Kernerman
DICTIONARY News

The Coming Boom in English Lexicography:
Some Thoughts about the World Wide Web
(Part One)

Charles M. Levine

Charles M. Levine was until recently the vice-president and publisher of the Random House Dictionaries. He is currently executive director and CEO of the Flora of North America, a collaborative nationwide project to catalog the plants of the continental US, Canada and Greenland. He has held a number of senior executive positions in digital and print content management in Asia and the USA. charlev@att.net

Not infrequently, people unfamiliar with the dictionary marketplace assume that spell checkers, given away free with computer word-processing programs, have made printed lexicons obsolete. "Why would anyone want a dictionary anymore?" some people chime, glossing over the difference between a spell checker and a dictionary, and oblivious as well to how poor spell checkers usually still are. (Since spell checkers are bundled freely, there is no money to be made and no incentive in developing truly better, more intelligent spell-checking software.)

Even those who are more knowledgeable about the dictionary market, such as sales and marketing people at large publishing houses, often think it is just a matter of time before people stop buying printed dictionaries. After all, one is often reminded, printed, multi-volume general English encyclopedias for all intents and purposes have gone extinct – from nearly 750,000 sets worldwide in 1990, to hardly 100,000 today. How can printed dictionaries be far behind?

According to friend and colleague Joseph Esposito, former CEO of Britannica and the driving force behind the first launches of www.britannica.com: "In talking about the relative size of the encyclopedia market today, you may want to think about two other metrics in addition to unit sales: dollar volume (down by 90% since its peak), and the number of people who actually use an encyclopedia (also down, but hard to measure). Encarta's ubiquity has not resulted in more widespread use. And thereby hangs a tale."

Or, as someone summed up in a Purdue University education course: "How does a company [like Britannica] that survived over 200 years, through the American revolution, the industrial revolution, WWI and WWII, find itself suddenly on the verge of being completely eliminated by three simple letters: www?" (Google search 2000, undated and unattributed.)

In such an environment, no one in their right mind would start up a new consumer encyclopedia company, or a new general-purpose dictionary operation, I have been firmly told.

You could even say that Bloomsbury in England, who created the new Encarta World English dictionaries (a brand new college edition of which was just shipped), are no exception: Microsoft footed the bill for a very specific reason: to avoid paying American Heritage license fees for its dictionary which once graced the Encarta digital reference suite. And, in the end,

Microsoft does not mind giving away for free a large percentage of the Encarta digital dictionaries in promotions to sell its Wintel hardware and software.

But surprisingly, the printed dictionary market in the United States seems to have held steady during this same past decade, while the market for printed encyclopedias collapsed. Each year, about 2 million college-level dictionaries and 2 million mass market pocket dictionaries are sold regularly like clockwork.

The American domestic dictionary market seems as steady and as uneventful as a rock. And, at the same time, Random House, American Heritage and Merriam-Webster have found a new area (for American lexicographers, at least, who came late to the game) in which to grow – the ESL/EFL (English as a Second/Foreign Language) world market.

In fact, ESL/EFL will probably help keep the wolves from the American lexicographers' door, and prevent dictionary operations from going the way of their encyclopedic cousins. English learning is booming around the world: "Today, 350 million people are native English speakers, but by some counts more than a billion speak at least some English as a second language. Most of them are in Asia...

"Now it's not native speakers that are moving English forward,' said Larry E. Smith, an expert on international English at the East-West Center in Honolulu. 'It's the non-native speakers, the people in Singapore, the people in Malaysia.'" (Seth Mydans, *The New York Times,* July 1, 2001.)

Internet use is also driving English use. "More than 80% of the home pages on the Web are in English, while the next greatest, German, has a mere 4.5% and Japanese 3.1%," according to John Simmons of Interbrand, which specializes in valuing brands. (In "Global usage of English speaks volumes in trade," Robin Young, *The Times,* 19 March 2001.)

By calculating all the commerce carried out in various languages, Simmons came up with a rather imaginative and, no doubt, utterly wild estimate of what he calls the "gross language product (GLP)", which for English "is the biggest at £5,455 billion, followed by Japanese at £2,960 billion, German at £1,714 billion and Spanish at £1,249 billion."

