

---

WINTER 2001



## EURALEX NEWSLETTER

Editors: Carla Mareello and Rosamund Moon.

Address for correspondence and contributions: Carla Mareello, Corso Unione Sovietica 115, I-10134 Torino, Italy.

Email: [mareello@cisi.unito.it](mailto:mareello@cisi.unito.it) and [rosamund@moon.eu.org](mailto:rosamund@moon.eu.org)

---

### The EURALEX Newsletter

This quarterly Newsletter is intended to include not only official announcements but also news about EURALEX members, their publications, projects, and (it is hoped) their opinions, and news about other lexicographical organizations. Please try to support this by sending newsletter contributions to Carla Mareello at the above address, or emailing the editors. The deadlines for winter (December), spring (March), summer (June), and autumn (September) issues are respectively 15 September, 15 December, 15 March, and 15 June annually.

### The EURALEX Web Site

The URL of the EURALEX web site is <http://www.ims.uni-stuttgart.de/euralex>.

### News from Afrilex

On the 28th and 29th of May 2001, Afrilex hosted a two-day Seminar on Terminology in Pretoria. The seminar was presented by S. Tarp and R. H. Gouws, and was well attended by a delegation from the National Language Service of the Department of Arts, Culture, Science and Technology, and by members of the various National Lexicography Units (NLUs). Even though the seminar was primarily theory-driven, the practical implications for the South African lexicographical landscape were given due attention and proved very enlightening indeed.

The 6th International Conference of Afrilex was held in Pietersburg from 2 to 4 July 2001. Some 30 papers were presented and the conference, although taking place in the deep hinterland of South Africa, was attended by over seventy people. Attendees came from various countries in Africa and Europe, with especially strong Gabonese and Zimbabwean delegations. Keynotes were presented by G.-M. de Schryver (Lexicographers' Dreams in the Electronic-Dictionary Age) and D. J. Prinsloo (Corpus-Based African-Language Lexicography). The programme consisted further of special sessions and parallel

sessions during which the conference theme ('Computational Lexicography and the Development of Corpora') and other issues of lexicographical importance were evenly balanced. The discussions, both those following the presentations and those during the unofficial programme, were lively, and there was a general consensus that African-language lexicography has taken its rightful place in the digital era.

The newly elected Executive Board consists of D. J. Prinsloo (Chairperson), R. H. Gouws (Vice-Chairperson), K. J. Mashamaite (Secretary), M. J. Mojalefa (Treasurer), M. Alberts (Registrar), and G.-M. de Schryver (Organizer). The other elected board members are L. M. Mphahlele, M. H. Mpungose, and F. M. Shoba.

The 7th International Conference of Afrilex will be held in Grahamstown, South Africa, from 8 to 10 July 2002 (immediately following the world-renowned Grahamstown Arts Festival). The conference theme is 'Culture and Dictionaries'. H. Bergenholtz, who has done extensive research in Madagascar and edited a Malagasy-German dictionary in 1991, has been invited to read one of the keynote speeches. Looking further ahead, the 2003 conference is planned to take place in Windhoek, Namibia, and the 2004 conference in Libreville, Gabon. For the latest news, please visit the Afrilex web site at <http://www.up.ac.za/academic/libarts/afrilang/homelex.html>.

Gilles-Maurice de Schryver, Afrilex Organizer  
([gillesmaurice.deschryver@rug.ac.be](mailto:gillesmaurice.deschryver@rug.ac.be))  
Ghent University, Belgium

