on the range of application, register, pronunciation, et cetera. The default presentation seems to be as article-internal microstructural entries within a typical relation of lemmatic addressing. Whereas some text boxes present data relevant to only the specific article, other text boxes, i.e. those with a synoptic assignment, also have relevance for other articles, namely a hybrid addressing relation, presenting both immediate and distant addressing. As devices employed in an extended compulsory microstructure care should be taken that text boxes do not become part of the compulsory microstructure and in so doing lose their significance and decrease the emphasis on the data included in the text boxes. The added value of text boxes may never be undermined by an over exposure of this device. Using both micro- and macrostructural text boxes offers exciting possibilities. Where dictionaries have a text production function data could be included in a text box to emphasise the use or non-use of certain combinations and collocations as well as proscriptive guidance. Of real importance is that lexicographers should realise that text boxes are lexicographic devices that can really enhance the data transfer in dictionaries. Lexicographers should think out of the box and they should get out of the box of tradition and employ text boxes in bold, innovative and functional ways.

> Guiding principles for the elaboration of an English-Spanish dictionary of multi-word expressions

GREGORIO-GODEO, EDUARDO DE

2 – The Dictionary-Making Process

So-called *word combinations* – also referred to as *multi-word combinations* or *multi-word expressions* – take shape when certain words regularly combine with certain other words or grammatical constructions. When exploring the word combinations of a language, both collocations and idiomatic expressions to a large extent examined. Collocations and idioms are usually taken to be multi-word expressions whose meaning is more than the sum of the meaning of their components.

Focusing on multi-word expressions in bilingual dictionaries, this contribution accounts for an ongoing research and editorial project guiding the elaboration of an English-Spanish dictionary of multi-word combinations. After presenting the lexicographic process leading the elaboration of the dictionary as such, this contribution will proceed to describe the principles determining the inclusion of entries and their presentation in the dictionary.

The rationale for this project is based on current lexicographic practices (Hartman 2001) having comprised four stages: (1) pre-lexicographic work, which consisted of a thorough examination of the market of English-Spanish dictionaries given the lack of specific dictionaries dealing with multi-word expressions in this area; (2) the research undertaken for the elaboration of the macrostructure of the dictionary and the use of various sources (e.g. existing English monolingual or multi-word dictionaries; bilingual dictionaries, and corpora), especially as far as usage examples, equivalents and their idiomaticity is concerned; (3) description issues, with a special emphasis on both the description of the multi-word expressions included in the dictionary, and the actual structure of dictionary entries; and (4) final formatting, which entails final presentation and revision prior to editing and publishing the dictionary.

Considering Spanish-speaking students of EFL and – to a lesser extent – translators as the potential users of this dictionary, this contribution will conclude with some final remarks of the educational implications of the project herein presented.

> La segunda y tercera ediciones del *Diccionario Básico Escolar*

MIYARES BERMÚDEZ, ELOÍNA; ARTOLA ZUBILLAGA, XABIER; ALEGRÍA LOINAZ, IÑAKI; ARREGI IPARRAGIRRE, XABIER; RUIZ MIYARES, LEONEL; ÁLAMO SUÁREZ, CRISTINA AND PÉREZ MARQUÉS, CELIA

2 – *The Dictionary-Making Process*

En julio del 2003 se publicó la primera edición del *Diccionario Básico Escolar* (DBE), obra desarrollada en el Centro de Lingüística Aplicada de Santiago de Cuba y orientada a un mejor dominio del idioma español por parte de sus destinatarios: estudiantes del segundo ciclo de primaria (5to y 6to grados), secundaria básica y preuniversitario.

Gracias a la inestimable colaboración del Grupo IXA de la Universidad del País Vasco y al Instituto Cubano del Libro de Cuba, se presentó la posibilidad de realizar la segunda y tercera ediciones del DBE, por lo que nuestro grupo lexicográfico emprendió la laboriosa y complicada tarea de mejorar y arreglar algunas entradas, además de agregar nuevos artículos a esta importante obra de consulta.

El *Diccionario Básico Escolar* está disponible en tres soportes: impreso, en CD y en INTERNET (http://ixa2.si.ehu.es/dbe/index.html) y la segunda y tercera ediciones del mismo incluyó la completa revisión de sus tres soportes.

Los diccionarios son 'organismos vivos'; un diccionario que posea varias ediciones tiene que revisarse constantemente, pues siempre habrá entradas que mejorar, otras que añadir y corregir los errores humanos, hasta llegar a una obra casi perfecta.

En este trabajo pretendemos analizar el entorno de edición de diccionarios *leXkit*, las características de la segunda y tercera ediciones del *Diccionario Básico Escolar,* sus resultados y una comparación con la primera edición, donde se demuestra la 'vitalidad' de esta herramienta lingüístico-pedagógica.

> The Living Lexicon: Methodology to set up Synchronic Dictionaries

NAZAR, ROGELIO AND AZARIAN, JENNY

2 – *The Dictionary-Making Process*

In this paper, we want to investigate the subset of the vocabulary of a given language or dialectal variant which is in actual use in the discourse of a linguistic community in order to set up a synchronic dictionary. The aim of this article is, thus, to develop a methodology for acquiring the nomenclature of synchronic dictionaries in a systematic way. To do this, we consider two kinds of operations: addition of entries –the birth of words, or Neology- and removal -the death of words, Desuetude, as we

call it here. The methodology consists in contrasting dictionaries of a language (or dialectal variant) to find the intersection of the vocabulary, and to compare the vocabulary of the dictionaries with the vocabulary of a diachronic corpus. Such a methodology enables us to answer the following research questions: 1) what proportion of the vocabulary is shared by most dictionaries, 2) what proportion of units of each dictionary is no longer in use and 3) what proportion of the vocabulary units in use today is still not registered in the dictionaries. These three questions are central to the definition of the ideal headword. In a pilot experiment in Peninsular Spanish, we combine the study of the main dictionaries of this language variant with diachronic studies using corpus statistics on Spanish newspaper archives.

> Lingvo Universal English-Russian Dictionary: Making a Printed Dictionary from an Electronic One

ANOKHINA, JULIA

3 – Reports on Lexicographical and Lexicological Projects

Lingvo Universal English-Russian Dictionary (*Lingvo UERD*) was the first electronic English-Russian dictionary published in Russia. It appeared in 1990, as part of the *Lingvo* software produced by the company *ABBYY*. Unlike many other dictionaries available on *Lingvo*, which are licensed electronic versions of high-quality paper editions, *Lingvo UERD* is the fruit of the company's own lexicographic research. As the dictionary database grew further, it was transformed into a multifunctional database, used to produce different kinds of dictionaries. The first printed edition based on its content was the *ABBYY Lingvo Comprehensive English-Russian Dictionary*, published in 2007. It was designed as an English-Russian dictionary for professional users and advanced learners; the dictionary entries were edited in the in-house DWS *ABBYY Lingvo Content*. The 2nd revised edition of it was initiated in 2008 by the publishing house *ABBYY Press*; the current article reports on this project.

While preparing the dictionary lexicographers faced a whole range of problems related to different access to electronic vs. printed dictionary data, to different user tasks while accessing them and to the specific character of the *Lingvo* software format. Many difficulties were solved by *ABBYY* programmers who adjusted export algorithms of the DWS and improved its interface. All those improvements were added to the latest version of the DWS *ABBYY Lingvo Content*.

Present-day dictionary databases tend to include as much linguistic data as possible in order to be used as a basis for different kinds of dictionaries, including printed editions. As an electronic database is a big hypertext comprising multiple links and different kinds of specific data which cannot be exported to the 'paper' format, making a paper dictionary from such a database may be quite a challenging task. Working hand in hand with the publishing house editors enabled us to minimize the inevitable losses resulting from such a procedure. The other result of this work was the creation of a printed dictionary more in line with the needs of modern users, presented in a more convenient and user-friendly way.

> Database of ANalysed Texts of English (DANTE): the NEID database project

ATKINS, B.T. SUE; KILGARRIFF, ADAM AND RUNDELL, MICHAEL

3 – Reports on Lexicographical and Lexicological Projects

DANTE is a lexicographic project where the end product is not a dictionary but a lexical database resulting from in-depth analysis of corpus data. The users of DANTE are not the dictionary-using public but the lexicographic teams who will develop dictionaries and computer lexicons from it. This project is the source-language analysis stage of the New English-Irish Dictionary (NEID: http://www.forasnagaeilge.ie/), being developed for Foras na Gaeilge, Dublin (FnaG: http://www.forasnagaeilge.ie/). The project was designed and carried out by the Lexicography MasterClass (http://www.lexmasterclass.com). The database covers approximately 50,000 headwords and 45,000 compounds, idioms and phrasal verbs, using over 40 datatypes in their lexical description. It was created in the course of 2.5 years by LexMC's 15-strong lexicographic team, managed by Valerie Grundy, Managing Editor; the project administration is in the hands of Diana Rawlinson, Project Administrator.

What makes DANTE special is the application of an existing methodology across the whole lexicon, extremely systematically and at an unprecedented level of detail. Amongst other aspects of this project, we describe:

- improving the reliability of schedule and workflow by classifying, before the compiling started, over 50,000 headwords according to type and complexity;
- the systematic use of 68 model 'template' entries;
- a new approach to quality control, combining conventional entry-editing by senior team members with the use of complex search scripts that list all entities of a specific type and allow rapid checking for accuracy;
- the customisation of the Sketch Engine (http://www.sketchengine.co.uk/) corpus query software, with a corpus of 1.7bn words;
- the use of IDM's Dictionary Production System (DPS: http://www.idm.fr/products/dictionary_writing_system/27/)..

The DANTE database is a rare, possibly unique, beast: a rich and comprehensive lexicographic analysis on linguistic principles, prepared on a substantial budget by a large team of professional lexicographers, and uncompromised by the needs of accessibility to non-linguist users.

> Quo Vadis Lexicography at the Institute for Dutch Lexicology?

BEEKEN, JEANNINE

3 – Reports on Lexicographical and Lexicological Projects

In this paper, we will first introduce the Institute for Dutch Lexicology. We will present an overview of the INL-dictionaries online, being the Dictionaries of Old Dutch (ca. 475 – 1200), Early Middle Dutch (1200 – 1300), Middle Dutch (1250 – 1550), the Dictionary of the Dutch Language (WNT, 1500 – 1970s), the Etymological Dictionary of Dutch, the General Dutch Dictionary (ANW, 1970s till 2015). Thirdly, we will present the Language Bank (Taalbank Nederlands) and its main tasks. Finally, we will elaborate on three U-turns, namely a first U-turn: from manual labour and printed material to computational linguistics and the internet, a second U-turn: from single functionality to multiple functionality, a third U-turn: from stand-alone product to spin-offs, linking and integration. We will finish with some thoughts and ideas answering the following question: quo vadis lexicography at the INL?

> Time to say goodbye? On the exclusion of solid compounds from the Swedish Academy Glossary (SAOL)

BERG, STURE; HOLMER, LOUISE AND SKÖLDBERG, EMMA

3 – Reports on Lexicographical and Lexicological Projects

The Swedish Academy Glossary, SAOL (short for *Svenska Akademiens ordlista*) is a monolingual glossary, first published in 1874. The latest edition, SAOL13, was published in 2006 and the next edition, SAOL14, is planned for 2015.

This article concerns the revision of the lemma list in SAOL, with special focus on the exclusion of transparent solid compounds. There are about 88,000 solid compounds in the 13th edition of the Glossary, i.e. 70 % of the total number of lemmas (125,000). Since there are almost infinite possibilities of creating new words in Swedish, the printed Glossary obviously only includes a sample of the contemporary Swedish vocabulary.

With improved lexicographic tools and an enlarged text corpus, the editors of SAOL14 have great possibilities of making more accurate decisions when including new solid compounds and excluding others from the lemma list. The discussion is above all based on the solid compounds including the noun *kalkyl* ('calculation', 'estimate', 'calculus').

> FACKELLEX – Zur Struktur des Schimpfwörterbuches

BREITENEDER, EVELYN

3 – Reports on Lexicographical and Lexicological Projects

In 2008, the work on the 'Schimpfwörterbuch' (Dictionary of Insults and Invectives), the second part of the Fackellex dictionary, was brought to an end. FACKELLEX is a so called dictionary in the field of textlexicography. The 'Schimpfwörterbuch' was compiled in the planned tripartite structure, developing the dictionary volumes named ALPHA, CHRONO and EXPLICA. The paper will show different methods of presenting text information in a dictionary.

During the work on the dictionary, some 200.000 invective expressions were identified on the 22.586 pages of the 'Fackel' ("The Torch') edited and written by Karl Kraus, 2775 of which were selected to be represented as keywords in the ALPHA volume – the alphabetic list of the dictionary. Selection was performed according to linguistic and semantic criteria, keywords were furnished with short excerpts from the original text. Main tasks during this phase of the project were the constitution of a list of candidates and the presentation of the extensive material. The focus in the work on the ALPHA section was the description of invective terms, particular constructions and examples of Karl Kraus's creativity in coining new words making use of text lexicographic methods. ALPHA is made accessible through three different indexes which were created making use of up-to-date IT technology as part of a cooperation within the 'Centre for Cultural Research' between the departments of AAC and Fackellex.

CHRONO – the chronological list of the 'Schimpfwörterbuch' – displays in chronological order roughly a fifth of the data contained in ALPHA and offers the reader a larger context. This part of the dictionary is designed to pursue the development of the author's creativity in coining and using words within the ›Fackel‹ as a whole and to display these phenomena in the context of a particular page of the journal.

EXPLICA – short for explanatory notes – is to fulfill two requirements: it contains the dictionary editor's explanatory texts on which the project was based, which were written as part of the seperate 'PARATEXTE' project. In addition, it contains ›Wichtiges von Wichten‹, the last article of the ›Fackel‹ which can be seen as the primary source of inspiration for this ›Schimpfwörterbuch‹. Passages of this text that were identified as invectives where highlighted and commented upon in a selective way.

> Improving the representation of word-formation in multilingual lexicographic tools: the MuLeXFoR database

CARTONI, BRUNO AND LEFER, MARIE-AUDE

3 – Reports on Lexicographical and Lexicological Projects

This paper introduces a new lexicographic resource, MuLeXFoR, which aims to present word-formation processes in a multilingual database designed for both language specialists (e.g. linguists, terminologists, lexicographers, NLP specialists) as well as second-language (L2) learners and trainee translators. Morphological items (e.g. affixes, compound parts, combining forms) and processes (prefixation, suffixation, compounding, conversion, etc.) pose major challenges for lexicographic work, especially with respect to the design of bilingual and multilingual resources. It is well-known that derivational affixes can take part in several word-formation rules and that, conversely, rules can be realised by means of a variety of affixes. In view of this complexity, it is often difficult to (1) provide enough information to help users understand the meaning(s) of an affix and the (near-)synonymy relations between affixes and (2) become familiar with the most frequent strategies used to translate the meaning(s) conveyed by these affixes. In fact, traditional dictionaries often fail to achieve this goal. The MuLeXFoR database tries to take advantage of recent advances in morphological description and the development of electronic multi-access database systems. The database relies on the lexematic approach to word-formation, which is especially helpful to represent morphological processes cross-linguistically. In addition, it has been entirely implemented in a multi-access database interface. The prototype described in this paper so far centres around prefixation in English, French and Italian. Two interfaces are currently available: a comprehensive interface aimed at morphological and lexicographic investigations by language specialists (*MuLeXFoR-Linguists*) and a second interface designed for second-language learners or trainee translators (*MuLeXFoR-Learners*).

> Author Dictionaries Revisited: Dictionary of Bohumil Hrabal

ČERMÁK, FRANTIŠEK AND CVRČEK, VÁCLAV

3 – Reports on Lexicographical and Lexicological Projects

With a view to continue the line of author dictionaries, started by that devoted to Karel Čapek (2007), a second dictionary, basically following the first, has been compiled, namely that of Bohumil Hrabal (2009), an influential and major figure of the contemporary literary scene. The idea to have more of comparable and corpus-based dictionaries of this type that would ultimately enable comparison and through the prism of some of the best masters of the language to view the Czech language in development, has been made possible only recently, with the existence of corpora and thanks to techniques developed by corpus linguistics. A number of new lexicographic and computational features, never used before (with the exception of K. Čapek's dictionary), have been tried verifying options how to best put into practice general theoretical ideas, such as when finding best collocations that could be included in the dictionary.

> The Faroese-Italian Dictionary - An attempt to convey linguistic information concerning the Faroese language as well as information about the culture of the Faroe Islands

CONTRI, GIANFRANCO

3 – Reports on Lexicographical and Lexicological Projects

In 2004 Føroya Fróðskaparfelag, the Academy of the Faroe Islands, published the *Dizionario Faroese-Italiano / Føroysk-Italsk orðabók*, the first bilingual dictionary of the Faroese and Italian languages. The dictionary has 632 pages and includes 14.850 headwords, plus a few hundred subheadwords within the relevant single entry. It was Professor Jørgen Stender Clausen of Pisa University in Italy who suggested that I compile this dictionary, and I carried it out working at the Department of Faroese Language and Literature of the University of the Faroe Islands. The dictionary is the result of some years' work and of the indispensable advice I was given by the lexicographical consultant Jógvan í Lon Jacobsen, and of the help of several Faroese and Italian collaborators. The dictionary is an attempt to create a practical means to help an Italian-speaking visitor or student to make acquaintance with both the Faroese language and the culture of the Islands, and also useful for Faroese meeting Italian-speakers on their travels. The differences between the two languages (with some structural features showing diversity), and the cultural differences between the linguistic areas (the respective everyday vocabulary is different), had to be dealt lexicographically. The compilation of the dictionary has not followed any existing lexicographic model: the list of headwords, the structure of the entries and the graphics are the result of research and experiment. One result, among others, is that many terms are related to the needs of a visitor or student interested not only in the language of the Faroe Islands but also their culture, and that therefore some entries are a combination of linguistic and 'cultural-encyclopedic' information.

> Covering All Bases: Regional Marking of Material in the New English-Irish Dictionary

Convery, Cathal; Ó Mianáin, Pádraig and Ó Raghallaigh, Muiris

3 – Reports on Lexicographical and Lexicological Projects

The New English-Irish Dictionary is a government-sponsored project that began in 2000 and is due for completion in 2012. The aim is to produce a modern bilingual dictionary containing c. 40,000 headwords which is to be published in both printed and electronic formats. When published this dictionary will be the first major dictionary published for Irish in over 40 years. The project is currently at the translation phase, and this paper focuses on the approach taken to attempt cover dialect variations in the modern spoken language. The methodology employed was divide the headword list into three distinct categories, each requiring a different level of translation.

Given the time and budgetary constraints of the project it was decided that only the 1000 (approx) most frequently occurring lemmas could receive a full dialectal profile. Translators from each of the three main dialects translate each entry, passing the entry on to a translator from the next dialect as they complete their part of the process. This translation work is carried out without reference to written sources. Once a translator from each main dialect has completed their work the entry is checked for completeness against set sources and labelled accordingly.

The main advantages of this process are as follows.

– It captures current translations that may not be covered in existing outdated sources.
– It provides a dialectal profile of words, phrases and usages.
– It enables an element of dialectal marking in the final product, particularly in the electronic version.
– It enables the option to customise the electronic version, fronting any particular dialect.
– A given dialect may be selected as the default pronunciation in the electronic version.
– It enriches the bilingual database creating a useful research resource for other academic research projects.

> Software Demonstration of the Dictionary of the Flemish Dialects and the pilot project Dictionary of the Dutch Dialects

De Tier, Veronique and Van Keymeulen, Jacques

3 – Reports on Lexicographical and Lexicological Projects

A. Dictionary of the Flemish Dialects

The relational database of the *Dictionary of Flemish Dialects* works under

Oracle. The WVD input database (*bronsoorten* = sources) consists of subdatabases of one or more questionnaires. Once all the data have been put into the correct subdatabases, the lexicographer can start compiling the dictionary by selecting the concepts that are to be put in one particular fascicle, e.g. all the selected concepts for 'sheep'. This is done in a new database structure, called 'publications'. After selecting the data from the different sources, the lexicographer automatically can generate a dictionary article with all the results for one concept. From this point onwards, the dictionary article can be compiled and lexicological decisions have to be made. This database also forms the basis for drawing the word maps in MapInfo and for making a text file ('Wetenschappelijk Apparaat' (Scientific Database)) in which every entry with the different lexemes for that particular concept is followed by codes that indicate the location of the words.