This is a wild but highly suggestive estimate, but indeed by every measure there is a boom in English usage and commerce fostered by the World Wide Web, and this seems to be having just the opposite effect on lexicography as it did on "encyclopedio-graphy." The World Wide Web seems to be creating new markets for ESL/EFL printed dictionaries and instruction materials — and for linguists themselves.

In fact, technology companies are competing to hire linguists, despite the downturn in the Net economy: "Suddenly, linguists have their pick of jobs as lexicographers, 'knowledge engineers' and 'vocabulary-resource managers.' For those with doctorates, the typical starting salary is around $60,000, plus some stock. The more highly trained talent is drawing more than $100,000.

"Linguistics experts help e-businesses improve customer service by building so-called natural-language processing systems that can respond meaningfully to requests for help or information. With linguists developing the database or 'lexicon,' a system can distinguish between multiple meanings of words, relate groups of words by concept, and narrow the scope of a search by asking questions of the site visitor." (Daniel Goldin, *The Wall Street Journal,* first read sometime in 2000, but undated on Google, where I retrieved the article.)

I hope in the next installment to talk more about this boom in English lexicography and how it relates to new developments coming to the World Wide Web.

PASSWORD for Hungarians

Tamás Magay

Tamás Magay was born in 1928 in Kaposvár, and graduated in English language and literature from the University of Budapest. He studied lexicography under L. Országh, and wrote his theses on the history of English lexicography in Hungary and the principles of bilingual lexicography. He was co-author and later editor-in-chief of a series of English/Hungarian dictionaries, and edited, among others, a major phraseological bilingual dictionary and a pronunciation dictionary of Hungarian proper names. He is a founding member of Euralex and organized its third congress (Budalex '88), and founded and heads the English Department of the new Gáspár Károli University, including lexicography and lexicology studies. Dr Magay edited the first semi-bilingual dictionary in Hungary (*Magyar Chambers,* 1992) and is the editor of the forthcoming second version.

1. Introduction

Plenty, though by far not enough, has been said and written by outstanding scholars and lexicographers on the subject of bilingualizing the monolingual learner's dictionary. Apart from typological relevance, the problems raised and solved so far have usually been in close relation to one of the most challenging questions of (practical) lexicography: *dictionary use* (Atkins

1998, Cowie 1999). It is this aspect that I would like to stress in the present article: how to proceed from the present state of the art, and further refine and improve this type of dictionary.

Gabriele Stein distinguished three stages of dictionary use in foreign language acquisition, the order being of major importance: (1) the bilingual dictionary, (2) the monolingual learner's dictionary, and (3) the native-

language dictionary (1990, 405). At the same time, I could not agree more with Lionel Kernerman who emphasized that the "vital element in the acquisition of a new language is associated with one's native tongue," (1994), another cardinal aspect.

The scene in Hungary, with Hungarian as one of the minor world languages spoken by 15 million people (about 10m in Hungary, 3.5m in neighbouring countries, and l.5m across the ocean), is simple. While struggling for the preservation of their mother tongue, Hungarians throughout history have always been bound to rely on foreign languages in order to keep in contact with other nations, and to survive and withstand pressure from all sides.

2. Hungarian past

The bilingual dictionary type was dominant in Hungary until the mid-twentieth century, and *a dictionary* was tantamount to a bilingual dictionary (with German as top priority foreign language alongside Latin, and English taking the lead as late as the 1980s-90s). The publication of Országh's monumental *Dictionary of the Hungarian Language* (DHL, 1959-62) turned the tide, and in the minds of the general Hungarian readership *a dictionary* soon began to mean a monolingual dictionary as well.

From the end of World War II until 1990, practically the sole producer of dictionaries, both scholar and commercial, was Akadémiai Kiadó (AK), publisher of the Hungarian Academy of Sciences and lavishly subsidized by the state. The output was enormous. A dictionary *trilogy* was launched, consisting of unabridged/comprehensive, medium/concise and pocket/school dictionaries in the major languages (eg German, English, Russian). Bilinguals in languages from Albanian to Vietnamese also appeared, and specialized, multilingual and technical dictionaries, on a commercial basis too, were turned out by the score.