### News from Asialex

The Second Asialex International Congress was held at Yonsei University, Seoul, Korea, from 8 to 10 August, 2001. Fifty-six papers by scholars and lexicographers from sixteen different countries, including South Africa and Kuwait, were presented on such topics as dictionary use, compilation of mono- and bi-lingual dictionaries, computational lexicography, problems in entry forms, corpus use, spoken corpus, dictionary and language teaching, and Korean dictionaries. Languages used were Korean and English, with the latter predominating. Patrick Hanks delivered a keynote speech titled 'The probable and the possible: lexicography in the age of the Internet'. Minoru Murata of Chiba University, Japan, and Zhang Yihua of Guangdong University of Foreign Studies, China, were invited plenary speakers. Sangsup Lee, President of the Asialex, gave the opening speech and, at the end of the first day of the congress, hosted a reception dinner for all the participants. At the Asialex business meeting, held on the second day, Professor Minoru Murata was elected next president. He invited all the participants to the Third Asialex International Congress to be held in Tokyo in 2003. The whole congress programme ended with foreign participants touring historic Seoul. The handsomely bound 360-page *ASIALEX 2001*

*Proceedings*, containing both English and Korean papers, is available upon request.

Sangsup Lee, Former President of Asialex  
 Professor of English and Director of the Center for  
 Linguistic Informatics Development,  
 Yonsei University

### Conference on lexicography in the Nordic countries, Tórshavn, 2001

The Sixth Conference on Lexicography in the Nordic countries took place in Tórshavn, the capital of the Faeroe Islands, from August 21 to 25, 2001. These conferences are arranged every second year in one of the Nordic countries: next time we will be in Norway where the first conference took place in 1991.

The Tórshavn conference gathered together some 110 participants and about fifty contributions. The conference was efficiently and hospitably arranged by the Department of Faroese at the University of the Faroe Islands in cooperation with The Nordic Language Council (“Nordisk språkråd”) and the Nordic Association of Lexicography (NFL). The papers were presented in Danish, Norwegian, and Swedish.

There was no common theme of the conference and the contributions ranged over many different topics. However, a few main areas could be mentioned: Internet dictionaries (as well as “dictionaries of the future”), historical dictionaries and their special problems (especially the big Nordic historical dictionaries), constructions and phraseology, bilingual dictionaries, and Faroese lexicography (unique in relation to the number of inhabitants!). Other presentations touched upon subjects like the Danish wordbase for language technology applications, and a Nordic network for coordination of computational lexicons. A number of papers discussed some central topics of lexicography, such as the relation between lexicography and linguistics, the connection between corpus and dictionary, dictionaries and spoken language, gender perspectives in dictionaries, comparative forms of adjectives, etymology, polysemy and cognitive semantics, value statements in dictionaries, a young people’s dictionary, etc.

There were three plenary lectures. Jón Hilmar Jónsson from Iceland discussed phraseology in dictionaries and stressed the unclear connection between phrases and the headwords under which they appear. He suggested a phraseologic index, especially in onomasiological dictionaries where the headwords are related to semantic areas. Jógvan í Lon Jacobsen from the Faroe Islands discussed the relation between lexicography and terminology, two sides of the same coin in Faroese tradition. The lecture brought up a theme that was touched upon in other papers as well, namely the relation between descriptive and prescriptive lexicography. Ebba Hjorth, chief editor of the big Danish dictionary project *Den Danske Ordbog*, summarized her experience from ten years of dictionary work that will be finished next year. She discussed – among other things – the possibility of combining description, guidance and prescription in one dictionary, the size of the corpus (40 million words) and the treatment of spoken

language. She also touched upon difficult practical questions like attracting and keeping together a qualified editorial staff and keeping an eye on time and money. A conference volume will be published with brief summaries in English.

Martin Gellerstam (gellerstam@svenska.gu.se)  
Föreståndare för Språkbanken  
Göteborgs Universitet  
Institutionen för Svenska Språket  
Box 200, 405 30 Göteborg, Sweden

## **Europhras**

Europäische Gesellschaft für Phraseologie  
European Society for Phraseology  
Société Européenne de Phraséologie

Information September 2001

The European Society for Phraseology (EUROPHRAS) with its seat in Zurich was founded in Bielefeld in January 1999.