B. Pilot project : the *Dictionary of the Dutch dialects*

The software of the *Dictionary of Dutch Dialects* works under the oracle platform as well. After digitizing dialect dictionaries by scanning and ocr-ing and after correcting the Word files of these dictionaries, the headwords are put into bold and two Hard Returns are inserted after each dictionary article. This Word document, converted into a standard XML file, can be imported into the database through the application built for this database. For each dictionary it is necessary to write a new custom script, which generates the XML-file by means of typographical conventions. Once the XML-files are uploaded, the database of the *Dictionary of the Dutch Dialects* can be made. The editors then may enrich the database with dutchification, translation and markers. The next step is to connect this database to a website with search facilities.

> The Style Manual for Monolingual Lụgbarati Dictionary

DRAMANI, SAIDI

3 – Reports on Lexicographical and Lexicological Projects

To compile a monolingual general-purpose Lụgbarati dictionary, a Style Manual based on the format of Makerere Institute of Languages was developed (Kiingi, 2004). It was the blue print for the process of compilation. Lụgbarati terminology for linguistics has hitherto been lacking. Words were coined using functions of the word classes. The coinages were used to give ancillary information on the lexical items being defined. The research involved developing a style manual, compiling the dictionary, testing it for acceptability, and analysing the testing outcomes. The corpus used was a

198-page list of vocabulary in Crazzolara's book; *A Study of Lugbara (Ma'di) Language* (1960:175-373), and a 25-page list of words in Dalfovo's collection of Lụgbara proverbs; *Lugbara* (sic) *Proverbs* (1984:249-274).

> An inverted loanword dictionary of German loanwords in the languages of the South Pacific

ENGELBERG, STEFAN

3 – Reports on Lexicographical and Lexicological Projects

The paper reports on a dictionary of German loanwords in the languages of the South Pacific that is compiled at the Institut für Deutsche Sprache in Mannheim. The loanwords described in this dictionary mainly result from language contact between 1884 and 1914, when the German empire was in possession of large areas of the South Pacific where overall more than 700 indigenous languages were spoken.

The dictionary is designed as an electronic XML-based resource from which an internet dictionary and a printed dictionary can be derived. Its printed version is intended as an 'inverted loanword dictionary', that is, a dictionary that – in contrast to the usual praxis in loanword lexicography – lemmatizes the words of a source language that have been borrowed by other languages. Each of the loanwords will be described with respect to its form and meaning and the contact situation in which it was borrowed. Among the outer texts of the dictionary are (i) a list of all sources with bibliographic and archival information, (ii) a commentary on each source, (iii) a short history of the language contact with German for each target language, and perhaps (iv) facsimiles of source texts.

The dictionary is supposed to (i) help to reconstruct the history of language contact of the source language, (ii) provide evidence for the cultural contact between the populations speaking the source and the target languages, (iii) enable linguistic theories about the systematic changes of the semantic, morphosyntactic, or phonological lexical properties of the source language when its words are borrowed into genetically and typologically different languages, and (iv) establish a thoroughly described case for testing typological theories of borrowing.

> The development of scholarly lexicography of the Estonian Language as a Second Language in an historical and a theoretical perspective

KALLAS, JELENA

3 – Reports on Lexicographical and Lexicological Projects

This paper aims to provide an overview of the development of scholarly lexicography of the Estonian language as a second language in an

historical and a theoretical perspective. The paper describes what kind of information is presented traditionally in dictionary entries on the level of morphology, derivation, syntagmatic relationships and paradigmatic relationships. In addition, taking into consideration theoretical and practical viewpoints of modern lexicography on what kind of information should be presented in a dictionary entry so that the dictionary could be classified as a production dictionary (Apresjan (ed.) 2006; Atkins & Rundell 2008; Bo Svensén 2009; Novikov 2001; Siepmann 2006), the author is going to illustrate what kind of information should be added into the entries of a learners' dictionary of the Estonian language as a second language so that they could be used as production dictionaries.
In an historical perspective the analysis of the learners' dictionaries, which were published during the last 160 years, indicated that dictionary compilers provide dictionary users mostly with information about inflectional formation; meanwhile, the information about word formation (derivatives, compounds), syntagmatic and paradigmatic relationships is almost neglected. On the other hand, learners' dictionaries meant for speakers of Estonian as a first language provide much more information: the information about inflectional formation, word formation, synonyms, antonyms, paronyms is presented explicitly. The information about syntagmatic relationships is presented mostly implicitly by means of examples at the level of phrases, clauses and sentences.
The author puts forward detailed proposals for what kind of formal (inflectional formation, derivatives, compounds), semantic (mostly content-paradigmatic information) and syntagmatic (syntactic valency, collocations, idioms) characteristics should be given in a dictionary of the Estonian language as a second language and demonstrates practical implementations of explicit systematic description of syntactic valency and collocations of different parts of speech (nouns, adjectives, adverbs, verbs, quantifiers).

> WikiProverbs – Online Encyclopedia of Proverbs

KATS, PAVEL

3 – Reports on Lexicographical and Lexicological Projects

The WikiProverbs project was envisioned as a free online multilingual dictionary of proverbs, edited by the community. The idea behind the project was to address the difficulty of translating proverbs across the languages and to create a public repository of multilingual equivalents of proverbs that will serve language professionals, such as: writers, translators, journalists, as well as language enthusiasts. From its inception the project was conceived as a non-profit humanitarian enterprise for the sake of Internet users.

> Stichwort, Stichwortliste und Eigennamen in elexiko: Einflüsse der Korpusbasiertheit und Hypermedialität auf die lexikografische Konzeption

KLOSA, ANNETTE; SCHNÖRCH, ULRICH AND SCHOOLAERT, SABINE

3 – Reports on Lexicographical and Lexicological Projects

Die Überschrift des Beitrags impliziert dessen Gliederung in zwei größere thematische Abschnitte: der erste, allgemeine widmet sich Überlegungen zu Stichwort und Stichwortliste, der zweite, speziellere erörtert die Behandlung von Eigennamen in *elexiko*.
elexiko (www.elexiko.de) ist ein am Institut für Deutsche Sprache in Mannheim entstehendes Online-Wörterbuch zur deutschen Sprache. Die methodische Basis für die redaktionelle, lexikografische Erarbeitung von Wortartikeln ist das Prinzip der Korpusbasiertheit. Voraussetzung für dessen methodische Umsetzung ist, dass für jedes Stichwort (und seine Lesarten) Belege in ausreichender Anzahl und Qualität im *elexiko*-Korpus vorhanden sind. Um das zu gewährleisten wurde auch die Stichwortliste komplett neu erstellt, und zwar auf der Basis von Korpora des geschriebenen Deutsch seit 1946. Im ersten Teil des Beitrags werden grundsätzliche Gedanken zur Erarbeitung einer adäquaten Stichwortkonzeption im Rahmen eines Online-Wörterbuches dargelegt, Sonderfälle und Ausnahmen vorgestellt sowie die Vorgehensweise bei der korpusbasierten Erstellung der *elexiko*-Stichwortliste skizziert.
Ausgehend von der Definition von Eigennamen erörtert der zweite Teil des Beitrags die gängige lexikografische Behandlung von Eigennamen in allgemeinsprachigen Wörterbüchern und stellt Überlegungen dazu an, wie Eigennamen in Abgrenzung zu Gattungsbezeichungen lemmatisiert werden sollten. Dabei stellt sich für ein Online-Wörterbuch wie *elexiko*, dessen Schwerpunkt der lexikografischen Beschreibung auf der Bedeutung und Verwendung von Stichwörtern liegt, die Frage, in welcher Form die lexikografische Behandlung von Eigennamen erfolgen soll. Außerdem thematisiert dieser Beitrag die Behandlung von Eigennamen in *elexiko* hinsichtlich ihrer Erfassung, Klassifizierung und Darstellung und erläutert unterschiedliche Angabetypen. Ein Ausblick auf Suchoptionen zu den Eigennamen schließt die Überlegungen ab.

> Orthographical Dictionaries: How Much Can You Expect? The Danish Spelling Dictionary Revis(it)ed

LORENTZEN, HENRIK

3 – Reports on Lexicographical and Lexicological Projects

Orthographical dictionaries constitute a particular and rather specialised subclass of dictionaries. This contribution offers a presentation of the ongoing revision of a spelling dictionary (for Danish) and a discussion of some of the general and specific issues that have arisen during the project. Firstly, the historical background is described, a brief overview of the many editorial changes is provided, and lemma selection, variant forms and definitions are discussed in some detail. Particular interest is paid to the number and character of the included headwords, to the problems of (too many) variant forms and to the difficulties involved in providing definitions in a dictionary whose main purpose is to inform about correct spelling. Secondly, the field of official and unofficial orthographical dictionaries in Denmark is compared to that of some other countries of northern Europe: Sweden and Germany, and it is shown how the forthcoming edition of the Danish spelling dictionary is inspired by the other dictionaries. Finally, the conclusion engages in a discussion of the necessity of this particular type of dictionary, which to this author seems somewhat questionable.

> A language on the back foot: The Afrikaans lexicographer's dilemma

LUTHER , JANA

3 – Reports on Lexicographical and Lexicological Projects

Afrikaans originated in the variants of Dutch that developed at the southern tip of Africa during the 17th and 18th centuries. In the 19th century, when English began to overtake Dutch as the high-function language in the Cape, proponents of Dutch and Afrikaans put up a resistance, and during the 20th century the functions of Afrikaans expanded until it could take its place alongside Dutch and later stand with equal status next to English. As an official language Afrikaans reached back to Dutch a second time to develop into a full-fledged language. But its heyday could not last indefinitely. In recent decades the milieu of Afrikaans speakers has changed radically. Political upheaval, technological advances, new areas of specialisation, the lightning pace of new developments have thrust Afrikaansers into the thick of the world-wide explosion of knowledge which demands efficient communication. A third reversion to Dutch is out of the question. The path between Afrikaans and Dutch has become overgrown; few present-day users of Afrikaans still walk along it. Likewise, to the average Dutch man and woman, Afrikaans today is a distant language. In the multilingual South Africa, where English dominates, the effect of the contact with English on Afrikaans is undeniable. A serious threat to Afrikaans is its loss of status in the judiciary, the administration, education and as a scientific language. Against this backdrop the *Handwoordeboek van die Afrikaanse Taal* (*HAT*) – a household name among Afrikaans speakers, comparable to the Dutch

‘Dikke van Dale’ – is subjected to scrutiny: After its ‘golden age’, how well has the *HAT* kept pace with Standard Afrikaans in transition? Can it keep in step with the unstoppable, irreversible changes of the time and in the language today? Or will Afrikaans’s flagship dictionary, in a decade or so, lose its relevance for the Afrikaans user?

> Phonetic Transcriptions for the New Dictionary of Italian Anglicisms

MAIRANO, PAOLO

3 – Reports on Lexicographical and Lexicological Projects

This paper describes the work that has been done concerning the phonetic transcriptions for the *New Dictionary of Italian Anglicisms* directed by Prof. Pulcini (University of Turin): the dictionary contains both transcriptions of how Italians pronounce anglicisms and of how the corresponding English words are pronounced by native speakers of English. We shall explain how different pronunciation variants were selected for inclusion in the dictionary and how the transcriptions of anglicisms had to be adapted to the phonology and phonetics of Italian. A discussion will follow about the effects caused by the juxtaposition of English and Italian transcriptions. In fact, because of the intereference of the two phonetic and phonological systems, traditional conventions were in some cases abandoned in favour of more accurate phonetic transcriptions: this has been done with the aim of illustrating the most remarkable differences between the pronunciation of the words by Italian and English speakers.

> Centre for Bilingual Lexicography at Tbilisi State University, Georgia. Projects, Methods, History

MARGALITADZE, TINATIN

3 – Reports on Lexicographical and Lexicological Projects

The first bilingual dictionary of the Georgian language, Georgian-Italian was compiled and published in 1629 in Rome. Between 1629 and 1870 approximately ten European-Georgian dictionaries were compiled.

At the beginning of the 19th century Georgia became a part of the Russian Empire. Since that time the major emphasis has been placed on Russian-Georgian lexicography. As a result of such an approach, bilingual lexicography of the Georgian language suffered in respect to European languages.

Even when the first English-Georgian, or other European-Georgian dictionaries appeared from the 1940s, they were mere translations of European-Russian dictionaries, which led to numerous inaccuracies and even gross mistakes.

The same erroneous lexicographical principles became the basis of compilation of A Comprehensive English-Georgian Dictionary (CEGD), initiated by the Department of English Philology of Tbilisi State University back in the 1960s. The decision was made to translate the New English-Russian Dictionary, edited by I. Galperin.

After examining the existing material, the Editorial Staff of CEGD (established in the 1980s) arrived at the conclusion that it was impossible to edit the material in the form in which it was executed. The Editorial Staff developed entirely new principles for the creation of CEGD.

The method of the analysis of definitions of English Dictionaries was identified as the basic technique for the investigation of the semantic structure of English lexical units.

The process of revision and editing of the material of CEGD has continued for 25 years.

The publication of CEGD started in 1995 in fascicles on a letter-by-letter basis. By now, thirteen fascicles of CEGD have been published, from letters A to O.

The Internet version of CEGD was launched in February 2010.

Other projects of the Lexicographic Centre include: 'English-Georgian Learner's Dictionary';

'English-Georgian Military Dictionary' (first publication of the series of specialised English-Georgian Dictionaries).

> Crossing borders in lexicography: How to treat lexical variance between countries that use the same language

PARQUI, JAAP; BOON, TON DEN AND HENDRICKX, RUUD

3 – Reports on Lexicographical and Lexicological Projects

In the past decades the identity of Belgian Dutch has changed considerably. It no longer tries to copy Netherlands Dutch, but is following its own course. This development should be reflected in Dutch dictionaries. For different dictionaries (e.g. bilingual and explanatory dictionaries) and for different categories of words (e.g. juridical or informal words) different strategies should be adopted.

> Better Nicely Linked than Poorly Copied. Historical and Regional Dictionaries of Dutch Digitally United

SCHOONHEIM, TANNEKE AND DE TIER, VERONIQUE

3 – Reports on Lexicographical and Lexicological Projects

The *Woordenboek der Nederlandsche Taal* (Dictionary of the Dutch Language, WNT) has been freely available online since January 2007 (http://wnt.inl.nl). Compared to the original (printed) dictionary, the search facilities

have been considerably expanded. For instance, you can now search for a headword using modern spelling, and submeanings and citations can be displayed or omitted on demand.
Another innovation is that headwords in the WNT are now linked to external information, for example, to language maps, pictures and etymological information. At the moment, we add links to the available dialect material, starting with the large dictionaries of the dialects of Flanders, Brabant and Limburg. In this contribution we describe how this is done.

> Dutch Lexicography in Progress: the *Algemeen Nederlands Woordenboek* (ANW)

SCHOONHEIM, TANNEKE AND TEMPELAARS, ROB

3 – Reports on Lexicographical and Lexicological Projects

The *Algemeen Nederlands Woordenboek* (ANW – Dictionary of Contemporary Dutch) is a project of the Institute for Dutch Lexicology in Leiden, the Netherlands. It is an online corpus-based, scholarly dictionary of contemporary standard Dutch in the Netherlands and in Flanders, the Dutch speaking part of Belgium. It describes the Dutch vocabulary from 1970 onwards.
The ANW is aimed at a large audience, ranging from professional linguists to students and puzzlers. It provides information on form, content and use of words belonging to the general vocabulary of Dutch. It has an elaborate structure which aims to simplify the retrievability of words and meanings for the user compared to existing digital dictionaries. The semagram plays an important role in this, but so do various other innovative elements in the structure.

> Wurdboek fan de Fryske taal/Dictionary of the Frisian Language online: new possibilities, new opportunities

SIJENS, HINDRIK AND DEPUYDT, KATRIEN

3 – Reports on Lexicographical and Lexicological Projects

The Wurdboek fan de Fryske Taal (Dictionary of the Frisian Language, WFT) describes the vocabulary of the Modern West Frisian language and consist of 25 volumes of 400 pages each. The dictionary contains more than 100,000 entries. This paper is intended to show that an electronic version of the WFT, once the data have been converted to state-of-the-art standards and made available to the public by means of an advanced retrieval application, will be a modern lexicographical resource of significant value.

Integrating the WFT into the dictionary component of the *Geïntegreerde Taalbank Nederlands* (Integrated Language Database of Dutch, GTB) of the *Instituut voor Nederlandse Lexicologie* is the obvious means to reach this goal. In order to create more ways of searching the dictionary entries, data accessibility has to be enhanced by explicit tagging of information categories which can be exploited by a retrieval application.

The process of implementing the online version of WFT took place in several stages: First the existing database had to be repaired and optimised. Mistakes and inconsistencies had to be repaired. The logical structure had to be parsed and tagged with XML mark-up. Furthermore the newly created XML database had to be enriched with TEI encoding. And, finally the dictionary was incorporated into the GTB application.

The WFT has been incorporated into the online dictionary application of the Dutch language bank, and so is freely available to a large audience, allowing interested parties to search in one of the most complete Frisian dictionaries, and to explore the Frisian language in relation to Dutch.

> The principles and structure of the Estonian Etymological Dictionary

SOOSAAR, SVEN-ERIK

3 – Reports on Lexicographical and Lexicological Projects

The Estonian Etymological Dictionary (EED) has been a project of the Institute of the Estonian Language (IEL) since 2003. Due to the urgent necessity for an etymological dictionary it was decided to start from a short and not too detailed version tailored for the general public with no philological background and to broaden this version later in order to compile a scientific dictionary. The next step involved concrete decisions about the material to be included into the first version of the dictionary.

> Economicus: A New Conception of the Bilingual Business Dictionary

STORCHEVOY, MAXIM A.

3 – Reports on Lexicographical and Lexicological Projects

The paper is devoted to the *Economicus* project – English-Russian Dictionaries in Economic, Management and Finance – which is built on the new conception of bilingual business dictionary with rich, reliable and user-friendly lexicographical information for ordinary users (students, managers, translators etc.) as well as for researchers of language. The Economicus uses a rather sophisticated and advanced concept of entry with multiple zones which relies heavily on advantages of electronic entry demonstration and especially on the possibility of hiding and showing zones and subzones at user discretion. The latter feature creates enormous opportunities for the lexicographer to develop a rich but still user-friendly content of the entry.

The project is based on the alliance of economists and linguists. The linguistic expertise for the project was provided by the ABBYY software company who gave Economicus lexicographers access to a database specially designed for building dictionaries, its proprietary markup language and its linguistic corpus. ABBYY linguists took active part in developing the conception of Economicus entry and helping economic lexicographers to find a correct and effective approach to developing an up-to-date terminological dictionary.

The economic and business expertise for the project was provided by several dozens of professors of various educational institutions in Russia and abroad. The most important role was played by professors of Graduate School of Management (GSOM), St-Petersburg State University who took active part in evaluating and improving entries in corporate finance, management, marketing, international business and other business fields. In 2007 Graduate School of Management established the Translation and Lexicography Department where Economicus project has been developed since that time.

Economicus dictionaries are distributed with ABBYY Lingvo (as part of its basic dictionary collection and as additional downloads) and are accessible through a web-site http://dictionary.economicus.ru. The number of entries in Economicus is currently about 75 000.

> The ANW: an online Dutch Dictionary

TIBERIUS, CAROLE AND NIESTADT, JAN

3 – Reports on Lexicographical and Lexicological Projects

The *Algemeen Nederlands Woordenboek* (ANW) is a comprehensive online scholarly dictionary of contemporary standard Dutch in the Netherlands and in Flanders, the Dutch speaking part of Belgium (Moerdijk 2004, 2008; Moerdijk, Tiberius & Niestadt 2008). The dictionary focuses on written Dutch and covers the period from 1970 onwards. The dictionary was conceived as an online dictionary right from the outset and offers a range of search possibilities supporting both semasiological and onomasiological queries. A demo version of the dictionary[1] was launched at the end of 2009 (http://anw.inl.nl). This paper discusses the search application of the ANW dictionary. It focuses on the access strategies that are offered and on FunQy, the query language that was specifically developed for the project to facilitate implementation and future extensions to the search options offered by the ANW. Currently the demo version of the dictionary has just over 2000 registered users.