Scholarly dictionaries, such as Országh's seven-volume DHL, or etymological, synonym, dialect and other high-standard dictionaries (compiled by Országh, S. Eckhardt, L. Hadrovics, L. Gáldi, E. Halász, G. Bárczi, L. Benkő and others) were edited by the Acadamy's Linguistics Institute and also published by AK.

3. European presence

With the disappearance of the Iron Curtain and Communist dictatorship, exchange between East and West in both the economic and cultural spheres, and particularly in language learning and teaching, has gradually been easing. As the choice in consumer products widened, 'western' articles, including a great variety of dictionaries, became more and more available to the general public. Although the use of bilingual dictionaries by learners at all levels has not abated, firstly teachers then students of English began to prefer almost exclusively monolingual dictionaries (ALD, LDOCE, COBUILD, Chambers, etc), consciously or subconsciously inspired by the late

A.S. Hornby who was "firmly committed to the pedagogical principle that English should be learned through the medium of English" (Marello 1998, 292). Hornby was likewise unwilling to accept that the learner's mother tongue "could be used for the initial presentation of meaning" (Marello ibid). Hornby's principles were readily shared in Hungary by a significant section of the English teaching profession. However, bilingual dictionaries continued to be sold in ever-increasing numbers.

I have tested university students on dictionary use several times in the past few years, and it turned out that bilingual dictionaries were top priority both for the comprehension and production of English texts, and that among the monolingual dictionaries, learner's dictionaries were preferred to a great extent, mostly for their contextual examples which users badly miss in bilingual dictionaries.

Then again, testing students during the past semester (2000/2001) surprised us with the finding that 39% of them discovered in our department library *Magyar Chambers* (MC, AK 1992), the Hungarian semi-bilingual version of *Chambers Concise Usage Dictionary* (CCUD). The answer as to why they liked it was usually its happy mingling (or constellation) of the source-language (English) headword+definition+examples with the Hungarian equivalent.

In this chain of criteria the definition proved the weakest link, which called for another test: what can a student of English make of a dictionary definition when, for instance, reading or translating from English into Hungarian (L2–L1). The following examples may perhaps support my reservations concerning definitions, especially in the notional sphere of words. In the test, students could use any or all three dictionaries referred to below. The figure in square brackets indicates the number of students who were able to find the adequate Hungarian equivalent of the English word, or its relevant meaning or sense:

(1) **commune** [3 out of 40 students]:
ALD: a group of people, not all of one family, living together and sharing property and responsibilities
CCUD: a group of people living together and sharing everything they own
LDOCE: a group of people who live together and who share the work and their possessions
(2) **communion** [2 out of 40 students]:
ALD: the state of sharing or exchanging the same thoughts or feelings
CCUD: the sharing of thoughts and feeling
LDOCE: a special relationship with someone or something in which you feel that you understand them very well
(3) **dazzle** [8 out of 40 students]:
ALD: to impress sb greatly through beauty, knowledege, skill, etc
CCUD: to affect the ability of making correct judgements
LDOCE: to make someone feel strong admiration

In another test, half of the students were allowed to use a bilingual English-Hungarian dictionary (EHD 1999), and the other half used MC. In the first group, 15 out of 20 found adequate Hungarian equivalents, while all the MC users found the best equivalents. The tests proved statistically what the actual practice was: English students began to prefer using MC because (a) it saved them time, and (b) with Hungarian equivalents in the context of English it was much easier to disambiguate meaning and reference in finding the adequate native-language equivalent.

Another advantage of the bilingualized dictionary is that the user is not forced to step out of the English context, since the text is about 85% *mono*lingual and only 15% *bi*lingual. In Kernerman's words, [the semi-bilingual dictionary] "contains the advantages of the monolingual learner's dictionary, combined with the native tongue translation found in the bilingual dictionary. The ambiguity of the bilingual dictionary is thus automatically eliminated. Learners are encouraged to read the definitions and examples of usage in English, since only the headwords [with their various senses] are translated" (1994).

4. Devil in details

If we look at a dictionary entry under a magnifying glass, discrepancies, mistakes, omissions and inconsistencies become obvious. In the second part of this article we deal with such shortcomings concerning both the macrostructure and the microstructure found in the course of revising and updating the original edition of MC. The new edition will be published as *Password for Hungarians* (PH) by Nemzeti Tankönyvkiadó (NTK, National Textbook Publishing House, Budapest).