The Society conceives the notion of phraseology in its broadest sense. Attempting to cover a wide range of research perspectives and methods, it advocates structural and discursive analyses from language-specific, contrastive, synchronic and diachronic viewpoints and includes lexicographical considerations, corpus-based approaches, and analyses in the fields of language acquisition and cognitive linguistics.

The Society's first conference (and general meeting) took place in Uppsala, Sweden, from 15 to 18 June 2000. The second conference (and general meeting) takes place in Loccum, Germany, from 2 to 5 June 2002. The first official conference was followed by a number of smaller conferences and workshops (Münster, Germany; Ascona, Switzerland; Stettin, Poland) organized by members of the society.

Two annual bulletins have been published containing, among other things, information on research projects.

The Society has approximately 200 members from Europe as well as a number of non-European countries.

Further information on the Society (statutes, publications by members, publication series, application for membership, executive committee, advisory board, forthcoming conferences) can be found on our homepage: <http://www.europhras.unizh.ch/>.

“Phraseologie” wird innerhalb der Gesellschaft im weitesten Sinne verstanden. Die struktur- und textbezogenen Gebiete der Phraseologie in synchroner und diachroner, einzelsprachlicher und kontrastiver Perspektive sollen ebenso

zur Geltung kommen wie phraseologische Aspekte der Lexikographie, der Korpus-Linguistik, der Spracherwerbsforschung und der kognitiven Linguistik.

Die erste Tagung (und Mitgliederversammlung) fand vom 15.–18. Juni 2000 in Uppsala, Schweden, statt. Die zweite Tagung (und Mitgliederversammlung) wird vom 2.–bis 5. Juni 2002 in Lokkum, Deutschland, stattfinden.

Es sind zwei Jahresbulletins erschienen (u.a. mit Informationen über laufende und geplante Projekte).

Die Gesellschaft hat inzwischen gegen 200 Mitglieder aus Europa sowie einigen aussereuropäischen Ländern.

Nähere Informationen über die Gesellschaft (die Ziele der Gesellschaft, die Statuten, Publikationen der Mitglieder, Möglichkeit der Mitgliedschaft, den Vorstand und Beirat sowie bevorstehende Tagungen finden Sie auf der Homepage: <http://www.euophras.unizh.ch/>.

Prof. Dr. Harald Burger, Chairman, Europhras  
([Burger.Harald@access.unizh.ch](mailto:Burger.Harald@access.unizh.ch))  
Deutsches Seminar der Universität Zürich  
Schönberggasse 9  
CH-8001 Zürich, Switzerland

### Bibliography of Phraseology

Europhras (see the preceding item) and EURALEX have been collaborating on a Bibliography of Phraseology, to be found on the EURALEX web site <http://www.ims.uni-stuttgart.de/euralex/bibweb>. The Bibliography is updated periodically: a new version was loaded in September 2001. We are heavily reliant on contributions from people working within the field, so please send additions, corrections, and suggestions to me, at the email address below.

Rosamund Moon ([rosamund@moon.eu.org](mailto:rosamund@moon.eu.org))

### Two Workshops on Collocation

Collocation studies are to all appearances a thriving sector, particularly in computational and corpus lexicography. Following the successful workshop organized at the 22. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft in Marburg in March 2000, two workshops have been organized in France this year looking at recent collocational work in NLP and Computational linguistics.

The first was held in Paris on the 13th January 2001 as one of the one-day workshops organized by ATALA, the French association for natural language processing. This meeting was organized by Béatrice Daille (IRIN, Université de Nantes) and Geoffrey Williams (CRELLIC, Université de Bretagne Sud). There were seven speakers in all. After an introduction to the background to collocation studies from linguistic and computational points of view by the organizers,