> Pilot project: A Dictionary of the Dutch Dialects

VAN KEYMEULEN, JACQUES AND DE TIER, VERONIQUE

3 – Reports on Lexicographical and Lexicological Projects

The lexicon of the traditional dialects in the Dutch language area is disappearing at a rapid pace. Three major regional dialect dictionaries, the *Dictionary of the Brabantic dialects* (WBD), the *Dictionary for the Limburgian Dialects* (WLD) and the *Dictionary for the Flemish Dialects* (WVD) inventory the vocabulary of the southern Dutch dialects. They are thematically arranged following the lexicographic idea's of A. Weijnen, which also are at the basis of still other dictionaries for some eastern dialect groups in the Netherlands. Because of their onomasiological arrangement, however, the dictionaries of Weijnen's school cannot render detailed semantic information. Therefore, professional lexicography has to call in the help of 'amateur' lexicography, i.e. the huge amount of alphabetical regional and local dialect dictionaries, made by non-professional lexicographers. In this paper a pilot project is presented, which aims at the creation of a lexicographical database for the alphabetical amateur lexicography, including both the old alphabetical tradition of the end of the 19th / beginning of the 20th century and the new tradition, rooted in the so-called dialect renaissance of the 70s and afterwards. It is defended that such a database – if enriched with the dutchifications of the dialect headwords – will prove to be an indispensable tool for lexicological research with regard to the history of the Dutch lexicon.

> Towards the completion of the Dictionary of the Flemish Dialects

VAN KEYMEULEN, JACQUES AND DE TIER, VERONIQUE

3 – Reports on Lexicographical and Lexicological Projects

The Dictionary of the Flemish Dialects is a major regional dialect dictionary for the Flemish dialect area, i.e. the provinces of West and East Flanders (Belgium), Zealand Flanders (the Netherlands) and French Flanders (France). The project began in 1972 at Ghent University (Belgium). It is a thematically arranged dictionary, set up along the lines proposed by A. Weijnen for the Dictionary of the Brabantic Dialects (1960-2005) and the Dictionary of the Limburg Dialects (1960-2008), its two sister projects. It combines a dictionary with a word atlas. This paper describes the state of affairs of the Flemish Dictionary with regard to data collection, data processing, presentation and publication. (The specialised software program used for the dictionary is presented in a separate paper, in which much attention is paid to the cartographic tools).

> Österreichische Pflanzennamen. Eine Webapplikation für ein thematisches Korpus

WANDL-VOGT, EVELINE AND PIRINGER, BARBARA

3 – Reports on Lexicographical and Lexicological Projects

Im Institut für Österreichische Dialekt- und Namenlexika (DINAMLEX) befindet sich eine onomasiologisch angelegte Pflanzennamensammlung, die geschätzte 31.000 mundartliche Pflanzennamen für geschätzte 2.000 botanisch-wissenschaftliche Stichwörter enthält. Neben den wissenschaftlich-botanischen Namen wurden überregionale deutsche Standardbezeichnungen und mundartliche Pflanzennamen gesammelt. In den Jahren 2000-2005 wurden sie im System TUSTEP digitalisiert. In den Jahren 2007-2010 wurde ebd. das System dbo@ema entwickelt. Es besteht aus der eigentlichen Datenbank, die zur Speicherung heterogener Dialektdaten geeignet ist, einer öffentlichen Website, einer Desktopanwendung zur Dateneingabe und eine Javascript Applikation zur Visualisierung geografischer Daten.

Seit 2008 werden die digitalen Pflanzennamen wissenschaftlich überarbeitet. Über die Website erfolgt der Zugriff auf die Datenbank und verschiedene Contentbereiche des Systems, z.B. Lemmata, Belege, Bibliographie, Personen, Multimedia. Eine interaktive Karte, wie sie beispielsweise von Google Maps bekannt ist, stellt eine lokationsspezifische Navigationsmöglichkeit dar. Über ein Popup können die raumbezogenen Daten auch über die Karte abgefragt werden und kommt der Benutzer wieder zu unterschiedlichen Contentbereichen der Datenbank. Damit werden Fragen wie '*Welche Belege aus dem Ort XY gibt es in der DBÖ / in dbo@ema*?' oder '*Wo sagt man Gelbling zum Pfifferling*?' per Mausklick beantwortbar. Mitte des Jahres 2010 soll eine Pilotversion des Systems mit den Bezeichnungen für österreichische Pilze unter wboe.oeaw.ac.at online gestellt werden.

Durch die Einbindung wissenschaftlich-botanischer Pilznamen und die Geocodierung der Belege wird die Vernetzung mit anderen Datenbanken sichergestellt. Folgende Verknüpfungsmöglichkeiten sind projektiert und demonstrieren beispielhaft den durch Standardisierung und Geocodierung zu erreichenden Mehrwert: Verlinkung mit der Online Flora von Österreich (http://62.116.122.153/flora/Hauptseite [Access date: 14 April 2010]) und der Datenbank der Pilze Österreichs (http://austria.mykodata.net/ [Access date: 14 April 2010]).

> The Dictionary of Lithuanian (LKŽ) and its Future in Databases and Electronic Versions

ZABARSKAITĖ, ELENA JOLANTA AND NAKTINIENĖ, GERTRŪDA

3 – Reports on Lexicographical and Lexicological Projects

The paper deals with the Dictionary of the Lithuanian Language (Vol. 1-20, 1941–2002): electronic release, 2005 (renewed version 2008) and its new version on the CD. A three-level lexical database: exhaustive for academic purposes, medium for the broad public, and more narrow for schools, is being created at the Institute of the Lithuanian Language. Its core consists of an electronic version of the Dictionary of the Lithuanian Language (about 0.5 million dictionary entries) and its card index (about 5 million cards), which is in the process of being computerized.

> The organization of entries in Spanish-English/English-Spanish bilingual dictionaries

DECESARIS, JANET

4 – Bilingual Lexicography

This paper discusses the organization of equivalents and presentation of fixed expressions in six bilingual dictionaries of Spanish and English. The dictionaries studied were published over the last forty years (1971, 1983, 2003, 2004, and 2008), and we compare the information contained in the older dictionaries with more recent ones. In addition, we compare frequency data taken from the Corpus del Español with information on fixed expressions contained in the dictionary entries. The focus of the study is on the representation of the Spanish words *cuadro* and *hoyo*, and the English word *poison.* The discussion herein would be of benefit to those planning a new bilingual dictionary or major overhaul of an existing one.

> Lexin – a report from a recycling lexicographic project in the North

HULT, ANN-KRISTIN; MALMGREN, SVEN-GÖRAN AND SKÖLDBERG, EMMA

4 – Bilingual Lexicography

In the late 70s, the Swedish Board of Education initiated a project (the *Lexin* project) aiming at production of dictionaries between Swedish and many immigrant languages. A monolingual Swedish dictionary was compiled, serving as the common base of the bilingual dictionaries. In the 90s, the project was exported to other Nordic countries. Since the Nordic languages are closely related, much of the work carried out in Sweden could be reused in Norway, Denmark, and Iceland. Today, there are many learners' dictionaries between Nordic languages and 'exotic' immigrant languages, especially with Swedish and Norwegian as source languages.

In this paper, we account for some aspects of this – in some respects probably unique – project. At the end, we give a description of the revision and updating of the Swedish database that has been going on since 2008

> Word-formation in English-French bilingual dictionaries: the contribution of bilingual corpora

LEFER, MARIE-AUDE

4 – Bilingual Lexicography

Research on the representation of word-formation in dictionaries is scarce and tends to be restricted to learners' dictionaries and monolingual dictionaries intended for native speakers. Nor is the issue of word-formation in bilingual dictionaries often discussed in lexicographic studies. This study, intended as a step on the way to rectifying the situation, reports the results of a comparison of the strategies adopted in four influential English-French dictionaries, focusing more particularly on derivational prefixes. The study shows that prefixes and word-initial elements in general receive very scant treatment in English-French dictionaries, which seems hardly justifiable when one thinks of the major role they play in the interpretation and translation of complex words. In my presentation I will highlight and illustrate a number of shortcomings, such as the lack of consistent criteria for the selection of affix entries and the misrepresentation of affix polysemy. More importantly, the presentation will also show how bilingual dictionary-making could benefit from bilingual corpora (both comparable and translation corpora) to improve the description of word-formation. I will propose a corpus-based list of the most productive and frequent prefixes in English and French. This list would seem to be a promising starting point for selecting more systematically and more rigorously the affixes to be included as headwords in bilingual dictionaries. To illustrate the usefulness of corpus data, I will also present a model bilingual entry for the French prefix *dé*– based chiefly on data extracted from an English-French translation corpus.

> Problems of Dialect Non-Inclusion in Tshivenḓa Bilingual Dictionary Entries

MAFELA, MUNZHEDZI JAMES

4 – Bilingual Lexicography

Language is human speech involving the use of words in an agreed way. However, a language is not absolutely homogeneous since there is variation. In any language one can expect to come across instances where certain speech differences may exist due to the influence of a language in an adjoining area. Tshivenḓa is characterised by a number of dialects, among them Tshiphani, Tshiilafuri, Tshimbedzi, Tshironga, Tshimaanḓa and Tshinia, which exhibit some linguistic features different from those of other groups. The standard dialect in Tshivenḓa is Tshiphani. This dialect is spoken in the areas of Tshivhase and Mphaphuli. The selection of the Tshiphani as a standard dialect in Tshivenḓa did not cause the other dialects to die out as they are still used by the Vhavenḓa as spoken language. However, there is non-inclusion of dialectal entries in some dictionaries, whereas in others, very few dialectal entries have been included. Some dialects differ from the standard dialect in vocabulary, whereas others differ from the standard dialect in pronunciation.

A lexicographer must always take into consideration that there is a variation in language. Lexicographers should not see the inclusion of non-standard dialects in a dictionary as corrupting the standard language. The inclusion of non-standard dialects in dictionaries, especially bilingual dictionaries, will assist dictionary users to know more about variants in the language. A dictionary is expected to accommodate all dialects of a language because they have equal value in spoken language. It is important for a lexicographer to first carry out research regarding the existence of dialects in a language if one intends to compile a dictionary. This paper seeks to show that it is necessary to include lexicons from non-standard dialects in lexicography works such as bilingual dictionaries because there is no dialect which is better than others. The addition of non-standard dialects in dictionaries will enrich the languages.

> Approche historique et sociolinguistique de la lexicographie bilingue missionnaire et les langues minoritaires en Algérie coloniale (1830-1930): le cas du berbère

MAHFOUD, MAHTOUT AND GAUDIN, FRANÇOIS

4 – Bilingual Lexicography

Notre propos prend place dans le cadre de l’histoire culturelle des dictionnaires. Nous nous proposons de mettre en lumière les circonstances qui expliquent et déterminent le développement de la lexicographie bilingue missionnaire dans l’Algérie colonisée. Nous traiterons plus particulièrement du cas du berbère.

La création, en 1868, de la Société des missionnaires d’Afrique en Algérie marque une nouvelle étape dans l’action missionnaire africaine. Depuis leur installation, les missionnaires ont œuvré pour faire sortir de l’anonymat la langue minoritaire du peuple berbère.

Le point de départ de notre étude relève d'un constat: même si les instructions, claires et rigoureuses, des supérieurs de la mission exigeant de leurs missionnaires une étude assidue et une connaissance approfondie de l'arabe, nous constatons que leur production lexicographique n'inclut aucun dictionnaire bilingue ayant pour objet la langue arabe. Or, toute la production lexicographique des missionnaires porte sur les différents dialectes berbères. Dès lors, cette orientation de la lexicographie-missionnaire ne manque pas de soulever des questionnements : a) Quelles étaient les instructions données aux missionnaires concernant l'étude des langues locales? b) Pourquoi les missionnaires se sont-ils penchés sur les différentes variétés berbères plutôt que sur l'arabe algérien alors que c'était la langue d'intercompréhension entre les communautés indigènes? La lexicographie bilingue-missionnaire a-t-elle participé à la valorisation des langues minoritaires berbères?
Le zèle missionnaire s'est centré sur la langue des berbères. Soumis aux choix de leur hiérarchie, influencés par le mythe berbère que cultive Lavigerie, confrontés aux nécessités de leur travail d'évangélisation, les missionnaires ont utilisé la langue berbère comme instrument pour répandre la bonne parole.
En composant des outils fondamentaux pour la vulgarisation de la langue berbère, les missionnaires ont contribué à la grammatisation de la langue berbère et au recueil d'un lexique devenu précieux pour les études sociolinguistiques.

> The TRANSVERB project – An electronic bilingual dictionary for translators: theoretical background and practical perspectives

SÁNCHEZ CÁRDENAS, BEATRIZ AND TODIRASCU, AMALIA

4 – Bilingual Lexicography

TRANSVERB is a lexicographic resource conceived for novice and professional translators who need assistance when translating texts into a foreign language. It is a semi-bilingual dictionary which can also be used for text production into a foreign language. The case study analyzed in this article pertains to the translation of verbs from French to Spanish. This dictionary is organized onomasiologically in terms of categories. Based on the hypothesis that human cognition organizes concepts in semantic categories (Tranel et al. 2001; Damasio et al. 2004), TRANSVERB is configured in lexical domains (Martín Mingorance 1985, 1987, 1900, 1995; Faber & Mairal 1999). The syntactic information in verb entries includes its combinatory potential, more specifically, its number of arguments as well as their semantic restrictions. This is established through corpus study.

> The retrieval of data for Slovene-X dictionaries

ŠORLI, MOJCA

4 – Bilingual Lexicography

The article reflects on the linguistic issues concerning the preparation of text for a new Slovene-English dictionary. Discussion is based on concrete examples from the reversed database of the Oxford-DZS Comprehensive English-Slovene Dictionary (2005/2006) and lexicogrammatical data from a corpus-based Slovene lexical database in the making. It is yet to be established how successful retrieval of data from a reversed bilingual database is. However, the first attempts to use information from the reversed database for the purposes of the compilation of a new Slovene-English dictionary indicate that the automatically generated database is a vast fund of information on the contrastive relations between English and Slovene, which should at no cost be overlooked. The user has instant access to the potential direct translation candidates, and to the more contextually-bound potential translations, many of which would have been inaccessible to a non-native speaker without an insight into what could be called the mirror image of the language. On saying that, it is important to stress that a reversed database as we understand it is in no way to be confounded with the actual 'reversed' dictionary itself, but merely to be seen as a bilingual framework in which no solution is automatically transferred to a Slovene-English dictionary. We come to the conclusion that while a corpus-based monolingual database is needed to provide a fresh and authentic image of the source language, it is also important to explore and exploit the data obtained in the reversed bilingual database because that will add an extra dimension to the Slovene-X dictionary text. The key question is how to proceed with the compilation of a new, bilingual, dictionary database, using both sources but avoiding a distorted lexical analysis of Slovene in use, while also ensuring a thorough contrastive analysis of the relationships between the two languages.

> OMBI bilingual lexical resources: Arabic-Dutch / Dutch-Arabic

TIBERIUS, CAROLE; AALSTEIN, ANNA AND HOOGLAND, JAN

4 – Bilingual Lexicography

In this paper we present the OMBI reversible bilingual lexical resources for Dutch-Arabic and Arabic-Dutch. These resources have been derived from a bilingual lexical database which has originally been produced with OMBI, a special tool for creating and editing bilingual dictionaries, within the framework of the project 'Woordenboek Nederlands-Arabisch,

Arabisch-Nederlands, Nijmegen' in the period of 1998 till 2002 at the Radboud University of Nijmegen. Printed dictionaries have been published on the basis of this database (Hoogland et al. 2003) and now the data has been converted to LMF (Maks et al. 2008) to ensure future interchangeability and interoperability.
OMBI-Arabic-Dutch and OMBI-Dutch-Arabic are part of a larger set of bilingual lexical resources which are available at the Dutch HLT Agency. The main strength of these bilingual computational resources is the high quality of the input data, which exceeds that of most existing computational resources, since it is based on the work of a team of professional lexicographers. In addition, most of these bilingual resources use the same Dutch component as a base, which offers interesting perspectives for linking the resources to each other following the hub and spoke model.

> Reversing a Bilingual Dictionary: a mixed blessing?

VELDI, ENN

4 – Bilingual Lexicography

The presentation focuses on the experience of reversing a general Estonian-English dictionary of about 49,000 entries and 93,000 equivalents by means of the Tshwanelex dictionary compilation software. The reversal served two purposes. First, it seemed appropriate to reuse the established cross-linguistic equivalents in the Estonian-English dictionary for the B part of a new English-Estonian dictionary. Second, one also expected to enlarge and improve the reversed Estonian-English dictionary in the course of the post-editing phase. So far the post-editing phase of the English-Estonian dictionary has been highly rewarding. In fact, it could be regarded as simultaneous cross-fertilization of both dictionaries, especially with regard to additional meanings and a more balanced treatment of synonyms. On the other hand, the post-editing phase of a general dictionary has been more time-consuming than expected. It is also argued that, on the one hand, the reversal mercilessly reveals the drawbacks of the B part of a bilingual dictionary, such as explanation-like equivalents, inaccurate equivalents, lexical poverty, etc. In fact, it appears that many dictionaries are not actually suitable for reversal. On the other hand, in the case of reversibly oriented dictionaries the post-reversal editing process may result in enriched target and source dictionaries – and will considerably reduce asymmetry in bilingual dictionaries.

> Management and use of terminological resources for distributed users in the translation hosting site Minna no Hon'yaku

ABEKAWA, TAKESHI; UTIYAMA, MASAO; SUMITA, EIICHIRO AND KAGEURA, KYO

5 – Lexicography for Specialised Languages – Terminology and Terminography

In this demonstration, we show the terminology management module of Minna no Hon'yaku (MNH: http://trans-aid.jp/), a translation hosting site with integrated translation-aid mechanisms, which was made publicly available in April 2009. As of February 8th, 2010, 1062 users have registered with MNH and more than 3400 documents have been translated, of which more than 1600 translations have been published on the site. On MNH, users can translate documents individually or can define groups and share the translation task. It provides users with functions such as lookup of high-quality dictionaries and terminologies, seamless access to Wikipedia and Google search, and reference to TM. There are two types of terminological resources on MNH, i.e. those provided by the system and those registered by users. The demonstration shows how terms are registered, shared and used.

> Adjectives and collocations in specialized texts: lexicographical implications

ALONSO CAMPOS, ARACELI AND TORNER CASTELLS, SERGI

5 – Lexicography for Specialised Languages – Terminology and Terminography

The *General Theory of Terminology* (Wüster 1979) states that all terms must be nouns, as the noun is the only category to designate a concept. For this main reason, adjectives and other grammatical categories are not considered as entries in most terminological dictionaries. The *Communicative Theory of Terminology* (Cabré 1999, 2000, 2002), on the other hand, has recently determined that predicative categories, such as adjectives, verbs and adverbs, can also become specialized lexical units (SLU). However, there are not enough empirical studies at this moment which confirm this hypothesis and examine the main characteristics of these predicative categories when they are used as terms. Specifically, our contribution studies the use of adjectives as terminological units in environmental texts. The study of Environment-related terminology is of special interest, as Environment is a new emerging domain with characteristics different from those of classical domains, such as Medicine or Chemistry. As it has been established in previous works (Alonso 2009, Bracho 2004), many Environment-related words are taken from the general language, but take on a terminological sense when they are used in environmental texts.

Our study focuses on adjectives which form a collocation $[N[A]_{SAdj}]_{SN}$, as this syntactic structure is frequently used in specialized discourse. Our main objective is to determine the 'terminological value' of these adjectives and their main characteristics. It is concluded from the data analysis results that the behaviour of adjectives depends mainly on the syntactic-semantic nature of the adjective. It is observed a general tendency to use as terms, either classifying relational adjectives (Bosque 1993, Bosque & Picallo 1996, Picallo 2002), or common qualifying adjectives that adopt a terminological sense in specialized texts. This fact brings about the need of different kinds of treatment for the representation of these adjectives in terminological dictionaries.

> 'Not Leaving Your Language Alone': Terminology Planning in Multilingual South Africa

BEUKES, ANNE-MARIE

5 – Lexicography for Specialised Languages – Terminology and Terminography

Status language planning has been one of the components of post-apartheid South Africa's transformation project that has managed to attract wide-spread attention. In 1994 South Africa moved from its former official bilingual language policy to a new constitution that enshrines official status to 11 of the languages spoken in South Africa. However, 16 years down the line there is widespread disappointment with organized language planning and management by government authorized agencies. The paper gives a brief analysis of terminology development in contemporary South Africa juxtaposed with a terminology development project at the micro level which, in Joshua Fishman's words, was initiated from the perspective of 'not leaving your language alone'.