(a) *Entry structure and headword arrangement*

Roughly speaking, there are two widely different ways of arranging entry words (headwords) and their derivatives, collocations and idioms: the cluster-type (or etymological) arrangement, and the strictly alphabetical arrangement wherein lexemes and their derivatives figure as separate entries. The first is more convenient for writing (or encoding), the other for reading (decoding) (Cowie 2000). There are, of course, possibilities and cases of combining the two methods. CCUD, which forms the basis for MC and various other adaptations known as *Password*, employs the cluster-type arrangement.

If one of the aims of the dictionary look-up process is quick accessibility, then the *Password* arrangement is inadqate. If, however, there is more to it, eg the pedagogical aspect, then it will win over generations of users, because it will help "grasp connections of meaning and form across the entries" (Cowie ibid). The etymologically-based cluster-type structure will enhance the lexicological awareness of the user with its information on derivatives, compounds, phrasal verbs and other idioms in a fixed order. Here is an example for this type of arrangement in MC:

> **heart**
> **-hearted**
> **hearten**
> **heartless**
> **heartlessly**
> **heartlessness**
> **hearts**
> **hearty**
> **heartily**
> **heartiness**   [so far derivatives]
> **heartache**
> **heart attack**
> **heartbeat**
> **heartbreak**   [etc, compounds]
> **at heart**
> **break someone's heart**
> **by heart**
> **from the bottom of one's heart**
> **have a change of heart**   [etc, idioms]

Once the user has taken up the 'rhythm' that is common to all entries (and if he/she is encouraged to 'learn' it from the front matter), the dictionary can be used more efficiently and profitably than with the strictly alphabetical (and perhaps quicker) method, because what happens here is—willy-nilly—vocabulary extension, an essential element of language pedagogy. Of course, this system, as any other system, is not flawless either, and has its weaknesses. Here are some examples:

> If **donor** is found under **donate**, why do **deceit** or **deception** and **deceptive** not come under **deceive**? And why are **defiance** and **defiant** separate entries, and not entered under **defy**? Again, **counter²** adverb, verb and **counter³** noun are separate entries while **counter¹** cross-refers to **count²** where it figures with other derivatives (**countable, counter¹, countless** etc). And why does **easily** come under **ease** and not under **easy**? Such inconsistencies are confusing.

(b) *Pronunciation and stress marking*

The phonetic notation in CCUD has become out of date and needed updating. By now the leading monolingual learner's dictionaries have come to terms, as it were, concerning the symbols used. While the new PH does indeed carry this out, there is a systemic flaw in the Chambers-based core which would call for major reshaping, namely to be somewhat more generous with *full* pronunciations given to derivatives, and to be more explicit (or less laconic) with *partial* pronunciations. A few examples will make this point clearer.

> **adaptation** has only [a–], **completion** [–ʃən.], **inferiority** [–'o–]; **adaptor, adequacy, dicey** or **hearty** receive no pronunciation; **apprenticeship** has no pronunciation, whereas it cannot be inferred from the pronunciation given to **apprentice**. Similarly, the pronunciation of **derisive** cannot be inferred from that of **derision**, nor can **malicious** be predicted from **malice**.

In the cases of zero pronunciation, the dictionary consistently marks the stress in bold headwords or sub-headwords. However, it might be considered an inconsistency that whereas pronunciations, full or partial, have only main stresses, sub-headwords (usually derivatives) receive secondary stress as well. For example, whereas **intermittent** has pronunciation with a main stress, **intermittently** is given without pronunciation but with both secondary and main stresses, as do **encyclop(a)edia** and **encyclop(a)edic**.

(c) *Run-on entries vs the 'one only' equivalent*
In the former edition, undefined run-on entries (mostly derivatives of main headwords) did not get a Hungarian translation. The reason for this was not sheer neglect but intentional. Since lexemes, ie entry words, often have more than one meaning, and so do most of their derivatives, how could we have found one equivalent only for an entry word containing, say, six senses, as with **circulation**, or seven senses, as with **closeness**, or the two widely distinct meanings of **engagement** or **collaborator**? The translator of the new edition could do nothing but give a single translation for the 'most important'(?) meaning, and perhaps one or two more, separated by a semicolon, to 'rhyme' with the semantic structure of the entry.