the invited speaker, Leo Wanner (University of Stuttgart) spoke on the identification of collocations in corpora using a semantic approach based on lexical functions, 'Sur l'identification des collocations en corpus: une approche sémantique'. This was followed by a system for the formal encoding of FLs by Sylvain Kahane (Université Paris 7) and Alain Polguère (Université de Montréal), 'Un langage formel d'encodage des fonctions lexicales et son application à la modélisation des collocations'. Three papers presented a conceptualist approach to collocation: 'Combinaison lexicales spécialisées. Regroupement des mots clés par classes conceptuelles', Marie-Claude L'Homme (Université de Montréal), 'Structuration thématique d'un réseau de collocations' by Olivier Ferret et Brigitte Grau (LIMSI-CNRS), and 'Cartographie de textes: une nouvelle approche pour l'exploration de grands corpora' presented by Isabelle Debourges (LIFO, Université d'Orléans). The previous papers looked at more theoretical aspects of collocation from an NLP perspective. The paper presented by Petra Ludewig (Institut für Semantische Informationsverarbeitung, Universität d'Osnabrück) 'Logo Tax – franchir le fossé entre dictionnaire et texte' took an entirely practical viewpoint with a demonstration of a tool, Logo Tax, that assists students learning German to master collocation. This tool can be visited online at: <http://cato.cl-ki.uni-osnabrueck.de/~logotax/>. The workshop was highly successful with an attendance of over sixty. The programme and abstracts can be consulted on the ATALA website (<http://www.atala.org/je/je-010113.html>).

The second workshop, again organized by Béatrice Daille and Geoffrey Williams, was held under the auspices of the 39th Annual Meeting of the Association for Computational Linguistics and 10th Conference of the European Chapter of ACL. This took place in Toulouse, France, on Saturday 7th July 2001. The invited speaker, Thierry Fontenelle, EURALEX, both opened and closed the workshop, starting with his paper on 'Collocational modelling: from lexical functions to frame semantics', and finishing the afternoon with an impromptu demonstration of an electronic bilingual collocational dictionary. Other papers were very varied, ranging from the purely computational as Sylvain Kahane's 'Formal foundations of lexical functions' to the corpus linguistic with Ton van der Wouden's paper 'Collocation behaviour in non content words'. A number of papers dealt with different aspects of the extraction issue, notably those by Gaël Dias, Heiki-Jaan Kaalep, and Kadri Muischnek, University of Beira, Portugal, University of Tartu, Estonia ('Automatic extraction of verb phrases from annotated corpora: a linguistic evaluation for Estonian'), Don Blaheta and Mark Johnson, Brown University, USA ('Unsupervised learning of multi-word verbs'), and Jean-Philippe Goldman, Luka Nerima, and Eric Wehrli, LATL/University of Geneva, Switzerland ('Collocations extraction using a syntactic parser'). Brigitte Krenn and Stefan Evert, OFAI/Vienna, Austria, and MS/University of Stuttgart, Germany, discussed the statistical basis for collocation extraction in a paper entitled 'Can we do better than frequency? A case study on extracting PP-verb collocations'. The paper by Janyce Wiebe, Theresa Wilson, and Matthew Bell, University of Pittsburg, USA, consisted of ongoing work into an unusual application of collocation 'Identifying collocations for recognizing opinions'.

A practical demonstration of a workbench for lexicographers, Word Sketch, was presented by Adam Kilgarriff and David Tugwell, ITRI/University of Brighton, UK. Such a tool would take much of the drudgery out of corpus-based lexicographical practice with automatic isolation of significant patterns. Thierry Fontenelle's practical demonstrations replaced the final speaker, who was unable to attend, and brought the date to a close with a great deal of interesting and fruitful discussion as to the practical aspects of collocation in lexicographical terms. The workshop organizers are grateful to the programme committee for their behind-the-scenes work in selecting texts and are especially grateful to Thierry Fontenelle for his major contribution to this event. The workshop was well attended and covered a number of languages, English, Dutch, German, and Estonian. It is hoped in future workshops to include work on an even wider range, including non-European languages. The programme can be consulted on the IRIN website, [http://www.sciences.univ-nantes.fr/info/recherche/Theme\\_TALN/index.html](http://www.sciences.univ-nantes.fr/info/recherche/Theme_TALN/index.html). The proceedings are available from ACL (<http://www.aclweb.org/>). The success of these workshops demonstrates the vitality of interest in collocation as a phenomenon in natural language. It is hoped that the impetus will be maintained with further meetings on this theme.

