The EURALEX Newsletter

This quarterly Newsletter is intended to include not only official announcements but also news on EURALEX members, their publications, career moves, and (it is hoped) their opinions. Please try to support this by sending newsletter contributions to Carla Marello at the above address. 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.

Verbatim

Verbatim Language Quarterly is resuming publication, with Erin McKean as editor, and the first issue is to appear in Fall/Winter 1998. Subscription, submissions, and back issues information can be obtained by e-mailing Erin McKean at mac@gerharz.com or editor@verbatimmag.com. Information can also be obtained by writing to Verbatim, 4907 N. Washtenaw Ave, Chicago, IL 60625, USA: tel. +1 773–878–7737. Verbatim has been purchased by Dr. Warren Gilson, from Madison, Wisconsin, and he is hoping to operate it as a literary non-profit corporation.

Erin McKean

Lexicographical Work in Progress

TRACTOR and the TELRI project

TRACTOR is the TELRI Research Archive of Computational Tools and Resources. TELRI is the Trans-European Language Resources Infrastructure,
a European Commission project (1995–7). TELRI aims to create a viable infrastructure between leading European language and language technology centres and to provide a platform for industry, research institutes, and universities to supply the NLP community with precompetitive/public domain monolingual and multilingual language resources. TELRI is particularly interested in encouraging minority languages or languages of limited diffusion, in providing language data and applicable software for multilingual applications, and in offering an open and living archive where everyone can deposit resources and everyone can use them for research purposes. The data and approaches required by the new generation of sophisticated multilingual sense-processing technology do not yet exist. TELRI is creating them through the TRACTOR project.

The first task of TRACTOR, mainly aimed at data acquisition, has just come to an end. This has involved negotiations with data suppliers, seeking permissions, legal copyright procedures, and a little financial bargaining. More than 2 gigabytes of resources (estimated to yield well over 100 million words) have been acquired so far, including monolingual and multilingual data (aligned, not aligned, translated, and parallel) in over 20 languages. Much of the data is of recent vintage, but there is a good range of historical material as well. The resources are currently held on a Solaris server at IDS-Mannheim.

Data-types are primarily from written sources, including newspapers, textbooks, dictionaries, literary works, grammars, academic, scientific and technical texts, popular fiction, poetry, jurisprudence, government and military records, and so on. But there is also some transcribed speech and a quantity of digitized sound files. Computational tools and resources present an even greater variety: de-asciifiers, spelling correctors, morphological analyzers and generators, concordancers, compound-recognizers, foreign-word-recognizers, general corpus retrieval and analytical tools, taggers, parsers, segmenters, editors, character-combination frequency lists, etc.

The second task of TRACTOR is already well under way. This consists of converting the data into the formats required by the available tools, finalizing documentation of data and tools, designing the web interface, and making the archive available via WWW, subject to access restrictions placed by resource providers. This will enable a wide range of user groups to benefit: language teachers, theoretical linguists, translators, lexicographers, terminologists, etc, as well as the NLP community itself, corpus creators and managers, software designers and programmers, corpus methodologists and evaluators, and so on.

Ramesh Krishnamurthy, Project Manager of TRACTOR for TELRI
Corpus Research, Department of English,
University of Birmingham, BIRMINGHAM B15 2TT, United Kingdom
E-mail: ramesh@clg.bham.ac.uk

Electronic dictionaries of African languages

The Linguistic Data Consortium (LDC) of the University of Pennsylvania, Philadelphia, is currently financing two electronic dictionary projects, among many others being planned, on African languages. The LDC is writing
programs that will enable these two dictionaries to go on-line in future, and they will be searchable.

1. An Electronic Multilingual Dictionary of the Manding Languages of West Africa (Bambara-Maninka-Jula-Mawu-English). The database will be multilingual, and it will be possible to extract a series of bilingual dictionaries at a later date from it. The dictionary is being developed by Dr Moussa Bamba, a linguist who speaks all the languages fluently. The project began in September 1996, and will probably continue into 1999.

2. An Electronic Bilingual Dictionary of Contemporary Yoruba, another West African language: this is being developed by Dr Yiwola Awoyale. For now, the database is Yoruba-English, but it will also be reworked into English-Yoruba later. The Yoruba dictionary will consist of about 100,000 entries. This project also began in September 1996, and will continue into 1999.