(d) *Peripheral entries and end matter*
CCUD, and thus MC, omitted important cultural, linguistic and geographical elements. This is provided for in the new PH, along with a more representative coverage of the commonest abbreviations and acronyms in its headword-list. Other innovations that will make PH more useful and colourful include full-page pictorial illustrations, tables of weights and measures, common irregular verbs with a brief grammar; and geographical names listing country names, their adjectives and pronunciation. A major addition of PH is a Hungarian-English index, to facilitate the two-way use of the dictionary and enhance the process of encoding (L1–L2 use), making this dictionary more user-friendly than its predecessor.

5. Conclusion
The relationship between pedagogical lexicography and dictionary use is treated here in terms of the bilingualized learner's dictionary. Tests among university students of English in Hungary proved for beginners and more advanced students alike the usefulness and advantages of the semi-bilingual dictionary over the bilingual on the one hand, and the monolingual learner's dictionary on the other. The most striking hindrances, eg the omission of example sentences from the bilingual, and definitions giving rise to disambiguity in the monolingual, are largely eliminated from this type of dictionary. However, a great deal still needs to be improved in the bilingualized adapatations, in particular the CCUD base of this revision, which could do with yet further (even structural) alterations.

References
(a) *dictionaries cited:*
ALD Hornby, A.S., J. Crowther (ed.) 1995. *Oxford Advanced Learner's Dictionary* 5/e. Oxford: OUP.
CCUD Schwarz C.M. and Seaton M.A. (eds.) 1985. *Chambers Concise Usage Dictionary*. Edinburgh: W&R Chambers.
COBUILD Sinclair, John et. al. 1996. *Collins COBUILD Learner's Dictionary* 2/e. London: HarperCollins Publishers.
DHL Országh, L. et al. (eds.) 1959-62. *Dictionary of the Hungarian Language*. Budapest: Akadémiai Kiadó.
EHD Országh, L. et al. (eds.) 1999. *English-Hungarian Dictionary*. Budapest: Akadémiai Kiadó.
LDOCE Procter, P. et al. 1978. *Longman Dictionary of Contemporary English* 3/e. Harlow and London: Longman.
MC Schwarz C.M., Seaton M.A., Magay T. 1992. *Magyar Chambers English Dictionary for Speakers of Hungarian*. Budapest: Akadémiai Kiadó.
PH Magay T. (ed.) 2001. *Password for Hungarians*. Budapest: Nemzeti Tankönyvkiadó. Forthcoming.
(b) *other references:*
Atkins, B.T.S. (ed.) 1998. *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*. Lexicographica, Series Maior 88. Tübingen: Max Niemeyer Verlag.
Battenburg, J.D. 1991. *English Monolingual Learner's Dictionaries. A User-Oriented Study*. Tübingen: Max Niemeyer Verlag.
Cowie, A.P. 1999. *English Dictionaries for Foreign Learners – a History*. Oxford: Clarendon Press.
Cowie, A.P. 2000. 'The EFL Dictionary Pioneers and their Legacies.' In *Kernerman Dictionary News*, 8.
Hartmann, R.R.K. 1992. 'Learner's References: from the Monolingual to the Bilingual Dictionary.' In H. Tommola et al. (eds.) *EURALEX '92 Proceedings*. Tampere: Tampereen Yliopisto.
Kernerman, L. 1994. 'The Advent of the Semi-Bilingual Dictionary.' In *Password News*.
Kernerman, L. 1996. 'English Learners' Dictionaries: How much do we know about their use?' In M. Gellerstam et al. (eds.), *EURALEX '96 Proceedings*. Göteborg: Göteborg University.
Marello, C. 1998. 'Hornby's Bilingualized Dictionaries.' In *International Journal of Lexicography*, 11.4.
Reif, J.A. 1987. 'The Development of a Dictionary Concept: An English Learners' Dictionary and an Exotic Alphabet.' In A.P. Cowie (ed.) *The Dictionary and the Language Learner*. Tübingen: Max Niemeyer Verlag.
Stein, G. 1990. 'From the Bilingual to the Monolingual Dictionary' in T. Magay and J. Zigány (eds.), *BudaLEX '88 Proceedings*. Budapest: Akadémiai Kiadó.