Geoffrey Williams (Geoffrey.Williams@univ-ubs.fr)  
 Département Langues Etrangères Appliquées  
 U.F.R. Lettres et Sciences Humaines  
 4, rue Jean Zay  
 B.P. 92116  
 56321 LORIENT Cedex  
 France  
 Web site: <http://www.univ-ubs.fr/crellic>

### **SENSEVAL-2: Second International Workshop on Evaluating Word Sense Disambiguation Systems**

An ACL SIGLEX Event: Also supported by EURALEX, ELSNET, EPSRC (grant GR/RO2337/01), and ELRA.

There are now many computer programs for automatically determining the sense of a word in context (Word Sense Disambiguation or WSD). The purpose of SENSEVAL is to evaluate the strengths and weaknesses of such programs with respect to different words, languages and tasks. The first SENSEVAL took place in the summer of 1998 for English, French, and Italian with 25 participating systems. The second SENSEVAL, the SENSEVAL-2 Workshop, was held on 5–6 July 2001 in Toulouse.

SENSEVAL-2 evaluated WSD systems on three types of task over 12 languages. In the “all-words” task, the evaluation is on almost all of the content words in a sample of texts. In the “lexical sample” task, first we sample the lexicon, then we find instances in context of the sample words and the evaluation is on those instances only. In the “translation task” (Japanese only), senses

corresponded to distinct translations of a word into another language. Some words for several tasks have been selected to be “translation-equivalents” to some English words to perform further experiments after the official competition.

Fourteen tasks have been designed:

All-words (Czech, Dutch, English, Estonian)

Lexical sample (Basque, Chinese, Danish, English, Italian, Japanese, Korean, Spanish, Swedish)

Translation (Japanese)

About 35 teams participated, submitting 94 systems. Some teams participated in several tasks allowing the analysis of the performance across tasks and languages. All of the results of the evaluation and data is now in the public domain including: results (system scores and graphs), data (system answers, scores, training and testing corpora, scoring software), system descriptions, task descriptions and scoring criteria.

We will publish proceedings of the workshop later this year (available through the ACL). A closely related event is the special issue of the *Natural Language Engineering* journal devoted to the evaluation of word sense disambiguation systems with a special emphasis on results and experiences gained through the SENSEVAL-2 competition from the point of view of the WSD task designer and system developer.

More information at <http://www.sle.sharp.co.uk/senseval2/>.

German Rigau

Co-organizer of the Spanish Lexical Sample task

TALP Research Center, UPC

### First Lexicom Workshop: [lexicom@itri](mailto:lexicom@itri) 2001

“Learning how to” rather than “learning about” was the keynote of the first Lexicom Workshop on Lexicography and Lexical Computing which took place at the Information Technology Research Institute, University of Brighton, England, from July 16–20, 2001. The focus of the tuition and practical work was twofold: the making of corpus-based dictionaries and the software that supports this activity.

### Participants

Fifty participants (see them at <http://www.itri.brighton.ac.uk/lexicom/photo>) from 21 countries spent a week of intensive training, partly in seminars and partly in hands-on practice in the Computer Suite, where everyone had their own machine and could choose to work on the practical tasks either individually or in informal small groups.

Members of the first Lexicom came from many different backgrounds: publishing houses (managers, senior editors and junior editors), universities



(linguists, computational linguists, computer scientists, academic lexicographers and graduate students), software houses (software engineers and computational linguists), government agencies (terminologists and translators) and various types of research institution. Their comments on the Workshop may be found on the website.