The practice of translation is an age-old activity, but translation studies is a fairly 'new' academic discipline and hence its terminology is still in its infancy. Translation studies has been taught in South Africa at higher education institutions for more than thirty years, but mainly through the medium of English and Afrikaans. The prod for this project was therefore the identification of fresh needs for terminology development in this area to contribute to facilitating the sustained development of specialized discourses in higher education. Terminology development is viewed as indispensable for creating and sustaining a dynamic environment for the use of South Africa's official indigenous languages as a medium of instruction and ultimately for scientific progress.

> Extension of a Specialised Lexicon Using Specific Terminological Data

CARTONI, BRUNO AND ZWEIGENBAUM, PIERRE

5 – Lexicography for Specialised Languages – Terminology and Terminography

The paper describes methods for acquiring lexical information to implement a 'Unified Medical Lexicon for French' (UMLF) that aims at being a reference resource for NLP in the medical domain. We address four issues of lexical acquisition in a specialised domain. First, to assess the 'desired coverage' of lexical information, we use a large collection of French terms as a reference resource for the medical domain sublanguage. The collection contains close to 300,000 terms organised around conceptual identifiers. Second, by looking through this large amount of terminological data, we highlight the different kinds of information that might be useful to deal with typical terminological processing tasks, like variant recognition. The terminological variation phenomena that are very frequent in these terms are of three kinds: graphemic, inflectional and derivational variations. Third, we propose a model for organising the lexical information. Most of this model is inspired from existing specialist lexicons, but special emphasis is put on derivational morphological information. Finally, different kinds of acquisition methods are described, at the two levels of linguistic description that are addressed here: inflectional and derivational morphological knowledge. These methods allow acquiring an important amount of lexical data. For inflectional knowledge, the full paradigm is recorded, to provide information about all the possible inflected forms of lexical units within terms. Regarding derivational knowledge, specific derivation processes are targeted, in order to handle particular term variations. The relevance of the gathered derivational information is also assessed.

> Bilingual Technical-Translation Thesaurus as a Reliable Aid to Technical Communication

FAAL HAMEDANCHI, MARYAM

5 – Lexicography for Specialised Languages – Terminology and Terminography

The article reviews the problem of technical terminology translation and the role it plays in technical communication. Despite the progressing attempts for standardization of terminology, there is still long distance to a perfect terminology system practically in all language societies. For an individual concept are used different variants even within a single text and technical dictionaries often fail to cover all these variants. Languages do not possess the same instruments for illustrating a definite concept, as a result, in translating different equivalents of a single concept the translated terms

may be considered as synonyms rather than variants or, on the contrary, partial synonyms of a term in the source language can be considered as variants or close synonyms in the target one. The problem gets even more complicated when it comes to languages, namely Persian and Russian, where the users are imposed to employ English as an intermediate language.
Technical dictionaries pay less attention to these differences, at the best, they may provide scope notes or short definitions to distinguish different senses of a term, which hardly suffices for a proper communication. On the other hand, users of a bilingual technical dictionary may look up different kinds of information besides definition and equivalents. They may look up cross-language synonymous or antonymous, allocations, homonyms and other information, which are rarely provided by a bilingual technical dictionary.
These facts imply the necessity of employing more onomasiological approach in compiling bilingual technical dictionaries. In our opinion, a revised structure of information-retrieval thesauri complies in a better way with the requirements of technical dictionaries.
A technical-translation thesaurus can reveal the basic structure of an information retrieval thesaurus, but compiles the necessary features of a common language thesaurus and provide approaches to equivalents of a term in different languages starting from the concept, which does not depend on the language.

> Introducing the Dutch Terminology Service Centre: a centre of expertise on practical terminology work

GÖRÖG, ATILLA

5 – Lexicography for Specialised Languages – Terminology and Terminography

The Dutch Terminology Service Centre (DTSC, Steunpunt Nederlandstalige Terminologie) was founded in 2007 by the Dutch Language Union, a Dutch-Flemish government institution. The DTSC functions as a non-commercial information center for all aspects of terminology and serves the entire Dutch-speaking community. We give advice on terminological research to anyone who is involved in terminology-related work (companies, organizations, translators, terminologists, teachers, scientists etc).

> The Tension between Definition and Reality in Terminology

HACKEN, PIUS TEN

5 – Lexicography for Specialised Languages – Terminology and Terminography

Whereas natural language concepts are based on prototypes, classical terminological definitions (CTDs) are based on necessary and sufficient conditions. The sociocognitive approach to terminology rejects both the possibility and the desirability of CTDs. In this paper, I argue that CTDs are needed for certain types of term, but not for others.

In a first step, I distinguish two overlapping classes of expressions, based on two different criteria for termhood. Specialized vocabulary is the set of items whose use is restricted to specialized communication. Specialization is a gradual property, so that the boundaries of this class are vague. Terms in the narrow sense have a concept with a clearcut boundary. In scientific and legal contexts, clearcut boundaries are necessary because classification is important. Only for terms in this narrow sense do we need CTDs.

Some of the problems raised for CTDs can be solved straightforwardly by restricting their domain to terms. Discussions about the best definition, e.g. for governing category in Chomskyan linguistics, indicate a search for the best concept rather than a prototype nature. In cases such as significant in statistics, the CTD is logically independent of the prototype concept associated with the word significant in general language. In such cases, finding CTDs does not pose insuperable problems.

More difficult problems arise when scientific concepts interact with our intuition about classification and with technological advances. The discussion of species in biology, compound in linguistics, and planet in astronomy demonstrates how these problems arise and how they can be addressed without recourse to prototype-based concepts. Terms in the narrow sense may be unnatural and their definition not straightforward, but they are crucial in scientific communication and depend on CTDs.

> Termania – Free On-Line Dictionary Portal

Krek, Simon

5 – Lexicography for Specialised Languages – Terminology and Terminography

Termania, a free on-line dictionary portal with integrated dictionary browsing and editing tools is being developed by Amebis software company from Kamnik, Slovenia, in cooperation with Trojina, Institute for Applied Slovene Studies. It provides an interface for dictionary browsing and a simple but reasonably versatile on-line dictionary editing tool. The portal is intended for general public users with no specialized computer or lexicographic knowledge, but with an interest to share terminological or general language knowledge, either by offering translations in a bilingual or multilingual environment or providing definitions in a monolingual context.

The portal is intended to serve as the central terminology data and opinion exchange node for Slovene terminology. Therefore, a discussion forum will be included in the portal and a possibility to comment on and evaluate each particular entry. Internationalization of the portal is

foreseen by providing language-specific interface for all widely used languages. General policy for the use of dictionary data on and off the portal is determined by the Creative Commons Attribution Non-Commercial Share Alike licence but also other licences can be used, according to users' preference.
The dictionary portal will be available at the web address: http://www.termania.net/ from July 2010.

> TermFactory: A Platform for Collaborative Ontology-based Terminology Work

KUDASHEV, IGOR; KUDASHEVA, IRINA AND CARLSON, LAURI

5 – Lexicography for Specialised Languages – Terminology and Terminography

TermFactory is an array of standards and tools based on Semantic Web ontology techniques. Its mission is to allow companies, organizations and individual contributors to collaboratively produce multi-domain special language vocabularies and ontologies.
Ontologization of terminological data has several benefits, such as global identification of concepts, automatic checks for logical errors, reasoning and data propagation, presentation of data in machine readable and -processable form and the possibility to substitute static entries with dynamic 'views' tailored according to the user's needs and preferences.
Collaborative work is a double-edged sword which potentially has many benefits but may also present serious challenges. In our poster, we describe challenges of collaborative terminology work and possible solutions to them. If well-organized, a collaborative project can be quite successful, as the example of Wikipedia and many other collaborative projects on the Internet demonstrate.

> The Focal.ie National Terminology Database for Irish: software demonstration

MĚCHURA, MICHAL BOLESLAV AND Ó RAGHALLAIGH, BRIAN

5 – Lexicography for Specialised Languages – Terminology and Terminography

In this demonstration, we will showcase the National Terminology Database for Irish focal.ie which was launched on-line in 2006 and immediately became very popular, attracting hundreds of thousands of searches every month. The Web site allows users to search a database of around 300,000 terms. It was developed to make the stock of terms of the National Terminology Committee available online, and was designed to be easy to use.
In addition to the public-facing Web site, the software comprises an

editorial interface (password-protected Web site), a relational database, and a library of objects and functions that acts as an interface between the two Web sites and the database.

The public-facing Web site allows users to search the database using a 'Quick Search', a 'Complex Search', or an 'Alphabetical Listings' function. The 'Quick Search' function returns 'Similar terms', 'Exact matches', and 'Related matches' from the database. The editorial interface allows users to search and edit the data contained in the database.

While the primary requirement for the public-facing Web site is user-friendliness, the primary requirement for the database is the ability to record complex linguistic data in a logical structure. The structure adopted for the Irish lexical database is based on the conceptual model, widely considered a standard in the terminology industry (ISO 704). The database is multilingual and contains rich grammatical labelling, usage examples, definitions, as well as other information. The database, editorial tools, and public interface were developed by Fiontar in-house using Microsoft technologies. The database and Web sites are hosted by Information Systems & Services, Dublin City University.

A new version of the public site was recently launched and can be accessed at the following URL: http://www.focal.ie/

> Towards a bilingual lexicon of information technology multiword units

MOSZCZYŃSKI, RADOSŁAW

5 – Lexicography for Specialised Languages – Terminology and Terminography

The article presents a proposal of an electronic, English-Polish translation dictionary covering the language of computer science. The dictionary will focus on multiword units and phraseology typical for this domain. It is supposed to answer the needs of technical translators, who can easily access simple terminological databases, but lack good production dictionaries that would go beyond single terms. The proposed dictionary aims at filling this gap by focusing on multiword units and their modifications, as well as on individual terms' collocational patterns.

The dictionary will be based on the idea of 'extended phraseology' proposed by Müldner-Nieckowski. According to this idea, phraseology is not limited to idioms in the traditional sense of the word, but also covers phrasemes (i.e. units with conventionalized structure, but without figurative meaning), as well as phraseograms (syntactically incomplete units that carry some semantic value). Such a broad approach to phraseology in the planned dictionary will allow translators to create texts that sound natural to computer science experts and to maintain consistency on the stylistic level on top of terminological consistency.

The dictionary will be created in electronic form, with the aim to make it available free of charge on the Internet as part of the Freedict project.

> Building on a terminology resource – the Irish experience

NIC PHÁIDÍN, CAOILFHIONN; Ó CLEIRCÍN, GEARÓID AND BHREATHNACH, ÚNA

5 – Lexicography for Specialised Languages – Terminology and Terminography

www.focal.ie is the national database of Irish language terminology. In this paper, we examine: (i) the impact achieved by this resource in the five year period since work commenced; (ii) the possibilities which have arisen from one project over a short time span, to develop sub-projects and related initiatives; and (iii) the advantages and opportunities arising from the creation of one high-quality electronic language resource. The Irish case shows that the development of high-quality resources for a lesser-used language can have interesting and unexpected knock-on effects.

We present eight stages and aspects of term planning: preparation/planning; research; standardisation; dissemination; implantation; evaluation; modernisation/maintenance; and training. Fiontar, in its work, has moved from its initial involvement in the *dissemination* of terminology, to take an active part in other aspects of term planning for Irish: *research, standardisation, evaluation, modernisation* and *training*. This has been achieved through *editorial and technological development*, in *partnership* with key stakeholders and always from a *socioterminological* point of view – that is, with an emphasis on terminology as an aspect of language planning and from the point of view of users in particular.

Particular projects described include Focal as a term management system and as a user resource; tools for translators; user links to a corpus; the development of a new sports dictionary; and research into subject field headings. Two related projects are the LEX legal terms project for term extraction and standardisation, and the development of terminology for the European Union.

> L'ancien et le moyen français au siècle classique: le *Tresor de Recherches et Antiquitez Gauloises et Françoises* de Pierre Borel (1655)

AMATUZZI, ANTONELLA

6 – Historical and Scholarly Lexicography and Etymology

Pierre Borel est 'le premier des savants qui se mirent à rédiger des recueils où les mots de l'ancienne langue étaient consignés' (G. MATORÉ, *Histoire des dictionnaires français*, Paris, Larousse, 1968, p. 132). En effet son

Tresor de Recherches et Antiquitez Gauloises et Françoises est un ouvrage qu'on peut compter parmi les premiers dictionnaires d' 'ancien français' et qui doit être également signalé pour les préoccupations étymologiques qu'il affiche et parce qu'il recense beaucoup de régionalismes.

Dans le cadre de ce colloque je focalise mon attention sur la manière dans laquelle Borel affronte la problématique de l'évolution du français et en particulier sur la façon dont ce dictionnaire présente et envisage la langue ancienne.

Dans sa préface longue et articulée, qui contient un discours métalinguistique explicite, Borel nous livre des considérations à propos du processus de transformation qu'investit les langues. Il s'agit des réflexions d'un homme cultivé du Grand Siècle qui entend valoriser la 'langue ancienne' et qui insiste sur le fait que l'évolution linguistique est tout à fait naturelle. Il affirme une vision de la langue où diachronie et synchronie sont complémentaires et indissociables, ce qui se traduit dans son souci constant d'aborder les phénomènes linguistiques (le lexique mais aussi la structure de la phrase et l'évolution phonétique) dans une perspective historique et culturelle.

À une époque où le français subit une épuration sévère et les archaïsmes sont proscrits, Borel va contre courant. Il produit un dictionnaire de la langue et de la civilisation anciennes,. utile pour accéder à la culture gauloise et française du Moyen Âge et pour comprendre ses évolutions successives, qui constitue un témoignage précieux de la valeur culturelle des mots, de la richesse et de la vivacité d'expression tant de l'ancien que du moyen français. Mais le *Tresor*, qui atteste aussi du développement du vocabulaire jusqu'au français classique, devient également 'mémoire' de comment la langue et la civilisation se sont construites le long des siècles.

> La compilation de dictionnaires de synonymes distinctifs: une démarche synonymique et lexicographique

FERRARA, ALICE

6 – Historical and Scholarly Lexicography and Etymology

En 1718 naissait un nouveau genre de dictionnaire: le dictionnaire de synonymes monolingue français appelé *a posteriori* dictionnaire de synonymes distinctif (par opposition aux dictionnaires de synonymes cumulatifs que nous connaissons aujourd'hui.)

Les dictionnaires distinctifs se composent de définitions, précisions, exemples qui justifient les propos de l'auteur. Quand les dictionnaires de synonymes se sont multipliés, à peine un siècle après la naissance du genre, apparaît un genre annexe, celui de compilation. Le but des auteurs de compilations était de garder ce qu'ils jugeaient être le meilleur des dictionnaires de synonymes les précédant. Dans notre article, nous nous interrogerons sur la place des compilateurs et si l'on peut dire que ce sont eux-aussi des synonymistes. Nous montrerons que compiler les dictionnaires de synonymes c'est faire preuve d'autant de travail synonymique et lexicographique que de composer directement le dictionnaire, car pour compiler des dictionnaires de synonymes il faut faire des choix multiples. Tout d'abord il faut choisir les auteurs dont on souhaite compiler les dictionnaires. Ensuite il faut sélectionner les articles à reprendre, ce qui représente un véritable choix du compilateur par rapport aux termes qu'il estime être réellement synonymes. Et enfin, il faut étudier ce qu'un compilateur garde des articles préalablement choisis. En effet, il gardera d'un article ce avec quoi il est en accord, et inversement, retirera tout ce à quoi il n'adhère pas. De plus, outre tous ces choix, le compilateur se fait également entendre puisqu'il peut aussi composer lui-même un certain nombre d'articles. Nous pouvons donc dire que la compilation de dictionnaire de synonymes distinctif est un véritable travail de réécriture fondé sur l'écriture et la création puisque les compilateurs n'ont de cesse de faire des choix lexicographiques.

> Lexicography, Printing Technology, and the Spread of Renaissance Culture

HANKS, PATRICK

6 – Historical and Scholarly Lexicography and Etymology

Historians of lexicography in the English-speaking world have implied that Robert Cawdrey's *Table Alphabeticall* (1604) is the first English dictionary. Landau (1984, 2001) makes this claim, adding that it is 'the least inspiring of all seminal works'. In this paper, I agree that the *Table Alphabeticall* is uninspiring, but I deny that it is a seminal work. Landau overlooks the rich 16th-century tradition of Renaissance and Humanist lexicography in Europe, in particular the *Thesaurus Linguae Latinae* of Robert Estienne (1536) and the *Thesaurus Linguae Graecae* of his son Henri Estienne (1572). These seminal works are astonishing achievements—breathtaking innovations—in terms of both scholarship and technology. They set standards for subsequent European lexicography. Two technological innovations made these great dictionaries possible: the invention of printing by Gutenberg in Strasbourg in about 1440 and the typography of Nicolas Jenson in Venice in 1462. These technological developments and the lexicographical achievements that were made possible by them contributed, in the first place, to the Renaissance programme of preserving the classical heritage of ancient Greece and Rome and, in the second place, to the role of dictionaries in spreading Renaissance culture and Humanism across Europe. The paper goes on to briefly outline the emergence of bilingual lexicography, replacing the polyglot lexicography that was standard in the 16th century. A comparison is made between the influence of printing technology on 16th century lexicography and the potential influence of computer technology on 21st century lexicography.

> Grammatical information in dictionaries

HOEKSTRA, ERIC

6 – Historical and Scholarly Lexicography and Etymology

A dictionary is an encyclopaedia of linguistic information about words. It presents to a target group of laymen and professionals general information about words belonging to various disciplines of linguistics such as

– semantics (the meaning of words and phrases)
– phonology (the pronunciation of words)
– syntax (the syntactic category of words and the collocations in which they partake)

My contribution discusses the question whether (new) insights from these disciplines may change the content of dictionaries, seeing that an evaluation of these insights does not take place very often.

It is a shortcoming of dictionaries that a paraphrase of the meaning of function words is often not very insightful with respect to their use (Coffey 2006).

– What is the meaning of Dutch *er* 'there'?
– What is the meaning of articles like *the*?
– What is the meaning of the complementiser *that*?

Some Dutch dictionaries muddle the description of the various uses of *er*, ignoring the distinctions drawn by de *Algemene Nederlandse Spraakkuns*t, the Standard Dutch Grammar (Haeseryn et al. 1997). Those distinctions are practical and well-motivated (Hoekstra 2000). It is proposed to use syntactic knowledge to structure articles about function words. In addition, dictionaries can covertly use example sentences to illustrate syntactic phenomena. Such measures strengthen the encyclopaedic character of a dictionary.

> Celtic Words in English Dictionaries and Corpora

ITO, MITSUHIKO

6 – Historical and Scholarly Lexicography and Etymology

The researcher has collected Celtic words from several English dictionaries and a few English etymological dictionaries and found that there are about 300 words in present English dictionaries. He has studied what words and how many of the 300 words native speakers of English know. The research method was giving matching tests of words and definitions and having subjects write appropriate words to definitions. The subjects were all adult voluntaries. Main purposes of the present study are: (1) to survey what words of the 300 words appear in BNC and Wordbnaks, and (2) to survey if well known words by native speakers are highly frequent words in BNC and Wordbanks. Two main results have deduced from the present study. One is that not all of the 300 words appear in BNC and Wordbanks and some words appear in the two Corpora and some others appear in either of the Corpora and the others do not appear in both of them. The other is that well known words in the research do not necessarily come to the top frequent positions of the Corpora.

> **Annotations in *Dictionarium Latino Lusitanicum, ac Iaponicum* (1595) in the Context of Latin Education by the Jesuits in Japan**

KISHIMOTO, EMI

6 – Historical and Scholarly Lexicography and Etymology

The Jesuits in Japan began establishing schools in the 1580s to mentor young native men in priesthood. In 1594, their students received a printed abridged edition of the Latin grammar, originally written by Manuel Alvarez, and the next year they received *Dictionarium Latino Lusitanicum, ac Iaponicum* (DLLI), a Latin-Portuguese-Japanese dictionary based on the Latin dictionary compiled by Ambrogio Calepino.