Teaching Lexicography
or
Training Lexicographers?

Michael Rundell

Michael Rundell has been a professional lexicographer since 1980, working on a wide range of English dictionaries. As Managing Editor at Longman Dictionaries (1984-94), he was responsible for running major projects including the *Longman Language Activator* (1993) and *Longman Dictionary of Contemporary English* (2/e 1987 and 3/e 1995). He had a leading role in the design of several lexicographic corpora (including the Longman Lancaster Corpus, Longman Learner Corpus, and British National Corpus) and was a member of the Advisory Committee for the forthcoming American National Corpus. As well as running training courses for Longman lexicographers, in both the UK and US, he has taught for many years on the Interlex and MA programmes at the University of Exeter. He is now a freelance consultant and divides his time between dictionary editing and lexicographic training. He is Editor-in-Chief of the forthcoming *Macmillan English Dictionary*. michael.rundell@dial.pipex.com

In an interesting paper at last year's Euralex Congress, Tamás Magay addressed the theme of teaching lexicography, and raised the issue of "where and how lexicography is taught in European universities" (2000). This is a highly relevant topic for almost any member of Euralex, but – as so often happens when this subject is discussed – no distinction was made between teaching people *about* lexicography, and training people to *be* lexicographers. The debate needs to move into this territory because there appears to be, in some quarters at least, an implicit assumption that the key to better dictionaries is simply more university-based training. To quote Magay: "Once we accept that dictionary-making is now a profession, it has to be taught at the highest level" (p.443). This is merely the latest airing of a view that has been floated at various times over the last 20 years (see for example the collection of papers edited by Robert Ilson in the mid-80s). But is it true?

The question is timely because there is currently a lot of activity in this area. The University of Exeter's Dictionary Research Centre (DRC) – which pioneered courses in lexicography in the UK, and has a well-established MA programme and a popular annual short course – is about to move to a new home at (very appropriately) the University of Birmingham. Meanwhile, the Lexicography MasterClass (LMC, consisting of Sue Atkins, Adam Kilgarriff and myself) is about to launch a one-week workshop at the University of Brighton's Information Technology Research Institute (ITRI), to be followed in 2002 by a new MSc programme in lexicography and lexical computing. Against this background, it seems incumbent on course-providers to be clear about the sort of knowledge and skills they are undertaking to equip prospective students with.

Lexicography is a legitimate academic subject in its own right, not least because of the wide range of "feeder" disciplines – lexical semantics, second-language acquisition theory, and computational linguistics, to name just a few – that supply a theoretical perspective against which lexicographic issues can be intelligently addressed. Consequently, people working in many fields can derive great benefit from learning more about the theoretical, pedagogical, and computational underpinnings of the lexicographic process. This does not, however, turn them into lexicographers. As any practising lexicographer knows, the only real way to learn how to write dictionaries is to write dictionaries. Most good dictionary publishers, whether operating in the commercial or academic sphere, provide on-the-job training for their staff – typically some combination of initial "basic training" followed by ongoing feedback and mentoring over a long period. Against this background, what is the role for courses in lexicography provided by academic institutions?

Perhaps another way of framing this question would be to look at the skills a contemporary lexicographer needs and consider how far these are (or can be) supplied within dictionary publishing houses through the traditional "apprentice" model. Probably the most central task in general lexicography consists of analyzing very large quantities of primary data (from a corpus) and imposing some sort of order on it – or, more accurately, *discerning* the underlying order within it and then describing this in a way that is both useful and relevant to a particular group of dictionary users. The better one understands every stage in this process, the more successful one is likely to be. If you know, for example, where your corpus data comes from, why the corpus is designed in the way it is and how is it linguistically annotated, you are in a better position to conduct sophisticated searches, to perceive patterns and regularities in the mass of data, and to distinguish between what is relevant and what is marginal. Or again, the process of discovering meanings is likely to be more effective when gut instinct is complemented by some understanding of lexical relations or frame semantics. Traditionally, many lexicographers have acquired high levels of competence as data analysts, seemingly by osmosis and without necessarily being able to articulate the criteria that underlie their decision-making. But now,

with so much data to process, lexicographers need all the help they can get, and linguistic theory has an important role to play in informing the judgments that we make.