Moussa Bamba (e-mail: mbamba@unagi.cis.upenn.edu)
Yiwola Awoyale (e-mail: awoyale@unagi.cis.upenn.edu)
University of Pennsylvania, PA, USA

Dictionary Research Centre, University of Exeter

There is now a Web site for DRC. The address is: http://www.exeter.ac.uk/drc.

RINT


New Dictionaries and Reference Books

Italian pupils: 6,000 headwords, with games and exercises interspersed amongst the definitions. Word-division and morphological information is provided for each headword, and there are 800 illustrations and 25 thematic tables, together with thematic illustrations for each letter. Torino, Paravia.


Robert K. Headley, Jr., Rath Chim, and OK Soeum (1998) Modern Cambodian Dictionary. This is the most up-to-date Cambodian-English dictionary available and at 1712 pages and with over 60,000 entries, one of the largest Cambodian dictionaries in any language. It has been compiled with an emphasis on current words and expressions used in written and spoken Cambodian; however, there are also many entries for literary and poetic forms so that the dictionary may also be used to read classic Cambodian texts. All of the main entries are provided with a phonetic pronunciation guide, and there are numerous examples. At the end of the dictionary, there are labeled drawings of important items in Cambodian material culture such as the loom. Dunwoody Press.

Kyrgyz-English Dictionary (1998). This dictionary of 12,465 entries is a partial description of the Kyrgyz lexicon based on a 275,493 word corpus taken from recent newspapers. The main feature of this dictionary is the inclusion of more than 16,500 example phrases and sentences. Approximately 10% of the entries are not found in other Kyrgyz dictionaries. Dunwoody Press.

Li Zhuqing (1998) Fuzhou-English Dictionary. This dictionary contains more than 7,100 entries for colloquial words and phrases in current use in the Fuzhou dialect of China. It does not include those non-colloquial words that share the same Chinese characters with the official language, Mandarin Chinese. The dictionary is based in part on Fuzhou Fangyan Cidian (1995) by Li Rulong et al. It has been revised by the author's fieldwork, in addition to other resources. The body of the dictionary is arranged by total strokes of the head character. Pronunciation is given in IPA. An entry index arranged by the Fuzhou romanization is included along with a pinyin index of head characters. Dunwoody Press.

D. Mantzel and B. Schulz, the Dictionary Unit for South African English (1997) Francolin Illustrated Primary School Dictionary for Southern Africa. This dictionary is targeted particularly at learners aged 10–12 from a rural or 'township' background, to assist in the acquisition of English in areas where the language is seldom heard. 3,441 headwords are defined using a simple, restricted vocabulary: there are illustrative sentences and over 300 line drawings. The dictionary covers basic general English vocabulary and includes 800
curriculum words and a number of specifically South African English terms. Introductory sections help the teachers use the dictionary in the classroom, and appendices list numbers, contracted forms, days and months, shapes, colours, and abbreviations of measurement. Cape Town: Francolin.


_Il CdRom del Castiglioni Mariotti Vocabolario della Lingua Latina_ (1998). Each of the 50,000 entries of the printed edition can be looked up, including any inflected forms. Searches can be made of the full text, etymologies, translations into Italian, or authors quoted, or of combinations of fields. Torino: Loescher.

**Other New Publications**


Recent volumes in the series ‘Strumenti di lessicografia letteraria italiana’ (Firenze: Olschki) include:


**Forthcoming events**

1998

August

4–8, Liège, Belgium: EURALEX '98. Info: Congress Organizers EURALEX '98, University of Liège, Department of English Language and Linguistics, Building A2, 3, Place Cockerill, B-4000 Liège, Belgium. Tel: +32 4 3665360
10–14, Montréal, Canada: COLING/ACL-98, 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics. E-mail: coling-ac198@iro.umontreal.ca. Web site: http://coling-ac198.iro.umontreal.ca

15, Montréal, Canada: Computer '98, the first Workshop on Computational Terminology. This will immediately follow COLING/ACL-98, and provide a forum to bring together researchers from the fields of computational linguistics, terminology, automated translation, information retrieval and lexicography who share an interest in computational aspects of terminology processing. E-mail: Didier.Bourigault@lli.univ-paris13.fr or Christian.Jacquemin@iut-nantes.univ-nantes.fr or lhommem@ere.umontreal.ca. Web site: http://tornade.ERE.UMontreal.CA:80/~lhommem/coring/computerm.html

17–21, Vienna, Austria: 15th Terminology Summer School, Terminology Management in Specialized Environments. Info: TermNet, Simmeringer Hauptstr. 24, A-1110 Vienna, Austria. E-mail: termnet@termnet.at.