One of the features, when comparing the DLLI with the original, is that it cites the names of Latin classical writers without quoting sentences in several entries. This paper attempts to clarify the reasons for these annotations in this edition and reflects on the purpose of the DLLI.

Plautus is cited in about 70 entries, the most citations among all the names found in the DLLI. However, this number does not reflect the number in the original, which includes many classical writers, especially Cicero, whose works were regarded as a model for Latin prose. We also have no evidence showing that Jesuits in Japan regarded Plautus's writing as more important than Cicero's in teaching Latin.

The editors of the DLLI cite Vergilius most frequently after Plautus; we also find many annotations from the original showing the differences in usages such as 'apud veteres' (used by ancient people) or 'apud poetas' (used by poets). Similarly, it is reasonable to suppose that the editors included notes on 'Plaut' to describe the differences in older usages. They appear to retain the citations of writers and other annotations on special usages in order to teach the various nuances of Latin vocabulary to students in Japan, many of whom had elementary or intermediate language skills and needed good Latin proficiency to work as priests.

> Vers un enrichissement raisonné de la rétroconversion du *Französisches Etymologisches Wörterbuch* (FEW)

MAZZIOTTA, NICOLAS AND RENDERS, PASCALE

6 – Historical and Scholarly Lexicography and Etymology

L'informatisation par rétroconversion du Französisches Etymologisches Wörterbuch (FEW) de Walther von Wartburg, oeuvre fondamentale de la lexicologie historique galloromane, est à présent en cours de réalisation et nous voudrions montrer en quoi elle est perfectible et comment elle pourrait être améliorée. Nous présentons brièvement le FEW et la nécessité de le rendre exploitable par l'ordinateur (attentes des utilisateurs et besoins de formalisation), puis nous exposons les principes qui ont gouverné sa rétroconversion (au format XML) et donnons l'exemple détaillé de l'article substantivus, dont nous faisons la lecture suivie. Ensuite, nous focalisons l'exposé sur la microstructure rétroconvertie des articles et l'enrichissement par annotation manuelle que nous préconisons. Nous montrons quelles informations ne sont pas accessibles à la machine en raison de l'omniprésence de conventions implicites dans l'oeuvre originale. Nous synthétiserons enfin le potentiel de notre approche: l'accès à l'intelligence du FEW, et non plus seulement à sa forme.

> The negation particle *ne* in the historical dictionaries of Dutch

MOOIJAART, MARIJKE

6 – Historical and Scholarly Lexicography and Etymology

In the historical stages of the West Germanic languages *ne* has been part of the negation system. In Old Dutch through Early New Dutch (c. 600 – 1600/1700) this particle had various specific functions and senses, depending on the sentence structure and on whether or not it co-occurred with other negation elements, such as negative indefinites and adverbs. The present paper focuses on the way in which this particle as function word is described in the four successive historical dictionaries of Dutch. As these dictionaries were compiled in different periods and on different editorial principles, one can expect differences in treatment of the particle. Sometimes shortage of material plays a role. The focus on translation in Modern Dutch rather than on a precise grammatical analysis, causes inadequate descriptions in other cases. Especially with respect to the conjunctional construction with *ne*, the lexicographer is in need of clear, insightful discussions of this complicated phenomenon in the linguistic literature as a basis to his description. In spite of these shortcomings, the dictionaries together contain a comprehensive survey

of the various uses of *ne*, in some of them with a detailed inventory of contexts, together with a large amount of mostly dated illustrative citations.

> Antedating headwords in the third edition of the OED: Findings and problems

PODHAJECKA, MIROSŁAWA

6 – Historical and Scholarly Lexicography and Etymology

The present paper describes problems involved in antedating headwords in the third edition of the Oxford English Dictionary (OED3). One of the elements in urgent need of revision in the previous edition was the dating of quotations but, despite the intensive labours of the lexicographers, some OED3 headwords and senses are still likely to be misdated. This paper is based on the premise that Google Books, a gigantic online resource, can be applied successfully for the verification of OED3 dating. Indeed, my findings indicate clearly that Google Books has a vast research potential, because half of the words in my sample (covering 129 items related to dancing) have been antedated in full-text sources. However, neither the search procedure nor the interpretation of data retrieved is straightforward, so a number of ambiguous and problematic examples have been provided to show that antedating is far more intricate than it seems at first sight. For example, one has to repeatedly distinguish between related senses, evaluate the relevance of similar word-forms and determine whether or not the words can be treated as fully-fledged loanwords. Even though many of my decisions were purely intuitive, I nonetheless hope that at least some of the antedatings found in Google Books turn out to be helpful for OED3 lexicographers in the on-going revision process.

> Le grand vocabulaire François, un ouvrage taxé de tous les maux

REY, CHRISTOPHE

6 – Historical and Scholarly Lexicography and Etymology

Notre communication propose une présentation du *Grand Vocabulaire François* (1767-1774), la première entreprise lexicographique du grand éditeur Charles-Joseph Panckoucke.

Nous esquissons ici les traits d'un ouvrage qui en dépit de son originalité scientifique et de ses choix linguistiques très intéressants est resté dans l'ombre de l'*Encyclopédie* et du *Dictionnaire Universel* de Trévoux.

Our paper provides an overview of the *Grand Vocabulaire François* (1767-1774), the first lexicographical work of the great publisher Charles-Joseph Panckoucke.

We present here a dictionary that despite its scientific originality and its very interesting linguistics choices remained in the shadow of the *Encyclopédie* and the *Dictionnaire Universel* of Trevoux.

> Frauen
Rollentypen in einem dialektlexikographischen Jahrhundertprojekt (1911-2010)

WANDL-VOGT, EVELINE

6 – Historical and Scholarly Lexicography and Etymology

Das Wörterbuch der bairischen Mundarten in Österreich (WBÖ) ist ein lexikographisches Großprojekt (Sammlung seit 1911-; Publikation seit 1963-; Digitalisierung der Datenbestände seit 1993-).

In der Geschichte des Wörterbuchunternehmens lässt sich feststellen, dass Frauen eine besondere Rolle eingenommen haben, die in Abhängigkeit von den Arbeitsbereichen und den zeitlichen Gegebenheiten gesehen werden muss.

Wichtige Rollentypen für das Wörterbuchprojekt sind:[1] Kanzlistin (DINAMLEX;1914-1942;nw);Gewährsperson(DINAMLEX;1913-lfd.;nw.|w.), Sammlerin (DINAMLEX; 1913-lfd.; nw.|w.); Praktikantin (DINAMLEX; 2009-lfd.; w.); Datenbankmitarbeiterin (DINAMLEX; 1993-lfd.; nw.|w.); Datenbankentwicklerin (DINAMLEX; 1993-lfd.; w.); Datenbankleiterin (DINAMLEX; 1993-lfd.; w.); Artikelverfasserin (DINAMLEX; 1963-lfd.; w.); Technische Redaktion (DINAMLEX; 1998-lfd.; w.); Redakteurin WBÖ-Online Edition (DINAMLEX; 2004-lfd.; w.); WBÖ-Gesamtredaktion (DINAMLEX; 1969-lfd.; w.); Stellvertretende Direktorin (2000-2005; w.); Direktorin(DINAMLEX;1998-lfd.,w.);Mitglied(Kuratorium;1994-2001;w.); Obfrau (SBT; 2006-lfd.; w.); Beiratsmitglied (SBT; 2006-lfd.; w.); Sprecherin des Beirats (SBT; 2006-lfd.; w.); Aktuarin der philosophisch-historischen Klasse (ÖAW; 1946-lfd.; w.); Mitglied der Gelehrtengesellsschaft (ÖAW; 1948-lfd.; w.); Vorsitzende der philosophisch-historischen Klasse (ÖAW; 2009-lfd.; w.); Vizepräsidentin (ÖAW; 2009-lfd.; w.).

Frauen spielten im Grundlagenbereich (Sammlung) u.a. auch zeitbedingt (frühes 20. Jahrhundert) eine untergeordnete Rolle. Nur 6% der ersten Sammlungen (und rund 90.000 Belege) gehen – wenn auch über das gesamte Bearbeitungsgebiet des WBÖ verteilt – auf Frauen zurück. Im Bereich der Archivierung und Sichtung des Materials spielten Frauen eine zentrale Rolle. Ihre aktive Mitarbeit im Schatten des Großprojekts hat von der Materialordnung und Materialerweiterung (Exzerption) bis hin zur Digitalisierung eine entscheidende qualitative und quantitative Rolle für das heutige Korpus. Die Digitalisierung steht von Beginn an bis heute unter weiblicher Konzeption und maßgeblicher Mitarbeit von

Frauen. Die Wörterbuchartikel werden seit den 70-er Jahren verstärkt von weiblichen Mitarbeiterinnen verfasst. Das derzeitige WBÖ-Team steht unter weiblicher Teamleitung. Bis heute spielen Frauen eine tragende Rolle und prägen das Bild des Projekts und Unternehmens entscheidend mit. Die Einträge in der Zeitspalte verdeutlichen auf anschauliche Weise, wie es gerade in den letzten Jahren vermehrt gelungen ist, Frauen als Entscheidungsträgerinnen zu etablieren.

[1] Die für das Wörterbuchprojekt quantitativ und qualitativ wichtigen Rollentypen sind verzeichnet. In der Klammer angegeben sind: Organisatorischer Rahmen: Institut oder Akademieebene, Zeitraum, in der der Rollentyp wesentlich ist, wissenschaftliche (w) oder wissenschaftlich-unterstützende (wu) Tätigkeit. DINAMLEX = Institut für Österreichische Dialekt- und Namenlexika; SBT = Zentrum Sprachwissenschaften, Bild- und Tondokumentation; ÖAW = Österreichische Akademie der Wissenschaften.

> Dictionary, lexicon, glossary, wordbook or thesaurus? The usefulness of OALDCE7 and OLT for choosing the right word

DZIEMIANKO, ANNA

7 – Dictionary Use

Monolingual English learners' dictionaries (MLDs) published in recent years have many features which make them better suited to the needs of the target user group. Among others, onomasiology has slipped into their design. Today, MLDs typically list synonyms and antonyms, or even offer synonym notes, where words close in meaning are compared and contrasted. On the other hand, thesauri have also changed. The year 2008 witnessed the publication of the *Oxford Learner's Thesaurus: A Dictionary of Synonyms*, which goes beyond clustering words close in meaning. It defines each synonym, exemplifies its usage, and even juxtaposes selected synonyms in special notes.
The aim of the present study is to investigate the usefulness of the *Oxford Advanced Learner's Dictionary of Current English* (7th edition, OALDCE7) and the *Oxford Learner's Thesaurus* (OLT) for discriminating between synonyms. The paper is underpinned by empirical research, in which 73 advanced learners of English took part. In the experiment, words appropriate for given contexts had to be indicated in different synonym sets. The results reveal that neither dictionary significantly shortened the time needed to complete the task. Nonetheless, the use of OLT much more often resulted in successful synonym selection. Interestingly, synonym notes, present in both dictionaries, did not affect the subjects' choices. Besides, different information was usually referred to in the two dictionaries. In OALDCE7 the subjects paid attention most often to definitions, while in OLT – to

examples. The results of the supplementary questionnaire suggest that the students' familiarity with the two dictionary types could not have affected their performance. They were nonetheless more satisfied with their results when they had OLT at their disposal rather than OALDCE7. Yet, they were critical of the arrangement of synonyms in the OLT synonym clusters, where the alphabetical order, rather than frequency, would be a better solution.

> Donner un accès aisé aux formes phoniques des mots décrits dans un dictionnaire: étude pour un dictionnaire monolingue français destiné à de jeunes utilisateurs

GASIGLIA, NATHALIE

7 – Dictionary Use

Dans le cadre de cette contribution, je me propose de réfléchir à ce qui pourrait évoluer dans les dictionnaires sur support électronique concernant les descriptions des formes phoniques des unités linguistiques décrites et les modes d'accès à celles-ci. En envisageant les consultations de dictionnaires à la fois dans le cadre d'une aide à la compréhension (de ce qui est entendu ou difficile à déchiffrer) et à l'expression (énonciation ou lecture à haute voix, ou graphie des mots respectueuse de l'usage), je me propose d'examiner comment améliorer l'accès aux articles par les formes phoniques et l'utilisation des indications phonétiques fournies. Les orientations qui se dégagent de cette étude sont établies dans le cadre d'une création de dictionnaire électronique destiné à des élèves francophones de 11 à 15 ans (plus autonomes que les lecteurs débutants mais dont la maîtrise linguistique doit encore progresser) ou allophones de niveau intermédiaire ou avancé. Elles s'appuient sur ce qui est proposé dans trois dictionnaires publiés par Le Robert, l'éditeur français qui a attaché le plus de soin au traitement des prononciations dans ses produits: le *Petit Robert* électronique (éditions 2001 à 2008), qui est le plus élaboré des dictionnaires généraux électroniques français quant à l'accès aux descriptions des formes phoniques, le *Robert junior* électronique (éditions 1998 à 2006 – la dernière sous le titre *Le Robert des enfants*), qui dispose des mêmes fonctions de recherche que le premier, mais dont le texte, destiné aux élèves de 8 à 11 ans, est moins riche, et le *Robert oral-écrit* (1989), dictionnaire imprimé novateur pour apprenants (natifs ou allophones) qui permettait un accès aux graphies à partir des transcriptions de formes phoniques. Complémentairement aux modalités de traitement et de consultation inspirées de ces dictionnaires, le recours aux technologies de reconnaissance et de synthèse vocales est envisagé. Il impliquerait des partenariats de recherche et développement.

> The ABBYY Lingvo Platform as a convenient tool for end users and a comprehensive solution for publishers

KUZMINA, VERA

7 – Dictionary Use

Current paper is devoted to ABBYY Lingvo electronic dictionary, which is not only a dictionary software, but a family of software products, jointly termed the ABBYY Lingvo Platform. The ABBYY Lingvo Platform includes a range of dictionary software components, which are aimed at meeting same dictionary users' needs – finding an appropriate translation to the word in the given context. The software parts of ABBYY Lingvo Platform are available in different technological realization and for different usage scenarios: as a mobile software, a software for desktop computers or laptops, and as a client-server software made both for professional translators and language learning beginners. The components of ABBYY Lingvo Platform are: ABBYY Lingvo Desktop, ABBYY Lingvo Mobile, ABBYY Lingvo Server, and ABBYY Lingvo Content dictionary writing system. These products operate as well as in connection with each other through ABBYY data centre and independently from each other, for example when the internet connection is unavailable. The applications enable quick and easy access to the dictionary and reference content provided by dictionary publishers, other content providers and also to the content created by the dictionary users themselves. The main needs of dictionary users are covered, such as the ability to use dictionary content stored on different media (online, desktop and mobile), convenient tools for its usage – wide range of search and lookup capabilities together with user-friendly interface, and the possibility to use dictionary content from different content providers in one format and in the same software 'shell'. We believe that the ABBYY Lingvo Platform will cover the main market needs and help end users to get answer to their questions and give new ideas to content providers how to manage the content.

> From Language-Oriented to User-Oriented Electronic LSP Dictionaries: A Case Study of an English Dictionary of Finance for Indonesian Students

KWARY, DENY

7 – Dictionary Use

The rapid development of Internet technology and the significant increase in the number of non-native English speaking college students urge lexicographers working on LSP dictionaries to create better electronic dictionaries to satisfy the needs of these dictionary users. This paper argues that better LSP dictionaries can only be created if lexicographers move from language-oriented to user-oriented lexicographical solutions. This paper shows that the traditional divisions of monolingual, bilingual and semi-bilingual dictionaries have confined the creation of lexicographical solutions that can thoroughly satisfy the needs of the users. The definitions given in monolingual LSP dictionaries are incomprehensible due to the use of difficult vocabulary. The equivalents given in bilingual dictionaries, though considered the quickest way for second language users to know the meaning of a term, do not really help the users when the equivalents relate to different concepts in L1 from the L2 due to cultural differences and when the equivalents are only the transfer of the L2 words. Combining the definitions and the equivalents, as in semi-bilingual dictionaries, may not work well either due to the overload of information presented to the users. Consequently, the shift from language-oriented to user-oriented has to take place in order to produce better lexicographical solutions. Better considerations on users' competences and characteristics are required in creating better electronic LSP dictionaries. In this paper, the implementation of this user orientation is only shown in the on-going project of an English dictionary of finance intended to give help to Indonesian college students to understand financial texts, but the proposed solutions may also be applicable to other LSP dictionaries with a similar type of users.

> Users Take Shortcuts: Navigating Dictionary Entries

LEW, ROBERT

7 – Dictionary Use

In the present paper we compare the effectiveness of two alternative meaning access facilitators in a monolingual learner's dictionary: a Menu system, placed at the top of a monolingual entry; and a Shortcuts system, where the cues are distributed throughout the entry. We test the two entry formats on 90 Polish learners of English at two CEFR levels, A2 and B1. The task which triggers dictionary consultation is guided translation from English to Polish. Three outcome measures are evaluated: access time to sense, accuracy of sense selection, and translation accuracy. While Menus and Shortcuts turned up no difference in terms of consultation speed, the task success was significantly better in the Shortcuts condition. Sense selection accuracy was also better, though not significanly so, for the Shortcuts. The overall conclusion of our study is that Shortcuts are more user-friendly than Menus, although this may also depend on the form of the cues and the medium of presentation.

> One, Two, Many: Customization and User Profiles in Internet Dictionaries

TRAP-JENSEN, LARS

7 – Dictionary Use

Recent textbooks in lexicography recommend the use of customization in e-dictionaries whereby users or dictionary-makers specify which information categories should be shown on the screen. In this paper I take a look at some online dictionaries and analyze how they solve the task. A few basic types are recognized, based on the answer to questions such as: are the user profiles specified by the user or by the lexicographer? Is the profile defined in relation to the look-up situation or to the user's general background and skills? Is the profile fixed or flexible? Must the profile be specified once and for all, before every look-up situation or can it be changed as the user navigates through the dictionary entry? For practical reasons, I confine myself primarily to English and Scandinavian dictionaries.
The analysis formed part of the preparatory phase of the online version of The Danish Dictionary. Four months after the introduction we can now observe from the log files how users manage the various options they are given. The experience so far is that user profiles that require deliberate action from the user are rarely used. The same holds for other kinds of customization such as advanced search possibilities. For the dictionary-maker there is all the more reason to be careful about configuring the default setting.

> Monitoring Dictionary Use in the Electronic Age

VERLINDE, SERGE AND BINON, JEAN

7 – Dictionary Use

The way in which a user consults a dictionary, navigates through a dictionary article and finds an answer to specific questions is a popular area of research in metalexicography. The successful development of online dictionaries opens new prospects in this area of research. Log files of online dictionaries may provide interesting 'free implicit feedback' (de Schryver and Joffe 2004:187). Thanks to its task- and problem-oriented interface, the *Base lexicale du français* (BLF) allows us to track all dictionary users' actions in a natural setting, outside any controlled research environment. Using these data, it should be possible to make well justified decisions on dictionary design.

> O uso de dicionários na compreensão escrita em italiano LE

ZUCCHI, ANGELA M.T.

7 – Dictionary Use

The use of dictionaries by FL students is a standard fact, but this use is frequently questioned by FL teachers. Based on lexicological and lexicographic studies, in addition to language teaching, the survey described was developed in which empirical research is used as a way to answer the question of whether the help of a dictionary leads to differences in the success of understanding pre-determined lexical units, or whether the context itself is sufficient for such comprehension. To achieve this goal, we invited volunteer students enrolled in the undergraduate program in Italian as foreign language at FFLCH, USP, Brazil to participate in the survey. We established three groups of volunteers, the first using an Italian monolingual dictionary; the second using an Italian – Portuguese bilingual dictionary, and the third not using any dictionaries at all. The test consisted of four reading texts in which forty lexical units were highlighted, and whose proper meanings were to be verified by a multiple-choice test. After being collated, the results were submitted to a statistical analysis carried out by the Center of Applied Statistics (CEA-IME, USP). In addition to the statistical results, the methodology allows, through a template, the various elements present in both macro and microstructures of the dictionaries, which really helped comprehension according to the students, to be examined. The results of this empirical research demonstrated the important role of the dictionary in comprehending lexical units, and therefore, also in the teaching and learning process of a foreign language. Furthermore, this survey supported the study of Pedagogical Lexicography and teaching of the use of dictionaries in FL classes.