Dictionary publishers remain the primary suppliers of training in the practical skills of lexicography. But it is not realistic to expect them to deliver the full range of training needs – partly because they will not always have expertise in all of the relevant theoretical disciplines, and partly for straightforward business reasons. Though the best publishers set aside time and resources to train their staff, the normal pressures of deadlines take their toll. Meanwhile, changing patterns of work have reduced publishers' capacity to nurture new editors: large in-house dictionary teams (which provide a supportive learning environment for new lexicographers) are becoming more of a rarity. The arrival of email, intranets and inexpensive high-powered computing has had the effect of dispersing editorial teams, as experienced people increasingly opt to work from home, sometimes for several different publishers. Though this has many benefits, for publishers and employees alike, one of its less positive side-effects is that few publishers can now provide all the necessary training.

In tandem with these changes, we are also seeing (and not before time) the development of closer links between lexicography on the one hand and the natural language processing community on the other. The benefits of collaboration between these two sectors are at last beginning to be recognized, but this in turn requires lexicographers to learn yet more skills. This, then, is the environment in which the LMC has begun to plan new courses. It became clear to us that the best model for an effective training programme would be one that combined a grounding in relevant theoretical subjects with a strong element of hands-on learning of practical skills. Over the past four years or so we have, either collectively or individually, run a number of customized short courses for institutions in various parts of the world. This July, however, we are launching our first *general* programme in Brighton – lexicom@itri – a one-week training workshop.

The term "workshop" is deliberate, because a key feature of the programme is that lecture sessions alternate with periods in the computer lab, where participants will do practical tasks that relate directly to the subject of the previous lecture. There is an analogy here with professional disciplines such as law or medicine, where a good theoretical grounding is a starting point for the development of practical skills. And just as a detailed knowledge of legal precedent cannot in itself turn someone into an effective courtroom advocate, so a familiarity with the minutiae

of metalexicographic theory does not *per se* make someone a good dictionary writer. The practical and theoretical go hand in hand. The three areas we aim to cover in detail are: designing, building and working with a text corpus; creating and using a dictionary database; and, the process of writing actual dictionary entries. Along the way, we will take in issues such as corpus annotation (POS-tagging and other ways of enriching raw text data), smart approaches to data extraction (including some programming skills), and the principles of writing definitions.

The workshop is not aimed at complete beginners: rather, we expect participants to have some grounding in one of the three main disciplines involved (lexicography, linguistics and computer science), so that they can use the course to learn more about the other subjects that have a bearing on their work, and of course to contribute their own perspective in those areas where they are already experienced.

Lexicom promises to be a great learning opportunity for everyone involved – not least the tutors themselves, of course, since the best thing about lexicography is that there is always more to learn. When we planned the course, we decided that we would need at least 15 participants to make it worth doing. In the event, we have had to cap the attendance at 50 and create a waiting list for lexicom 2002. This level of interest suggests there is plenty of demand for courses of this type. We are currently developing, with several other colleagues at Brighton, a new MSc programme in Lexicography and Lexical Computing, which has been approved to start up in October 2002.

To return, finally, to the question we touched on earlier: what skills can participants in the lexicom workshop expect to go away with? It would be unwise for us to claim that we can turn people into lexicographers in a week. A more reasonable objective, however, is to give aspiring dictionary-editors enough grounding to know whether they have the potential to go further, and to enable established lexicographers to "raise their game" through a deeper and broader understanding of the diverse range of factors that contribute to making great dictionaries.

Our other main objective, of course, is that we all enjoy ourselves in the process.

References
Ilson, R. 1986. *Lexicography: an Emerging International Profession*, Manchester: Manchester University Press.
Magay T. 2000. 'Teaching Lexicography.' In Ulrich Heid et al. (eds), *EURALEX 2000 Proceedings*. Stuttgart: Universität Stuttgart, IMS.