17–28, Saarbrücken, Germany: 10th European Summer School in Logic, Language and Information. Includes a one-week course on ‘Word Sense Disambiguation’, and other workshops and courses related to the lexicon. Info: DFKI, ESSLII X, Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany. Phone: +49 681 302 4933. Fax: +49 681 302 4929. E-mail: esslli@coli.uni-sb.de. Web site: http://www.coli.uni-sb.de/esslli

26–30, St. Andrews, Scotland: 31st Annual Meeting of the Societas Linguistica Europaea. Info: Dr. C. Beedham, Dept. of German, School of Modern Languages, The University, Buchanan Blg., St. Andrews, Fife KY16 9PH, Scotland. E-mail: cbl@st-andrews.ac.uk

September

2–4, Herstmonceux Castle, Sussex, England: Workshop, ‘Word Sense Disambiguation Evaluation Exercise’, sponsored by EURALEX/ACL. There will be specialized groups working on French and Italian (ROMANSEVAL), as well as on English. Info: Adam Kilgarriff, SENSEVAL, ITRI, University of Brighton, Lewes Road, Brighton BN2 4GJ, England. E-mail: senseval-coord@itri.bton.ac.uk. Info concerning ROMANSEVAL: Jean. Veronis@lpl.univ-aix.fr. Web site: http://www.itri.bton.ac.uk/events/senseval/euralex-cfp.txt

9–12, Leuven, Belgium: EUROCALL 1998. E-mail: eurocall@hull.ac.uk. Web site: http://www.arts.kuleuven.ac.be/euurocall98


University of Reading, Whiteknights, P.O. Box 218, Reading RG6 6AA, England. E-mail:p.e.kerswill@reading.ac.uk
21-23, Chur, Switzerland: Minority Languages in Context: Diversity and Standardisation. Info: Lorenza Mandada, Romanische Seminar, University of Basle, Stapfelberg 7/9, CH-4051 Basle, Switzerland. E-mail: mandada@ub-aclu.unibas.ch.
23-26, Brno, Czech Republic: A Workshop on Text, Speech and Dialog (TSD’98). Principal topics fall within the field of natural language processing, in particular Corpora, texts and transcription; Speech analysis, recognition and synthesis; and their intertwining within NL dialog systems. Info: Dana Komarkova, Faculty of Informatics, Masaryk University, Botanicka 68a, 60200 Brno, Czech Republic. Tel: +420 5 41 512 359. E-mail: dkomar@fi.muni.cz. Web site: http://www.fi.muni.cz/tsd98/
25-27, Edinburgh, Scotland: Endangered Languages – What Role for the Specialist? Info: Nicholas Ostler, e-mail:nostler@chibcha.co.uk. Web site: http://www.bris.ac.uk./Depts/Philosophy/CTLL/FEL

October
2-3, Savonlinna, Finland: Interdisciplinary Colloquium on Translation and Cognition. Contact: Riitta Jääskeläinen, University of Joensuu, Savonlinna School of Translation Studies, P.O. Box 48, 57101 Savonlinna, Finland. E-mail: riitta.jaaskelainen@joensuu.fi
8-10, Milano, Italy: IX International Congress of Linguists. Info: Prof. Giancarlo Bolognesi, Istituto di Glottologia, Università Cattolica, Largo Gemelli 1, 20123 Milano, Italy. E-mail:listglott@mi.unicatt.it
29-31, Budapest, Hungary: XXXII Congresso internazionale di Studi della Società di Linguistica Italiana, ‘Semantica e Lessicologia storiche’. Info:
Prof. G.P. Salvi, Olasz Nyelv es Irodalom Tanzek, Muzeum korut 4, Building C second floor, Budapest, Hungary. Tel./fax: +36 2667284. E-mail: salvi@isis.elte.hu.

November

26–27, Veldhoven, The Netherlands: 3rd International Conference on Maintenance and Loss of Minority Languages. Info: Conference Secretariat, Tilburg University, Research Group on Languages and Minorities, P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands. E-mail: language.loss@kub.nl.

1999
January

April
14–17, Newcastle upon Tyne, England: 2nd International Symposium on Bilingualism. Info: Mrs Gillian Cavagan, ISB Organising Committee, Department of Speech, University of Newcastle upon Tyne, NE1 7RU, UK. Fax: +44 191 2226518. E-mail: gillian: cavagan@newcastle.ac.uk. Web site: http://www.ncl.ac.uk/~nspeech/isb2.htm

May

August