> The Treatment of Lexical Collocations of Six Adjectives Related to Feelings in A Sample of Bilingual Dictionaries English-Italian

BERTI, BARBARA

8 – Phraseology and Collocation

The importance of lexical collocations is nowadays undeniable, especially from a SLA perspective. Besides grammar, learners of a second language also need to access information concerning the lexical environment of words. The knowledge of the restrictions on lexical combinability is part of a L2 competence and should be successfully master by learners.
Possibly, the most important source of information on lexis is the dictionary, thus it should also be (or become) a reference tool for the retrieval of collocations. In particular, bilingual dictionaries seem to be favoured by learners; this makes it increasingly more important for them to represent the combinatorial rules that organise lexis on the syntagmatic axis.
This study aims at investigating the presence and the treatment of some adjectival collocations in three bilingual dictionaries English-Italian. Six

adjectives related to the semantic area of emotions are singled out and their nominal, adverbial and verbal collocates looked up in both sections of the dictionaries. The data are then compared to those available from a dictionary of English collocations and from the British National Corpus. From a quantitative point of view, the study highlights an unsatisfactory presence of adjectival collocations, especially when the collocate is an adverb. The very few collocations found in the dictionaries are often closer to free combinations, therefore doubtfully useful to dictionary users. The collocations are not systematically organised and can be found interchangeably under either the base or the collocator, thus creating a feeling of confusion that leads to poor user-friendliness. On the whole, from the analysis of the data, it emerges that the issue of collocation should be taken into greater account from bilingual lexicography.

> Onomasiologisch angeordnete Idiomlexika und ihr Nutzwert für die Translatologie: das Forschungsprojekt FRASESPAL zur deutsch-spanischen Phraseologie

BUJÁN OTERO, PATRICIA

8 – Phraseology and Collocation

Die vorliegende Arbeit basiert auf einer praxisnahen und funktionalen Perspektive auf das Übersetzen von Phrasemen. Häufig werden diese als besonders problematisch beim Übersetzen empfunden. Für die intralinguale Äquivalenzanalyse ist aber erstens eine Unterscheidung zwischen Phrasemen im Text und Phrasemen im Sprachsystem nötig. Bei der Äquivalenzerstellung im ersten Fall müssen die Funktionen des Phrasems und deren Relevanz je nach Kontext und Kotext analysiert werden. Eine Hierarchisierung dieser Funktionen im Ausgangstext im Zusammenspiel mit den spezifischen Anforderungen des jeweiligen Übersetzungsauftrags und – vor allem – der Translatfunktion bestimmt dann eine oder andere Übersetzungsstrategie. Ziel dieser Arbeit ist eine Analyse dieser Aspekte, sowie der Vorteile, die die Erstellung von onomasiologisch angeordneten zweisprachigen Wörterbüchern für die Übersetzungspraxis hervorbringt. Die Analysebasis bildet der onomasiologisch angeordnete Thesaurus von Phrasemen aus dem semantischen Feld LEBEN/TOD, die im Rahmen des interuniversitäres Projekts FRASESPAL zur kontrastiven Phraseologie Deutsch-Spanisch zusammengestellt wurde. Idiomatik-Thesauri wie dieser eignen sich in besonderer Weise zu einem aktiven Gebrauch seitens des Wörterbuchbenutzers, was sich positiv auf die Übersetzungspraxis, sowie auf andere Bereiche, wie zum Beispiel die Didaktik einer Fremdsprache, bewirkt.

> A proposal for an electronic dictionary of Italian collocations highlighting lexical prototypicality and the syntactic-semantic relations between collocation partners

GIACOMINI, LAURA

8 – Phraseology and Collocation

My paper presents a corpus-based case study aimed at designing an electronic dictionary of Italian collocations and focussing on a small set of nouns belonging to the semantic field of *paura/ fear*. In the paper, all of the steps involved in data retrieval, automatic and non-automatic evaluation, collocation selection and lexicographic organization are explained in detail. Lexicographic data are represented as a three-dimensional lexical framework displaying ontological, semantic and syntactic relations among lexemes. On the ontological level, *paura* as an entity is connected to contiguous emotions, but it also serves as a prototype for the category of the lexemes selected, shaping their syntactic and semantic behaviour. Collocations are formally categorized through a set of analytic parameters which enable a detailed lexical description as well as more finely grained dictionary search results. On the microstructural level, substantival collocation partners of the selected nouns are described in terms of thematic roles and semantic features, whereas adjectival collocation partners additionally have a set of principles derived from psychological studies applied to them. Finally, analysis of verbal collocation partners focusses on the interplay of the grammatical function and the thematic role of the noun they cooccur with, and on verbal Aktionsart.

The intended exicographic description rests upon a coherent cross-reference network linking the prototype to the other lemmas, collocation partners to each other, and collocations belonging to the narrower semantic field of *paura,* as well as collocations belonging to other semantic fields (e.g., the semantic field of emotions). At the same time, according to an open source principle, corpus-based lexical data can be deductively expanded using framework information.

> Die Festlegung der Polysemie in einem phraseologischen Wörterbuch Spanisch-Deutsch

HENK, ELISABETH AND TORRENT-LENZEN, AINA

8 – Phraseology and Collocation

Seit ungefähr sieben Jahren ist ein Team von Linguisten, Übersetzern und Studierenden mit der Aufgabe beschäftigt, ein Spanisch-Deutsches Wörterbuch der Redewendungen des europäischen Spanischs zu erstellen. Inzwischen haben wir in mehreren Publikationen die von uns angewandten Kriterien bezüglich unterschiedlicher Aspekte wie zum Beispiel der Strukturierung des Definiens oder der Angaben über den eventuellen ironischen Gebrauch bekannt gemacht. Ziel unseres jetzigen Beitrages ist es, die Richtlinien zu erläutern, denen wir bei der Festlegung der Polysemie folgen.

Wir gehen in dieser Studie empirisch-induktiv vor und stellen unterschiedliche Probleme bei der Festlegung der Polysemie dar, die sich in unserer phraseographischen Praxis ergeben und die uns allmählich zu neuen Erkenntnissen führen. Wir vertreten die Meinung, dass das Wesentliche in einem phraseologischen Wörterbuch das Erfassen der übertragenen, phraseologischen Bedeutung ist und dass die Phraseographie aus diesem Grund eventuell andere Kriterien bei der Festlegung der Polysemie braucht, als diejenigen, die bei der monolexemischen Lexikographie gelten. Die Formulierung nützlicher Kriterien für die Festlegung der Polysemie in der phraseographischen Arbeit soll das Ziel einer anderen Publikation sein. Genauso wie es in der allgemeinen Lexikographie der Fall ist, gilt hier zu sagen, dass das Phänomen der phraseologischen Polysemie im Allgemeinen nicht durch streng systematische Kriterien erfasst werden kann. Gerade im Bereich der Phraseographie ist es jedoch wichtig, neue Technologien wie das Internet zu nutzen und die Methoden der linguistischen Pragmatik weiter zu entwickeln, damit nach und nach intuitive Vorgehensweisen durch wissenschaftliche Erkenntnisse ersetzt werden können.

> Von ‘hinkenden’ Stühlen, ‘tanzenden’ Zähnen und ‘verlorenen’ Verkehrsmitteln. Erfassung und Darstellung italienischer lexikalischer Kollokationen für deutschsprachige L2-Lerner (auf der Grundlage des *Dizionario di base della lingua italiana – DIB*)

KONECNY, CHRISTINE

8 – Phraseology and Collocation

Im vorliegenden Beitrag soll ein am Institut für Romanistik der Universität Innsbruck geplantes Forschungsprojekt (Projektleitung: Ch. Konecny) vorgestellt werden, dessen Ziel in der Erfassung und Darstellung italienischer lexikalischer Kollokationen im Vergleich mit ihren deutschen Äquivalenten besteht. Die Kollokationsglieder selbst sind dem italienischen Basiswortschatz entnommen, so wie er im *Dizionario di base della lingua italiana – DIB* von Tullio de Mauro und Gian Giuseppe Moroni festgehalten ist. Kollokationen sind besondere syntagmatische Wortverbindungen, die von Muttersprachlern meist intuitiv korrekt gelernt und verwendet werden, für L2-Lerner hingegen eine häufige

Fehlerquelle darstellen, weil sie oft von jenen der eigenen Muttersprache abweichen. Ein Italienisch-Lernender sollte z.B. wissen, dass in dieser Sprache ein wackelnder Stuhl ‘hinkt’ (*la sedia zoppica*), ein wackelnder Zahn hingegen ‘tanzt’ (*il dente balla*) oder man einen verpassten Zug oder Bus ‘verliert’ (*perdere il treno / l’autobus*). Dem Projekt liegt ein enges Kollokationsverständnis zu Grunde, demzufolge lexikalische Kollokationen hierarchisch organisierte binäre Konstruktionen repräsentieren, welche aus einem kognitiv übergeordneten Element (Basis) und einem kognitiv untergeordneten Element (Kollokator) bestehen: In *la sedia zoppica* und *perdere il treno* z.B. sind *sedia* und *treno* die Basen und *zoppica* und *perdere* die Kollokatoren. Im Projektbericht soll anhand konkreter Beispiele u.a. gezeigt werden, wie die lexikographische Darstellung metaphorische Weiterentwicklungen und damit den polysemen Charakter von Kollokationsgliedern (*piantare un chiodo nel muro – piantare un coltello nella schiena di qcn.*), antonyme (*un coltello affilato – un coltello smussato*) oder alternative Komponenten (*levare / estrarre / cavare / strappare un dente*) und schließlich die jeweiligen referentiellen deutschen Äquivalente berücksichtigen kann. Die geplante Arbeit richtet sich an Italienisch- und Deutsch-Lernende, Italienisch- und Deutsch-Lehrende, aber auch an ÜbersetzerInnen und DolmetscherInnen, ist also sowohl als Lern- wie auch als Lehrhilfe gedacht. Darüber hinaus soll sie dazu beitragen, ein Bewusstsein für die Bedeutung von sprachspezifischen Kollokationen für das Sprachenlernen auch im Bereich der Lernerlexikographie Italienisch-Deutsch zu schaffen.

> Football Phraseology: A Bilingual Corpus-Driven study

MATUDA, SABRINA

8 – Phraseology and Collocation

Football is the most popular sport in the present century. It has assumed a role beyond that of a national sport by becoming a cultural manifestation. It is now a battleground for several important issues such as economy, the resolution of conflicts, poverty as well as racial and minority awareness (Anchimbe 2008). Football relations across countries have increased significantly over the past decades. In order to regulate these relations, we need to express ourselves through language, that language being, in most cases, English. However, each culture has its own way of playing and supporting its teams, a way that is differently expressed according to each mother tongue . The problem arises when there is an urge to express these particularities in a foreign language.

Football constitutes first and foremost a technical domain, though usually not considered as such. Therefore it involves a specialized language. Aiming at understanding the football vocabulary we propose a detailed study of football phraseology. In order to do so the study is based on the notions of Corpus Linguistics (Bowker & Pearson, 2002; Hunston, 2002; Sardinha, 2004; Sinclair, 1991); *corpus-driven* translation studies (Tognini-Boneli, 2001) and terminology (Krieger & Finatto, 2004; Maia, 2002; Temmerman, 2000). The study relies on the assumption that a term is not likely to be used apart from other lexical items and also on the fact that the protected status that is often attributed to it changes according to the context and the words to which it co-occurs.

The *corpus*, still being compiled, consists of approximately 285 thousand words – 156,146 in English and 127,984 in Portuguese.

Due to the complexity of compiling a representative corpus just a preliminary account of the findings is presented.

> Coals to Newcastle or glittering gold? Which idioms need to be included in an English learner's dictionary in Australia?

MILLER, JULIA

8 – Phraseology and Collocation

English idioms and figurative expressions are used by native English speakers of all ages and from many different English speaking countries. The non-literal nature of idioms can pose a problem for non-native speakers, however, who wonder why taking coals to Newcastle should be a significant action, or where the back of Bourke might possibly be. Many non-native speakers of English in Australia are university-age students, aged between 16 and 22, whose first point of departure in finding the meaning of unfamiliar expressions is likely to be a monolingual English learner's dictionary (MELD). Since the MELDs available in Australia are mainly of British origin, learners of English may therefore not find in them the Australian expressions that are used in general conversation and in the media. Moreover, Australian native speakers of English who belong to different generations may not know or use the same idioms. Students who do learn the meaning of an idiom need to know with whom it is appropriate to use such an expression, and this information is often not available in a MELD.

This paper addresses five idioms and expressions taken from a larger study of 84 idioms in order to examine which of these expressions are known and used by different age groups in Australia and the UK. Native English speakers in Australia and the UK completed 2085 surveys indicating where they had first encountered the 84 idioms and where they would use them. The findings indicate that not all expressions given in the British MELDs are known and used by native speakers in the 16-22 age range in either the UK or Australia, and that Australians use idioms which are often not included in the British MELDs. It is therefore suggested that MELDs used in Australia include more Australian material, perhaps online or via a CD-ROM, and that a more appropriate labelling system be introduced to indicate age as a factor in usage.

> Part-Of-Speech Labelling and the Retrieval of Phraseological units

VRBINC, ALENKA

8 – Phraseology and Collocation

The paper presents some insights into the problems of PoS labelling of lemmata, paying special attention to locating phraseological units where it is essential to identify the correct part of speech of the word under which the phraseological unit is included and dealt with in monolingual learners' dictionaries. When studying the inclusion of individual words, senses and phraseological units in five learners' dictionaries (OALD7, LDOCE5, COBUILD5, CALD3, MED2), it was found that numerous lemmata are equipped with more than one PoS label. Consequently, the user no longer needs to identify each and every part of speech of the word in question. As far as the inclusion of phraseological units in the entry for a specific part of speech is concerned, another method of inclusion is proposed, which can be regarded as a further simplification of the microstructure: i.e., that all phraseological units with one common element belonging to different parts of speech are simply grouped together in one special idioms section without distinction between individual parts of speech. This method is certainly worth applying in monolingual learners' dictionaries.

> Phraseological false friends in English and Slovene and the metaphors behind them

VRBINC, MARJETA

8 – Phraseology and Collocation

The interest in false lexical equivalence reflects the interest in language contact, the observation of which always leads to the conclusion that formally identical and similar words and word combinations in different languages do not necessarily overlap semantically. Dictionaries of false friends deal with one-word lexical items, but false-friend relationship can also be established in phraseology. The aim of this paper is to look at phraseological components of English and Slovene lexicons with a view to identifying and describing the false semantic equivalence between idioms in these two languages.
When studying false lexical equivalence, the closeness or sameness of

form has been made *tertium comparationis*. Several phraseological units that are the same or similar in form but different in meaning in English and Slovene are analysed in the paper. Some of these pairs of idioms show certain common features, such as comparison, emotion, spoken or written communication. Phraseological false friends are illustrated by examples and similarities and differences between the idiom in English and the phraseological false friend in Slovene are commented upon.

Since phraseological as well as lexical false friends represent a great problem in communication, translation and lexicographic treatment, it is necessary to first raise awareness of the lexical traps into which non-native speakers of English as well as any other language may easily fall, regardless of their level of linguistic knowledge. It is, therefore, essential to find and treat these pairs of idioms appropriately and acquaint learners with them by including them in coursebooks, in bilingual, general and especially phraseological dictionaries.

> Going organic: Building an experimental bottom-up dictionary of verbs in science

WILLIAMS, GEOFFREY AND MILLON, CHRYSTEL

8 – Phraseology and Collocation

Choosing what headwords to enter in a dictionary has always been a major question in lexicographical practice. Corpora have greatly helped ease both the choice of words to add, and those to remove, by resorting to frequency counts so as to monitor usage over time. This has been particular valuable in the building of learners dictionaries as, however good earlier word lists may have been, they were built largely in intuition whereas, corpora allow the consultation of large reference corpora for a better picture of current realities. In specialised dictionaries dealing with terminological issues, pure frequency is not a feasible solution for headword extraction. However, linked with extraction patterns and statistical tools, corpora still play a major role in supplying information on terms in use. In this research we aim to tackle a situation that lies in between the needs of an advanced learners dictionary and those of a terminological dictionary in attempting to build a pattern dictionary for verbs used in scientific research papers. In order to select verbs for this dictionary and put them into classes, we propose to use collocational relationships as a tool for both selection and analysis of patterns. The principle here is that a series of high frequency verbs can provide the seeds from which prototypical patterns can be extracted. By moving backwards and forwards from verb to argument and back pattern are revealed that use the statistical selectionning to highlight verbs lower in the frequency list that would otherwise be overlooked. Thus patterns will naturally enlarge the word list by selecting what is statistically significant with a textual environment. These patterns not only illustrate typical usage in a specialised environment, but will also group verbs according to textual functions as authorial positioning and description of processes.

> Sampling techniques in metalexicographic research

BUKOWSKA, AGNIESZKA ANUSZKA

9 – Lexicological Issues of Lexicographical Relevance

Browsing through International Journal of Lexicography archives and other metalexicographic work one could easily notice that sampling techniques are generally neglected by metalexicographers, rarely described exhaustively by the authors themselves and almost never discussed, even though numerous researchers sample in order to make generalizations about the whole dictionary text, usually too large to be studied in its entirety. Not rarely samples consisting of one stretch only, usually selected judgmentally, are used to draw inferences about the whole dictionary text and serve as a basis for statistical analysis, which produces results of uncontrolled reliability. This study aims both at exposing the pitfalls of currently used sampling techniques and at proposing probability sampling instead.
Two basic probability sampling schemes were examined: simple random and stratified selection of pages. Censuses based on three dictionaries, three characteristics examined in each one, confirmed my concerns regarding one-stretch sampling. Simple random selection of pages produced, as expected, far more satisfying results in virtually all the cases. This can be, however, bettered by stratification in case of entry-based characteristics in larger dictionaries. Page-based characteristic, mean number of entries per page in this study, did not benefit from stratification. The smallest of my dictionaries presented a range of problems mostly connected with stratified sampling. Furthermore, empirical evaluation of sampling techniques proposed in Coleman – Ogilvie (2009) demonstrated that randomization within strata is also crucial.

> ‘Offensive’ items, and less offensive alternatives, in English monolingual learners’ dictionaries

COFFEY, STEPHEN

9 – Lexicological Issues of Lexicographical Relevance

This paper discusses lexical items which have been labelled as ‘impolite’, ‘offensive’ or ‘rude’ in monolingual learners’ dictionaries (MLDS). Such items may be grouped into three broad categories. Firstly, there is lexis

which relates to the human body and its functions (e.g. *knockers, dick, to crap, to screw*). Secondly, there are items which refer to people and which are potentially insulting (e.g. *bitch, dago, midget, queer*). Thirdly, there are words and phrases, with a variety of meanings, which have in common the fact that they make use of the potentially rude words referring to the human body. Examples are *to balls something up, not to give a shit, fucking, a piss artist* and *work your arse off.*

The precise aim of the paper is to draw attention to the fact that, wherever possible, learners should be provided with less offensive alternatives to the potentially offensive lexis.

In order to assess the current situation in MLDs, a study was carried out on over 200 such lexical items in recent editions of five dictionaries. The main conclusions reached were that in many cases learners are not being provided with alternative lexis, or else that the alternatives suggested are somewhat banal in nature. It is also proposed that in some cases a contextualized example of a lexical item could be rewritten in order to show learners what a less offensive version of the example would look like.

> Deiktische Konstruktionen des Deutschen aus lexikographischer Perspektive

DOBROVOL'SKIJ, D.O.

9 – Lexicological Issues of Lexicographical Relevance

In German, there are two quasi symmetrical constructions: *vor sich hin* and *vor sich her*. These constructions are relatively frequent; however, their meanings have not been sufficiently investigated and their lexicographic description remains rather poor. These constructions are highly idiosyncratic, at least from the perspective of other languages, i.e. speakers have no chance to use them properly if they do not learn them as idiomatic units of the lexicon. The reason is that the meaning of these constructions does not come about as a result of the composition of meanings of their constituent parts. In this sense, these two constructions have also to be studied within both phraseology and Construction Grammar.