### Programme

The programme is to be found in detail on the website (see above). Each seminar introduced and discussed a key topic, which was then further explored in practical exercises at the computer. As well as the basics of practical lexicography, the programme covered corpus design and annotation, the extraction of information from corpus data to make a dictionary entry, a brief look at various types of dictionary databases, and an introduction to frame semantics as a practical approach to corpus analysis. Participants received complete documentation of all the seminars in a bound copy of the Course Notes. One session consisted of presentations by six participants of their own software already in use – a rewarding experience for us all. And in the closing session, several groups presented dictionary entries they had compiled, initiating discussions of various theoretical and methodological issues.

The course was led by Sue Atkins, Adam Kilgarriff and Michael Rundell, the partners in the Lexicography MasterClass (<http://www.lexmasterclass.com/>). It was hard work but it was great fun, and we all learned a lot.

Brighton proved an ideal setting for spare time activities, from walks along the beach to an evening's clubbing for those with real stamina.

**Lexicom 2002** This will be held next summer. The programme and general approach for 2002 will be substantially the same, with a few enhancements suggested by this year's group. Pre-registration is already taking place on the website.

### New Master's degree in lexical computing and lexicography

The University of Brighton is offering a master's degree (MSc) in Lexical Computing and Lexicography, with its first intake in the autumn of 2002. All the topics introduced in Lexicom, and others such as linguistics and the lexicon, translation and terminology, will be treated in week-long modules. As well as students following the full Master's course, those interested in one or more individual modules will also be welcomed. Details of the degree programme, and how to apply, will be posted shortly, or email [Adam.Kilgarriff@itri.brighton.ac.uk](mailto:Adam.Kilgarriff@itri.brighton.ac.uk).

Sue Atkins

<http://www.itri.brighton.ac.uk/lexicom>

### New publication

Sidney I. Landau (2001) *Dictionaries: The Art and Craft of Lexicography*, 2nd edition. Cambridge: Cambridge University Press. ISBN: 052178512X.

## Forthcoming events

### 2002

#### May

29–June 2, Las Pas, Canary Islands: Third International Conference on Language Resources and Evaluation (LREC). Main Conference: 29–31 May, 2002. Workshops: 27–28 May and 1–2 June 2002.

Web site: <http://www.lrec-conf.org/>

#### June

2–5, Loccum, Germany: Europhras Conference on Phraseology. Information: Prof. Dr. Jan Wirrer, Fakultät für Linguistik und Literaturwissenschaft, Universität Bielefeld, D-33501 Beilefeld, Germany.

E-mail: [jan.wirrer@uni-bielefeld.de](mailto:jan.wirrer@uni-bielefeld.de). Web site: <http://www.europhras.unizh.ch/>

#### July

8–10, Grahamstown, South Africa: 7th International Conference of Afrilex. Info: Kathy Kavanagh ([k.kavanagh@ru.ac.za](mailto:k.kavanagh@ru.ac.za)).

Web site: <http://www.up.ac.za/academic/libarts/afrilang/homelex.html>.

15–17, Leicester, GB: Conference on Historical Lexicography and Lexicology.

Web site: <http://www.le.ac.uk/ee/jmc21/hll.htm>

#### August

13–17, Copenhagen, Denmark: Tenth EURALEX Congress. Information: Congress Organizers EURALEX 2002, Anna Braasch, Center for Sprogteknologi (CST), Njalsgade 80, DK-2300 Copenhagen S, Denmark.

E-mail: [elx2002@cst.ku.dk](mailto:elx2002@cst.ku.dk). Web site: <http://www.cst.ku.dk/elx2002>

#### December

16–21, Singapore: 13th World Conference of Applied Linguistics (AILA 2002). Information: AILA 2002 Conference Secretariat, c/o Conference & Travel Management Associates Pte Ltd, 425A Race Course Road, Singapore 218671.

Tel: +65 299 8992. Fax: +65 299 8983. Web site: <http://www.aila2002.org/>