Lexicom takes place on 16-20 July 2001. The other two members of the Lexicography MasterClass (LMC) are Sue Atkins and Adam Kilgarriff. Sue Atkins is an immensely experienced lexicographer, a past president of Euralex, and the driving force behind the establishment of the BNC. Adam Kilgarriff, a Senior Research Fellow at Brighton's Information Technology Research Institute (ITRI), is an authority on the interface between lexicography and language technology, and has published extensively in both areas. http://www.itri.brighton.ac.uk/lexicon
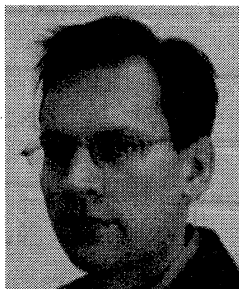
GLOBALDIX
a Unique Multilingual Dictionary
for the WorldWide Market

Mika Herpiö

Mika Herpiö has a master's degree from the Helsinki University of Technology. He is the director of business development and a partner in Kielikone Oy. mika@kielikone.fi www.kielikone.fi

Kielikone Oy is the leading language technology company in Finland, with more than 10 years of experience in electronic dictionaries. We started with diskette dictionaries in the late 1980s, and our latest products run on Linux and WAP telephones. Besides dictionaries we offer terminology management software, machine translation, morphological analysis and parsing software. Our focus is on products for corporate customers. We publish electronically works from more than ten publishing houses in Finland, Sweden and the UK*. They are all available in the MOT Dictionary Bookshelf, where users can select quality dictionaries and use them through a single, common user interface. MOT is available through varied platforms, such as Windows, Mac and Unix (Linux, Sun Solaris, etc), over the Internet and intranets, on mobile phones via WAP/GPRS service, and soon through Nokia Communicator memory cards as well.

Currently we are developing a new multilingual electronic dictionary based on the K Dictionaries semi-bilingual Password series. The first version, including well over twenty languages, will be launched this fall. It will be available both as part of the MOT Dictionary Bookshelf and as a stand-alone product named *GlobalDix*, that will be a shrink-wrapped software package for individual users, as well as for all the above-mentioned platforms.

All the multilingual dictionary solutions with which we are familiar list an item and its translations on a single line – just like database spreadsheet-tables filled with words from various languages. The problem with this kind of approach is, of course, the many grades of synonymy that words have in each language and the varying degrees of correspondence that a word and its translation(s) may have. For exmple, the structure of an English-German dictionary is not the same as that of an English-French one. Since a word may have several meanings, the structure of an article in the dictionary depends on whether the same translation can be used for all the meanings or whether a different translation is needed for each meaning.

There is an analogy between *GlobalDix* and multilingual terminology databases. Multilingual terminology databases are handy, useful and precise tools, since each entry is based on a concept that is common to all the languages covered. Password dictionaries are also based on descriptions of concepts and have a common set of definitions that are translated into various languages. As we learn from terminology, a well-defined concept can work as an anchor that binds together words from different languages. The English explanation in Password dictionaries may work as a similar anchor that connects words in many languages. Since most of the different Password versions are based on a fairly similar monolingual lexical core, in many cases the user can find how this meaning is translated into various languages. Combining these dictionaries together into a single software product enables cross-searches from any one language into any other language. For example, searching for the Finnish *erinomainen* would bring up its English definitions together with their translations in the other Password language versions based on the same core, and the translaions of the word in the Password versions not sharing the same core.

The first version of *GlobalDix* will contain the following language versions from the Kernerman Semi-Bilingual Dictionaries series: Chinese (traditional and simplified), Czech, Dutch, English, Estonian, Finnish, French, German, Hungarian, Icelandic, Indonesian, Italian, Japanese, Korean, Latvian, Lithuanian, Norwegian, Polish, Portuguese (Portugal and Brazil), Russian, Slovak, Spanish, Swedish and Turkish. It will cover a total of approximately 1.5 million headwords (all languages combined). The user is expected to have a basic knowledge of English, so that he/she can browse through the possible different meanings and choose the right one. The solution is aimed at helping the average person in her/his communicative needs, mainly reading foreign-language newspapers, websites, magazines and books. Another special application of *GlobalDix* is cross-language information retrieval (CLIR), where it will be possible to search for information in any of the languages covered through the user's mother tongue.

* WSOY, Gummerus and others, including over 20 titles for Finnish users. Gustava: Engelsk ordbok – Swedish-English-Swedish dictionary. HarperCollins: Collins English Dictionary, Collins English Thesaurus, Collins French Dictionary, Collins German Dictionary, Collins Italian Dictionary and Collins Spanish Dictionary.