The best way to deal with the semantics of *vor sich hin* lexicographically is to postulate a prototypical meaning of this construction and to impose semantic rules (in the sense of Apresjan), which modify the prototypical meaning according to the context. In other words, here we are dealing with coercion. The semantic features that are good candidates for the structure of the prototypical meaning are 'duration', 'introversion', 'week intensity', 'uncontrallability', 'not result-oriented'. In every single VP-construction, the prototypical meaning is specified through focusing some of the semantic features and/or deleting others.

The corpus-based analysis of the construction *vor sich her* showed similar results. From the theoretical perspective, this construction deserves special attention because of the semantic contribution of the deictic element *her*. It is obvious that here we are not dealing with the 'standard meaning' of *her*. The deictic element *her* focuses here the idea of the 'parallel movement'. Obviously, the deictic elements *hin* and *her* have a much richer semantic potential than is assumed in the traditional lexicological and lexicographic description.

> Onomasiological dictionaries and ontologies

FRANÇA, PATRÍCIA CUNHA

9 – Lexicological Issues of Lexicographical Relevance

There has been an increasing interest in ontologies over the last years by the computer science. This interest can be attributed to the advent of what has been known as Semantic Web. Consequently, a persistent interest in Onomasiological Llexicography has arisen and several authors have written about the relation between thesaurus and ontologies, determined to build bridges between the two representational instruments.

The aim of this paper is to light up the discussion between the similarities and differences between onomasiological dictionaries and ontologies. It also presents definitions for concepts such as onomasiology, onomasiological dictionaries, ontologies, formal ontologies, linguistic ontologies.

It intends to demonstrate that the differences pointed out by some authors to distinguish onomasiological dictionaries from ontologies are not quite evident.

> Terminology, Phraseology, and Lexicography

HANKS, PATRICK

9 – Lexicological Issues of Lexicographical Relevance

This paper explores two aspects of word use and word meaning in terms of Sinclair's (1991, 1998) distinction between the *open-choice principle* (or *terminological tendency*) and the *idiom principle* (or *phraseological tendency*). Technical terms such as strobilation are rare, highly domain-specific, and of little phraseological interest, although the texts in which such word occur do tend to contain interesting clusters of domain-specific terminology. At the other extreme, it is impossible to know the meaning of ordinary common words such as the verb blow without knowing the phraseological context in which the word is used.

Many words have both a terminological tendency and a phraseological tendency. In some cases the two tendencies are in harmony; in other cases there is tension between them. The relationship between these two tendencies is investigated, using examples from the British National Corpus.

> An outline for a semantic categorization of adjectives

HEYVAERT, FRANS

9 – Lexicological Issues of Lexicographical Relevance

The aim of this paper is to sketch some basic principles on which a full semantic categorisation of adjectives can be founded that will allow for constructing uniform description templates for the individual categories. The underlying idea is that there should be a one to one correspondence between the category and the template used, like it is the case for most of the existing categorisations of nouns. For that purpose adjectival categories need to be defined by 'adjectival' expressions (mostly with present or past participles as their head). Since this procedure generates only a limited number of very general categories (more or less parallel to the verb categories containing static relations), the templates are extended with some semantic features among which the feature *domain*, plays an important role since it offers the opportunity to specify on a subcategorial level the information that most existing proposals for adjective categorisation give: the conceptual field in which the adjective is to be situated. These conceptual fields for adjectives appear moreover to play an important part in the construction of noun templates, not in terms of form but also in terms of content. The ultimate aim of this proposal is to construct a kind of template building grammar the elements of which can be used equally for nouns, verbs and adjectives.

> Inflection and Word Identity

JANSSEN, MAARTEN

9 – Lexicological Issues of Lexicographical Relevance

This article argues that inflection should be seen as a (partial) criterion that defines homonymy: when two word meanings have different inflected forms, they have to belong to different lexical entries. If this were not the case, it could not be maintained that inflection is a property of lexical entries, but we have to rather say that each word sense has its own inflectional paradigm, even though in most cases all senses of a word inflect in the same way. Although there are apparent cases where it looks like inflection might be in fact dependent on word meaning, none of these cases really goes against the hypothesis that inflection is a property of lexical entries, and not of word senses.

> Defining Dictionary Definitions for EFL Dictionaries

KERNERMAN, ARI AND GEFEN, RAPHAEL

9 – Lexicological Issues of Lexicographical Relevance

The following are some of the issues involved in defining headwords, which I will touch on in this paper:
What is the definition of a dictionary definition?
What linguistic styles for definitions are in use today for specific types of dictionaries? (e.g., technical definitions, folk definitions)
Are there differences in style and technique for defining the various parts of speech?
Is it valid to explain a word in terms of a different part of speech for the sake of clarity?
Are definitions in corpus-based dictionaries different from those in non-corpus based dictionaries?
Can (and should) the viewpoint of the lexicographer be completely hidden?
Does saying what a word is not, adequately explain what it is?
Should only active voice be used?
Which is more important, accuracy or comprehensibility?
How do definitions in learners' dictionaries differ from those in general-purpose dictionaries?
Should the definition be translated for reinforcement, in FL learning?
How useful are illustrations?
How useful are synonyms? Can they replace definitions?

> Systematic Polysemy of Nouns and its Lexicographic Treatment in Estonian

LANGEMETS, MARGIT

9 – Lexicological Issues of Lexicographical Relevance

The focus of the study of the polysemy of the Estonian noun (Langemets 2010) was on identifying the systematic patterns of noun polysemy with further perspective to elaborate the principles to encode and represent systematic polysemy of nouns in the database of the one-volume dictionary of Estonian (to appear in 2015) and in the **ƎELex** (= EELex) dictionary management system of the Institute of the Estonian Language.
The analysis was based on the lexical perspective, i.e. on the lexicographic representation of polysemy in the academic six-volume monolingual dictionary of Estonian (1st ed. 1988–2007, 2nd ed. 2009), and the supportive theory of generative lexicon by means of a qualia structure (Pustejovsky 1995). The sample of study consisted of simple nouns (843 headwords in

all), the total of 1738 semantic units covered both the numbered senses and various subsenses. A hierarchy of the semantic types of nouns, adapted from the lexicographic projects SIMPLE[1] and CoreLex[2], as well as the Estonian Wordnet[3], was used as an ancillary means of analysis enabling, in a way, to 'measure' the regularity of alternating word senses.

A result of the analysis is the list of 40 systematically polysemous patterns, presented as the 'golden standard' of systematic polysemy in Estonian (named after Peters 2004). In total the sample (843 headwords) contained 305 sense alternations that could be interpreted as revealing systematic polysemy. Of those, nearly every fourth (72 patterns) involves an ARTEFACT sense, while half (!) of the patterns involve ACTIVITY.

[1] SIMPLE homepage, see http://www.ub.es/gilcub/SIMPLE/simple.html (31.03.2010).

[2] CoreLex Online, see http://www.cs.brandeis.edu/~paulb/CoreLex/corelex.html (31.03.2010).

[3] Estonian Wordnet, see http://www.cl.ut.ee/ressursid/teksaurus/ (31.03.2010).

> Une pratique lexicographique émergente: les dictionnaires détournés

LÉTURGIE, ARNAUD

9 – Lexicological Issues of Lexicographical Relevance

Parmi les nombreuses références bien installées auprès du public, certains auteurs agrémentent le paysage lexicographique français de dictionnaires particuliers: les dictionnaires détournés. Composés de néologismes construits par des procédés tant morphologiques que sémantiques, ces dictionnaires soulèvent des questionnements qu'il est bon de considérer. L'émergence de plus en plus prégnante de dictionnaires de mots inventés conduit des perplexicologues (les 'linguistes hagards devant la prolifération des mots-valises' selon Finkielkraut) à observer ce type de production lexicographique.

L'objet de cette étude sera donc d'exposer les principes du détournement lexicographique afin de mettre en évidence l'apparition d'un genre lexicographique à part entière: les dictionnaires détournés. Nous en présenterons la typologie pour en illustrer la diversité. Ces ouvrages ne sont en effet pas tous construits à l'identique et trois catégories de dictionnaires se dégagent au sein d'un corpus de 40 références, selon les méthodes de création lexicale employées pour bâtir leurs nomenclatures.

Nous présenterons quelques données quantitatives reflétant la multitude d'ouvrages de ce genre parus entre 1979 et 2010. Cette analyse sera l'occasion d'observer la prééminence d'une catégorie, celle des dictionnaires de mots-valises, sur les deux autres. Enfin, il nous sera également permis d'aborder (de façon superficielle) les différents apports théoriques et pratiques du détournement de dictionnaire. Bien qu'ils s'attachent à pasticher le modèle lexicographique classique, les dictionnaires détournés empruntent nécessairement des spécificités à leurs modèles. Ainsi, certains auteurs présentent leurs dictionnaires – de mots-valises essentiellement – comme des ouvrages didactiques permettant à des locuteurs étrangers ou aux enfants d'acquérir le vocabulaire du français. D'autres militent de façon parodique pour l'intégration de leurs néologismes dans le dictionnaire de l'Académie française.

Tous ces aspects font des dictionnaires détournés un vaste champ d'investigation que nous introduirons dans cet article.

> Multilexical units and headword status. A problematic issue in recent Italian lexicography

MARELLO, CARLA

9 – Lexicological Issues of Lexicographical Relevance

The paper will discuss the headword status of multilexical units in Italian monolingual dictionaries and will include a comparison of Italian and Spanish dictionaries. Twentieth century monolingual lexicographies of Romance languages recognized and registered multiword units, but did not promote them easily to headword status. Italian and Spanish monolingual lexicography in particular have very few multilexical units whereas French has a few more. The initial infiltrations through the 'one-word headword' wall came through Latin borrowings (*alter ego* 'second self', *aut aut*, 'forced choice', *tabula rasa* 'blank sheet'), through two (or more for French) centuries of French and Anglo-American multiword borrowings entering gradually into the Italian language and then into monolingual dictionaries macrostructures (for instance *ballon d'essai* 'trial balloon', *malgré lui*, 'despite him', *fair play, self-made man* are XIX century borrowings; *j'accuse*, 'denunciation', *au pair, best seller, on the road* are XX century borrowings), and in recent decades through the macrostructures of bilingual English-Italian dictionaries where English multilexical headwords are registered and brought to the attention of Italian monolingual lexicographers as multiword units with headword status in English monolingual dictionaries. A status which might determine them becoming multilexical headwords also in Italian monolingual dictionaries. Nowadays most Italian multiwords still remain registered under one-word headwords, even adjectival or adverbial phrases which cannot occur as single words (as for instance *alla carlona* 'carelessly', *a perdifiato* 'at the top of one's voice' registered under the headword **carlona, perdifiato,** words with combinatorial usage only. Italian corpora can help define the confines of the multilexical unit and establish possible variations, such as widespread elliptical uses. Coherent corpus-based decisions are in turn extremely valuable not only for lexicographers, but also for POS tagging of corpora in which the multilexical units are recognized and entered as a whole in addition to the single parts.

> A Semantic and Lexical-Based Approach to the Lemmatisation of Idioms in Bilingual Italian-English Dictionaries

MULHALL, CHRIS

9 – Lexicological Issues of Lexicographical Relevance

The aim of this paper is to propose a new semantic and lexical-based lemmatisation framework for the recording of idioms in bilingual Italian-English dictionaries. Many of the difficulties and inconsistencies characterising the lexicographic treatment of idioms stem from them being viewed as a semantic and lexically homogenous phrasal category. This incorrect generalisation typically motivates the traditional description of idioms as being non-compositional and lexically fixed units. Current bilingual Italian-English dictionaries treat idioms quite unsystematically, mainly due to their reliance on the subjective judgement of lexicographers and generic syntax-based listing strategies. The rationale for pursuing these methods remains unclear, particularly given the availability of substantive semantic and lexical information that could provide a more defined template for determining the position of idioms in a dictionary. This paper looks at two particular aspects of idioms in five current bilingual Italian-English dictionaries: *Il Ragazzini* (ZIR) (2009), *Grande Hoepli Dizionario Inglese* (GHDI) (2007), *Il Sansoni Inglese* (SI) (2006), *Oxford Paravia Italian Dictionary* (OPID) (2006) and *Hazon Garzanti Inglese* (HGI) (2009). The first is a semantic-based investigation, which analyses the entry procedures for 150 English and 150 Italian idioms across three categories: pure idioms, figurative idioms and semi-idioms. The second examines the listing strategies for 40 English and 40 Italian idioms with variable verb and noun components. Overall, two particular trends emanate from the analysis. Firstly, the arrangement of idioms is unsystematic and the allocated entry points do not reflect or emphasise their individual semantic or lexical features, which are central to their identity. Secondly, the English-Italian and Italian-English sections of certain dictionaries are disparate in their overall coverage with Italian idioms assigned a greater number of listings. These discrepancies call for a formulaic entry model that eliminates the subjectivity, inconsistency and unsystematic approach currently associated with the treatment of idioms in bilingual Italian-English dictionaries.

> Seeing through dictionaries: On defining basic colour terms in English, Japanese and Polish lexicography

PAKUŁA, ŁUKASZ

9 – Lexicological Issues of Lexicographical Relevance

It seems that despite the undeniable fact that colour research has received considerable attention for centuries resulting in more than 3000 publications during the last 150 years (MacLaury 1997 after Steinvall 2002), there still exists a niche to be filled. There has been no or very little research regarding colour terms conducted from the viewpoint of (meta) lexicography. The present study is meant to evaluate existing dictionary definitions of Basic Colour Terms (henceforth BCTs) from the colour lexicon of English, Japanese and Polish in order to detect any doubtful content which could be improved to equip the dictionary user with richer, more adequate information regarding the colour lexicon. The immediate aims of the study are to determine: 1) what definition types are used to define CTs 2) what prototypes extensional definitions point to when defining BCTs and how these relate to the data obtained from naive native speakers of the languages in question. To this end, two empirical investigations were conducted. The first one is devoted to dictionary definitions, while the second one is an experiment carried out among naive native speakers of the three languages.

> Getting through to phrasal verbs: A cognitive organization of phrasal verb entries in monolingual pedagogical dictionaries of English

PERDEK, MAGDALENA

9 – Lexicological Issues of Lexicographical Relevance

The abundance of English phrasal verbs along with their syntactic and semantic complexity has always been a stumbling block for learners of English. Some think of phrasal verbs as hallmarks of a native-like command of English but there is no universal method to learn their natural contexts or applications and no ready-made recipe to deduce their meaning is available. Therefore, more attention should be paid to the accurate lexicographic description of phrasal verbs in learners' dictionaries, which are often the first source of reference for students. Moreover, dictionary compilers should aim at such presentation of these structures as to guide the users towards working out the multiple meanings of phrasal verbs on their own by creating cognitive links in the entries or even offering spatial cognitive networks.

The paper looks at the organization of a phrasal verb entry in the most recent pedagogical dictionaries of English from the cognitive perspective.

The layout of the entries is examined with focus on the methods used to differentiate the many meanings of phrasal verbs, especially figurative ones and an attempt is made to find any cognitive links that are used to generate helpful associations and predictions about the meaning. In his recent paper on phrasal verbs, Brodzinski (2009) calls for such an associative approach to presenting phrasal verbs to learners, be it in class or in a dictionary. His claim is that for pedagogical purposes it is better to replace the multiple meanings of a given phrasal verb with one core meaning along with applications.
An alternative to the linear organization of a phrasal entry could be a network of meanings underlying any possible cognitive links between different senses. Such an approach might prove to be more stimulating for non-native users. Three examples of such networks, each with different semantic focus, are presented in the paper.

> The Frisian Language Database as a tool for semantic research

SLOFSTRA, BOUKE AND VERSLOOT, ARJEN

9 – *Lexicological Issues of Lexicographical Relevance*

In this paper, the authors present two examples from the field of body parts, illustrating the level of details, very often with a diachronic component, that can be detected in the study of semantics. The bilingual background of Frisian – Dutch is in one way or another a second language in Friesland already since the late Middle Ages – constitutes an extra trigger in the organisation and restructuring of the lexemes and their meanings. The given examples from Google and the Frisian Language Data Base illustrate that a diachronic and comparative corpus based approach can add several aspects that have remained uncovered so far in the more traditionally conceptualised WFT. This enhanced picture of meanings and their developments are an essential prerequisite for gaining a deeper understanding of the organisation and structure of human semantic concepts and their operationalisation in human speech.

> From lexicological to lexicographical issues: Italian verbs with predicative complement

STRIK LIEVERS, FRANCESCA

9 – *Lexicological Issues of Lexicographical Relevance*

Intransitive (e.g., *become*) and transitive (e.g., *consider*) verbs obligatorily requiring a predicative complement are an interesting, and at the same time problematic issue both at a theoretical and at a lexicographical level. In this paper we focus on Italian verbs, and on the way two computational semantic lexica deal with them. Both in ItalWorNet and in SIMPLE the treatment of these verbs shows to be problematic, since the information appears to refer to the 'verb + predicative complement' complex rather than to the verb itself. Recognizing that verb and predicative complement contribute to the construction of a unitary event, we believe that it is nevertheless possible, and useful, to isolate the role of the two components. The description proposed here is based on the Generative Lexicon model (Pustejovsky 1995), and it is in line with the recent project of a lexical resource for (sub)event structure (Im and Pustejovsky 2009). Verb and predicative complement codify each a different part of the subevent structure. To give an example, '*diventare* ('become') + predicative complement' is a transition, where *diventare* codifies the process subevent, and the predicative complement codifies the (result) state subevent. This kind of analysis can possibly be integrated into the SIMPLE lexicon, which is already built following the Generative Lexicon model.

> On defining the category MONSTER – using definitional features, narrative categories and Idealized Cognitive Models (ICM's)

SWANEPOEL, PIET

9 – Lexicological Issues of Lexicographical Relevance

This paper explores how the coherence between a lexical item which denotes a category and the lexical items that refer to individual members of the category can be expressed in explanatory dictionaries. A detailed analysis is provided of the relationship between the lexical item *monster* (which refers to a category) and the lexical items that refer to individual members of this category (e.g., Cyclops, dragon, mermaid, vampire, werewolf, Dracula, and zombie). More specifically, the goal of the paper is to determine whether the semantic explanation(s) for *monster* could function as a dictionary internal (as opposed to Fillmore's (2003) external) cognitive frame for the other lexical items in the monster set. If not, the question is whether and how the field of monsterology could assist one in designing such a frame and what the content, structure and function of such a frame would be.
In Section 2.1 the focus falls on current lexicographic practices and problems in defining the category monster and its members. The dictionary entries for *monster* and those of a number of its members in a selection of English explanatory dictionaries are surveyed to determine what cognitive models of the category monster underlie these definitions. In Section 2.2 the focus falls on the definitional features, ICM'S and narrative structures used to define the category of the monster in the field of monsterology and on the numerous meanings monsters may have as symbolic expressions (metaphors in particular). Section 3 shortly summarizes the contribution monsterology could make towards the definition of a monster frame.

> Metonymical Object Changes in Dutch: Lexicographical choices and verb meaning

SWEEP, JOSEFIEN

9 – Lexicological Issues of Lexicographical Relevance

The Dutch term *objectsverwisseling* (literally: 'object change') is a lexicographical label used to describe specific combinations of a verb with two qualitatively different direct objects. Illustrative examples are *de borden / de tafel afruimen* ('to clear the plates / the table'), *hout / een vuur / de haard aansteken* ('to light wood / a fire / the fireplace'), *riet / manden vlechten* ('to weave reed / baskets'), *gaten / sokken stoppen* ('to darn holes / stockings'), *sinaasappels / sap persen* ('to press oranges / juice'), *eieren / kuikens uitbroeden* ('to hatch eggs / chicks'), etc.

These examples are often analysed as specific instances of metonymy (cf. Adelung 1811; Van Dale 2005; Koch 2001; Waltereit 1998). Both possible direct objects are interchangeable because they are conceptually connected by their existence as a conceptual unity in the real world (such as a set table, a wood fire, reed baskets, etc.). There are, however, some discrepancies between linguistics studies of metonymical object changes (MOCs) and lexicographical choices in dictionaries. These basically concern the question of whether an object change affects the meaning of the verb.

On the basis of theoretical considerations as well as lexicographical descriptions I will try to clarify to what extent MOCs influence the meaning of a verb. To this purpose, I will evaluate the incorporation of MOCs in two standard Dutch dictionaries, i.e. *Van Dale Groot Woordenboek van de Nederlandse Taal* (2005) and *Woordenboek der Nederlandsche Taal*. Theoretically, it will turn out to be necessary to distinguish between grammatical-relational information and lexical meaning (cf. Brdar 2007: 181). I will argue that MOCs actually provide evidence for the fact that the verb has one lexical meaning. In this way, the present paper gives more insight into the object changes, into the underlying metonymy and also into verb meaning in general. These insights may subsequently be useful in the improvement of dictionary entries.

> Metonymy representation in English monolingual learners' dictionaries: Problems and solutions

WOJCIECHOWSKA, SYLWIA

9 – Lexicological Issues of Lexicographical Relevance

The paper aims to show how the tenets of the cognitive theory of metonymy can benefit the representation of metonymic lexemes in pedagogical lexicography, so that the semantic connections between

basic and derived meanings become more transparent and motivated. It reports the results of a lexicographic study into the representation of conventionalised metonymic lexemes in the five most renowned English monolingual learners' dictionaries (henceforth MLDs): CALD2, COBUILD4, LDOCE4, MEDAL2 and OALDCE7. The study focuses on three elements of the dictionary entry: sense arrangement, definition, and the correlation between noun codification and exemplification. These features are evaluated against the background of both the cognitive theory of metonymy and the widely accepted principles of lexicographic practice. Significant inconsistencies concerning the treatment of metonymy are found within each dictionary, as well as numerous cases where the semantic relationship between the source and target senses of a metonymic lexeme is broken. It is also noticed that in the case of metonymisation which results in change of noun's countability, noun codes are sometimes ambiguously assigned, and some examples of usage do not explicitly show the count-mass distinction. Solutions are offered to arrive at a more systematic, transparent and cognitively oriented representation of metonymy. These include using template entries in the compilation process, subsuming the definition of the metonymic target under the source definition, and defining the target as a semantic elaboration of the source.

> The German-Lower Sorbian Online Dictionary

BARTELS, HAUKE

10 – Lexicography of Lesser Used or Non-State Languages

After the publication of a new and comprehensive Lower Sorbian-German dictionary in 1999, the urgent need for an active learner's dictionary has been widely felt. Some specifics of the sociolinguistic situation of Lower Sorbian must have direct impact on the conception of such a dictionary: For almost all speakers of younger generation German is the first and better known language. German-Lower Sorbian interference, a very small or only partially elaborated vocabulary, and an often defective command of grammar, especially of those parts of it lacking in German, is widespread.
Since 2001 the Lower Sorbian Department of the Sorbian Institute works on a dictionary that tries to meet the requirements of that target-group. With respect to the fact that Lower Sorbian is highly endangered and there is no time to lose, all information is published on the internet as quickly as possible. In 2003 a first version of the online dictionary 'Deutsch-niedersorbisches Wörterbuch' (DNW) was launched. At the present the DNW contains about 70,000 entries, but it will continually be extended and corrected; it is still considered a draft version.

Apart from some technical background information, the paper gives an overview of the lexicographic description. In order to help to avoid typical L1-interferences and to actively use the minority language the dictionary offers, for example, additional information about the use of verbal grammatical and lexical aspect (*Aktionsart*). Also support verb constructions (so-called *Funktionsverbgefüge* in German), where direct translations of the German construction often lead to a non-idiomatic language usage, are taken into consideration. For a better integration of such and other important information, some new conventions have been introduced, hoping that the DNW will function as a learners' dictionary as well as a contribution to language documentation.

> WFT: The comprehensive Frisian Dictionary (*Wurdboek fan de Fryske taal / Woordenboek der Friese taal*)

BOERSMA, PITER

10 – Lexicography of Lesser Used or Non-State Languages

The *Woordenboek der Friese taal* is a dictionary of a regional minority language. Yet it may be compared to the big scholarly dictionaries of national languages like Dutch, German and English, not because of its size but with respect to its principles. The *Woordenboek der Friese taal* is, as a description of a minority language, in this sense unique. Its more modest size is partly due to the dictionary's design, but a more important reason is that the lexicographical description of Frisian is hampered by the absence of a large variety of written sources, because Frisian, characteristically as a minority language, especially functions as a spoken language.

In my paper I clarify how the position of a minority language – and in addition the scholarly infrastructure – are decisive for the lexicography of Frisian and the compilation and contents of the *Woordenboek der Friese Taal* in particular.

Before discussing some aspects of the WFT itself I will deal with three items. 1. The unfinished dictionary *Lexicon Frisicum (A-Feer)* (1872) by J. H. Halbertsma (1789-1869), the founding father of the lexicography of modern Frisian; 2. The continuation of Halbertsma's lexicographical work, resulting in the *Friesch Woordenboek* (1900-1911). 3. The preamble to the *WFT*.

I discuss the following aspects:

- the choice of the non-Frisian metalanguage in the dictionaries above
- the choice of only post-1800 Frisian in the WFT.
- the choice of regional variants in *Friesch Woordenboek* and WFT
- the choice of including the first attestation of each entry into the WFT

- the microstructure of the WFT
- etymology in the WFT

I finally mention the future of the WFT: with the completion of the the paper dictionary, the WFT is now ready to enter the exiting world of online electronic dictionaries.

> The language norm in a century of Frisian dictionaries

DUIJFF, PIETER

10 – Lexicography of Lesser Used or Non-State Languages

Since the Renaissance in Western Europe language builders have been making efforts to standardise languages. In the Netherlands and in Belgium the Dutch standard language became accepted generally. The standardising process of Frisian, the second official language in the Netherlands, was different and it still is. The standard for Frisian had not crystallized out yet. In practice, this means that several variant forms and pronunciations are accepted. A series of Frisian dictionaries have been published in the past hundred years. In this contribution the question will be answered whether these dictionaries contributed to standardising Frisian. Did the dictionaries reflect dialectal diversity or did they have a prescriptive design? In answering this question the position of four phonological variations in the dictionaries has been investigated. Also has been made an inventory of the editor's comments on their selection of dialect forms.

On the base of the results the conclusion must be that Frisian dictionaries did not use one and the same standard language consequently. A small number of dictionaries consciously prefer to include only one variant entry form. Just the more elaborate and widely used dictionaries show a rather tolerant standard of language, though not in consequent fashion. The electronic language databases of the Fryske Akademy show that in practice the choices made by the most frequently consulted dictionaries were followed generally.

In the paper also a picture of the Frisian language and a briefly description of the history of Frisian lexicography have been given.

> Can the new African Language dictionaries empower the African language speakers of South Africa or are they just a half-hearted implementation of language policies?

KLEIN, JULIANE

10 – Lexicography of Lesser Used or Non-State Languages

Language planning was always a very sensitive topic in South Africa, as languages was used to separate people during apartheid. This presentation analyses three different Sesotho sa Leboa dictionaries, which can be seen as examples of a successful implementation of language policies. The policies which are discussed here are the constitution of the Republic of South Africa form 1996, The National Lexicographic Units Bill from 1996 and the South African Languages Bill from 2000. The main objective of those language policies is the development and promotion of the eleven official South African languages. Dictionaries are one possibility to develop languages, .e. they describe the standardised variety of a language. They can be used as tools to promote the African languages, as they are the visible proof that the language has the words to be used in a specific situation, for example a dictionary of Maths shows that the language has words for mathematical concepts.

The three dictionaries which are discussed here are a Sesotho sa Leboa – English general dictionary which was published by the Sesotho sa Leboa National Lexicographic Unit, a bilingual Sesotho sa Leboa English school dictionary published by OUP South Africa and a Sesotho sa Leboa – English online dictionary published by TshwaneDJe HLT. This presentation discusses the advantages of each dictionary and shows that they all can empower their users but that none of the three dictionaries can cater for everybody in all situations because there is no such thing as THE dictionary that provides a solution for everything.Let's detail the opportunities offered by the online dictionary market in three areas:

- Search Engine Optimization (SEO): why dictionary content is a marvellous resource to answer a wide range of queries in search tools such as Google, Bing, Yandex or Baidu,
- Reaching local markets worldwide with bilingual content,
- User Generated Content: an unmissable resource.

> Dictionaries and their influence on language purification in minority languages. The case of Frisian

KUIP, FRITS VAN DER

10 – Lexicography of Lesser Used or Non-State Languages

In literature, scepticism on the effect of language propaganda is dominant. Researchers observe that it is almost impossible to stop lexical interferences from becoming current in standard languages such as Dutch (or Southern Dutch in Belgium) through language purification literature or through language-related articles and transmissions in the media or (last but not least) through concise dictionaries.

The question we have to ask ourselves is, whether the dictionaries' influence in a minority language such as Frisian is limited as well. Most speakers of Frisian are, as far as writing is concerned, illiterate in their

own language. They are not accustomed to written Frisian word forms and unsure when it comes to how their language should be written correctly. A Frisian speaker will be more inclined to consult a dictionary when writing something in his own language, than a speaker of a majority language or a national language would do. On the basis of that assumption, you would expect that including purisms and avoiding or marking interferences in dictionaries, would significantly affect the written language at least.
In this survey, I looked at four loan-words, including the loan-translations and purisms (if any) that go with them. I compared the occurrences (and non-occurrences) of these words as dictionary entries to their respective frequencies of occurrences in two major databases.
On the one hand, we see that, throughout the years, the purisms included in the dictionaries perform considerably better than the equivalent loan-words and loan-translations. The purisms not in the dictionaries perform considerably worse. On the other hand, we notice a trend among writers of Frisian to use interference words in the last few decennia. So, at first glance, dictionaries seem to have influenced language purification. However, one cannot tell for how long that will be the case. It will depend on speakers' attitudes towards their language. After all, it is very difficult to control a language as has been proven in the case of Dutch, and the same might hold for Frisian.

> Mobile phone dictionaries for small languages: the Whitesands electronic dictionary

McELVENNY, JAMES

10 – Lexicography of Lesser Used or Non-State Languages

This poster presentation reports on work to develop an electronic dictionary of Whitesands (Austronesian; Tanna Island, Vanuatu) which can be stored on and accessed through mobile phones. In the presentation we will outline some of the benefits of mobile phone electronic dictionaries for speakers of small languages, look at some of the difficulties that we had to overcome in preparing the dictionary, and discuss the reception of the dictionary in the Whitesands community.

> Scottish Lexicography: Major Resources in Minority Languages

PIKE, LORNA AND ROBINSON, CHRISTINE

10 – Lexicography of Lesser Used or Non-State Languages

This paper focuses on current aspects of the lexicography of two minority languages in Scotland, Scottish Gaelic and Scots, and looks at two projects at either end of the lexicographical spectrum: Faclair na Gàidhlig, an on-

line full historical dictionary of Gaelic and the new edition of the *Concise Scots Dictionary* (*CSD*, 1985), a one-volume derived dictionary of Scots.

A brief outline of the history of both languages is given. Each in turn was the dominant language in Scotland until both were replaced by English.

The paper looks at how Scotland's minority languages have benefited from the skills of the Scots who contributed to English lexicography. Sir James Murray, first Editor of the *Oxford English Dictionary* (*OED*), pioneered the application of historical principles to English lexicography and his colleague, Sir William Craigie, applied those same principles to the *Dictionary of the Older Scottish Tongue* (*DOST*) which covers the Scots language from the 12th century to 1700. These skills are now being transferred into Scottish Gaelic.

Faclair na Gàidhlig will be an on-line historical dictionary of Gaelic compiled on similar principles to *OED* and *DOST*. The major challenge in establishing a project of this magnitude is to create a lexicographical tradition as effectively and efficiently as possible. The paper outlines the approach adopted. A draft noun entry is examined with discussion of entry structure and organisation.

Scots is equipped with two historical dictionaries, *DOST* and its modern counterpart the *Scottish National Dictionary* (*SND*). *CSD* is a one-volume distillation of these works. The second edition will use a more user-friendly structure and update coverage to the 21st century. Sample entries are examined.

Scottish lexicography will continue to build on its historical tradition providing Gaelic and Scots with resources comparable to English.

> The Lexicographic Work of Euskaltzaindia – The Basque Language Academy 1984-2009

SAGARNA, ANDONI

10 – Lexicography of Lesser Used or Non-State Languages

The Academy started developing a standard variety of Basque Language in 1968. In 1983 the Academy created a commission of lexicography, and in 1984 approved a long-term plan for the development of dictionaries. That plan included the following projects:

1 The General Basque Dictionary (GBD), which should be a compilation of the lexicon used in the publications until 1970.

2 A lexicology project, whose aim was to study the formation of words in Basque.

3 A compilation of the lexicon used in current publications

The corpus of GBD contains 6 million text words, The first result of the project was a dictionary of 16 volumes in paper format. Since October

of 2009 this dictionary is available online at the address http://www.euskaltzaindia.net/oeh . The result of the compilation of the lexicon used in current publications is The Statistical Corpus of the Twentieth Century that contains 4,658,036 words from 6,351 pieces of text. It is available online at the address http://www.euskaracorpusa.net/XXmendea/Konts_arrunta_fr.html.

In 1992 the Academy created a commission to prepare The Unified Dictionary of the Basque Language. The General Basque Dictionary and The Statistical Corpus of the Twentieth Century are precisely the information sources that provide insight into the use of words. In 2000, was published a list of standardized 20,000 words and in 2008 a second edition collected a total of 29,000 words. By the end of 2011 the list should contain about 40,000 forms. Regardless of the publications on paper, the unified dictionary is available at http://www.euskaltzaindia.net/hiztegibatua

> Lexicography of a Non-State Language: The Case of Burgenland Romani

SCHRAMMEL, BARBARA AND RADER, ASTRID

10 – Lexicography of Lesser Used or Non-State Languages

Burgenland Romani (henceforth BR) is spoken in Burgenland, the easternmost province of Austria. Until recently BR was an exclusively oral language. However, active language use of BR has almost totally ceased in the second half of the 20th century. The self-organisation of the group from the 1990s onwards led to a new appreciation of the language, which is now accepted as the primary identity marker. This new interest in their own language and culture entails the desire for the revival, maintenance and spread of BR. One aspect of language planning in BR concerns the functional expansion of the language into acrolectal domains where it has never been used before.

BR is lexicographically documented in two different media, i.e. in ROMLEX (henceforth RL), which is an extendible multi-dialectal lexical database with a freely accessible web-interface (http://romani.uni-graz.at/romlex/) and a print dictionary. RL is intended as a tool for comprehensive lexical documentation of BR. At the same time, it is a practical, low-threshold tool for text producers. The print dictionary, on the other hand, primarily serves an emblematic purpose. Given the differing purposes of RL and the print dictionary, different strategies are used in lexicographic decision-making. Roughly speaking, RL favours an inclusive descriptive approach while the print dictionary is rather restrictive and follows normative principles. The paper discusses decisions taken with respect to orthography, lemma selection and meaning for RL and the print dictionary, respectively. We are highlighting lexicographic phenomena, such as increased polysemy, generic usage of terms and heavy borrowing, which are typical of the functional expansion process of stateless minority languages.

> An Overall View about Lexicography Production for the Friulian Language

TOFFOLI, DONATO

10 – Lexicography of Lesser Used or Non-State Languages

Researches about distinctive characteristics of Friulan language began in the second half of the Eighteenth Century. There was a need to have lexicography instruments that could help to understand and convert the lexicon heritage of the language actually spoken in large parts of Friuli, Friulian language, in the quite unknown linguistic code of the new born Italian state, the Italian language.

Some of these lexicographic works still exist and are nowadays very important: for example '*Il Nuovo Pirona*', a formidable tool for the dialectological work; also other interesting tools were printed such as, for example, '*Vocabolario della lingua friulana*' by Giorgio Faggin, or '*Dizionario pratico illustrato italiano-friulano*' by Maria Tore Barbina.

A meaningful change happened in the 90's when the first law to safeguard and promote the Friulian language was approved and the first body for the linguistic policy, l'*Osservatorio Regionale della Lingua e Culture Friulane* (OLF), was founded. The lexicographic work began to be more structured.

A great work has been done for computer medium: a spelling corrector: '*Coretôr ortografic furlan*' by '*Informazione Friulana*' Cooperative; a dictionary on CD '*Dizionari Ortografic Furlan*' (DOF) by Alessandro Carrozzo.

Lately The most important lexicographic work is '*Grant Dizionari Bilengâl Talian Furlan*' (GBDTF), by *Consorzio Friûl Lenghe* 2000, based on a prestigious Italian model such as the '*Grande Dizionario dell'Uso della Lingua Italiana*', by professor Tullio de Mauro that coordinated the Friulian version too.

> The Klagenfurt lexicon database for sign languages as a web application: LedaSila, a free sign language database for international use

KRAMMER, KLAUDIA

11 – Sign Language

The Klagenfurt online database for sign languages 'LedaSila' (Lexical Database for Sign Languages, http://ledasila.uni-klu.ac.at/) is designed in such a way that it is possible to present all the information which can be found in any good monolingual or bilingual (printed) dictionary. It offers the possibility to enter semantic, pragmatic as well as morphosyntactic information. Furthermore, a detailed analysis of the non-manual and manual parameters of a sign is possible. LedaSila offers the possibility to search for any information already contained in it (including single signs or formational parameters), to document a sign language, or analyse it linguistically. The search function is accessible to all internet users. When using the database for sign language documentation and/or analyses, an authorisation from the Centre for Sign Language and Deaf Communication in Klagenfurt is required.

When using LedaSila for documentation and/or analysis of a sign, a user does not have to follow a specific order when entering the data. Furthermore, the user is free to decide whether to enter data only in one field (e.g. semantics or region) or to do a full analysis of the sign. A special feature of LedaSila is the possibility to add new categories and values at any time. This is especially important for an analysis tool which is designed to be used internationally. This feature ensures that all categories and values needed for a specific sign language are available.

LedaSila can be used free of charge for non-commercial deaf and scientific issues. The database is hosted on a server of the University of Klagenfurt. All information (including videos) is stored directly on the web server. This means that using LedaSila comes with zero administration. The international sign language linguistic community is invited to use this easily manageable database.

> The Danish Sign Language Dictionary

KRISTOFFERSEN, JETTE H. AND TROELSGÅRD, THOMAS

11 – Sign Language

The entries of the The Danish Sign Language Dictionary have four sections:

Entry header: In this section the sign headword is shown as a photo and a gloss. The first occurring location and handshape of the sign are shown as icons.

Video window: By default the base form of the sign headword is shown. Other types of videos are rendered in this window, but activated by clicking play buttons in different sections of the entry.

Meanings window: In this section the meanings of the sign are shown. The meaning description includes: Danish equivalents, a description of the sign's usage (for function signs) and mouth movement, cross-references to synonyms etc., information about restricted use, and example sentences. Semantically opaque compounds with the sign are shown below the regular meanings.

Additional information: This section holds cross-references to homonyms and to common frozen forms of the sign (only for classifier entries). In addition to this, frequent co-occurrences with the sign are shown in this section.

The signs in the The Danish Sign Language Dictionary can be looked up through:

Handshape: Particular handshapes for the active and the passive hand can be specified. There are 65 searchable handshapes.

Location: Location is chosen from a page with 15 location icons, representing locations on or near the body.

Text: Text searches are performed both on Danish equivalents, sign glosses and example sentences (both transcriptions and translations). This enables users to find signs that are not themselves lemmas in the dictionary, but appear in example sentences.

Topic: Topics can be chosen as search criteria from a list of 70 topics.

> The first national Dutch Sign Language (NGT) Dictionary in book form: Van Dale Basiswoordenboek Nederlandse Gebarentaal

SCHERMER, TRUDE AND KOOLHOF, CORLINE

11 – Sign Language

In October 2009 the first national Dutch sign language (NGT) dictionary in book form was published by Van Dale publishers. (Schermer,Koolhof (eds) 2009). The content of the book is produced by the national centre for NGT and for sign language lexicography, the Dutch Sign Centre, and is based on 25 years of research into the lexicon of Dutch Sign Language (NGT) which we will describe briefly in our paper (Schermer 1990, 2004). Subsequently we will describe organisation and content of the Van Dale dictionary which contains 3000 standard signs with illustrations ordered alphabetically by using a glos as lemma.. In addition to the Van Dale dictionary in book form an online NGT dictionary is available on our website (Schermer,Koolhof,Muller 2010) which offers both search features: alphabetically and via handshape/location. Each entry in the Van Dale dictionary contains further information: an example sentence of how the sign is used and grammatical information about the non manual features and type of verb. We will show examples from both dictionaries, discussing the dilemma's we faced and the solutions we opted for in the making of this dictionary.