The Coming Boom in English Lexicography –
Reconsidered (Part Two)

Charles M. Levine

Recap. In July 2001, I optimistically crawled out on a limb in these pages talking about an imminent boom in English lexicography.[1] Then in the July 2002 issue, my good friend and colleague Joseph Esposito countered that traditional, legacy dictionary publishers, like OUP and Merriam-Webster, "will muddle along, with growth becoming harder to come by except at the expense of their smaller and declining rivals; eventually they will stop publishing for broad markets altogether and the remaining activity will be to focus on the scraps Microsoft leaves on the floor."[2]

As a graduate in the history and philosophy of science, I have always been amazed at the uncanny ability of Esposito, a literature graduate, to anticipate important trends in technology. Possibly I get too preoccupied chasing down such details as how much, when? For example, how much will online dictionary searches replace print look-ups in the next three, five, and ten years? Twenty percent or 50 percent within ten years? But, as Esposito wrote (in a more general point about the future of traditional dictionary publishing), "Who knows?…In the absence of growth, the old [dictionary] business will be strained for capital, which will beget smaller investments, which will in turn hasten the decline."

Recently in the States. Certainly during the past two years, American lexicography has shown many of the strains Esposito wrote about. One could say these wounds were self-inflicted through corporate ownership dramas – but no matter, the

Charles M. Levine was recently the vice-president and publisher of Random House Reference Division, and has held a number of senior executive positions in digital and print content management in Asia and the USA. He is currently consulting, agenting, and packaging for a number of publishing clients. charlev@att.net

Levine's 'The Coming Boom in English Lexicography: Some Thoughts about the World Wide Web (Part One)', was published in Kernerman Dictionary News, Number 9, July 2001. http://kdictionairies.com/newsletter/kdn9-1.html

Joseph J. Esposito's sequel, 'Dictionaries, another Netscape?', was published in Kernerman Dictionary News, Number 10, July 2002. http://kdictionairies.com/newsletter/kdn10-1.html

A view from **Microsoft Corporation**, referring particularly to Esposito's points and introducing their overall policy concerning dictionaries and lexicography, is in preparation by Julian Parish, the recently-appointed Business Manager of Microsoft's new Dictionary Central Group, and due to be published in issue Number 12, July 2004.

cause of lexicography in America was not advanced, and possibly permanently hurt.

For example, Random House Webster dismantled its lexicography staff, as its relatively new Bertelsmann-owned parent grew tired of dictionary making. *Webster's New World*, sold to Hungry Minds (formerly, IDG Books), was off-loaded to John Wiley. *American Heritage*, as part of Houghton Mifflin, went to Vivendi, which then sold it to a private American-led investment group (headed by Bain Capital and Thomas H. Lee). The new *Microsoft Encarta* print dictionaries (created by Bloomsbury in the UK and distributed by St. Martin's in the US), after creating a splash of publicity, retreated in sales at retail. And the recent US$55 *New Oxford American Dictionary*—by all appearances an excellent product—assumed a dignified, but quiet, presence on bookshelves.

Only Merriam-Webster seems to have gained market share in America at the expense of its smaller competitors, who were hobbled by corporate problems. Currently shipping the new eleventh edition of its flagship *Collegiate Dictionary* (bundled for the first time with an electronic version and a free introductory online subscription – all for US$25.95), M-W recently boasted a 17 percent increase in dictionary sales, while their website bustled with more than 150,000 daily visits. The private investor (Jacqui Safra) who now owns Merriam, however, in a possible sign of impatience, recently brought in an outside CEO (G. Macomber) to find ways to grow the business more quickly. We will have to stay tuned for what develops at America's leading dictionary publisher.

In the aggregate, though, the American print dictionary market seems to stay at about the same size, year after year, even as online look-ups are apparently booming.[3] Is the resilience of print dictionaries in America a good sign? Is overall dictionary use (counting both print and digital look-ups) increasing? Possibly, yes to both questions. If I unscientifically use myself as an example, I now routinely consult two online dictionaries in addition to printed standbys – the OED online (accessible for free as a member of the Quality Paperback Bookclub[4]) and the faithful friend I once published, the *Random House Webster's Unabridged Dictionary* on CD-ROM. If a word still perplexes me, I search Google[5] or visit well-developed reference sites like Webopedia[6] (for technical terms). Because of the availability of multiple sources in both print and online, I am now much more dictionary literate – as are, I extrapolate, other serious word users.

Another noteworthy development in the States is the long-overdue progress toward creating the first comprehensive American National Corpus (ANC) – the first 10 million words of which are now being made available. My friend and colleague Wendalyn Nichols says more about this elsewhere in this newsletter[7].

The Bigger Picture. Looking beyond American shores and around the world, lexicography seems very much alive and well, if not booming to my optimistic drumbeat. The continued use and exploration of corpora and the vigorous linguistic research into world Englishes are two important signs of continued vigor. While at Random House Webster, I helped initiate—with the assistance of Nichols and others—an all-too-brief foray into creating entirely American-bred ESL/EFL dictionaries, partnering with publishers like FLTRP (Foreign Language Teaching and Research Press[8]) in Beijing, under the innovative leadership of Li Pengyi[9]. (Houghton Mifflin created the *American Heritage English as a Second Language Dictionary* in 1997 and revised it in 2002, primarily I believe to reach American schools and colleges – without strong marketing internationally.[10]) Except for these efforts the major American dictionary companies still appear blissfully lackadaisical about the potential of the global ELT market – which is probably the single largest area for growth in the English-dictionary business.

Maybe it would be more accurate to say that the global reach and penetration of English—especially as reflected on the WWW—will keep linguists and lexicographers busier for some time to come analyzing and mediating exchanges in which English is the lingua franca, and helping build the next iteration of the WWW, called the Semantic Web.

In the absence of an American corpus, for several years I have relied on the Web as a surrogate. For example, in a pop-reference book on Yiddish I co-created, playfully entitled the *Meshuggenary*[11], the Web was the best source of finding current uses in English of Yiddish-origin words and phrases (see the sidebar).

I am grateful to Michael Rundell, editor-in-chief of the *Macmillan English Dictionary for Advanced Learners*[12], for alerting me to the forthcoming special issue of *Computational Linguistics*[13], edited by Adam Kilgarriff and Greg Grefenstette, which will be devoted to the question of the Web as a corpus. Rundell notes that there are "many computational linguists who are beginning to see the Web as

the only corpus worth looking at (well, maybe I exaggerate somewhat), and as the solution to the long-running problem of 'data-sparseness'."

Of course, as Kilgarriff and Grefenstette point out, search engines like Google still have significant limitations as lexicographic tools—for example, in giving too much weight to words in the titles and headings of Web pages, and in missing the vast volume of material now hidden from the eyes of search engines behind "fee walls," in archives that charge to retrieve documents—but nonetheless it is clear that the Web presents lexicographers with a whole new set of opportunities to research current language use and should be considered a valid linguistic corpus.

Rundell also pointed out the heating up of discussions in the UK and Europe about English as a lingua franca – a development that looms so large throughout the world that it can still actually seem invisible to many native (English-centric) speakers. I noted in my first installment that an estimated 80 percent of Web pages are written in English (though that percentage may actually decrease somewhat over time; for example, Internet Explorer now supports the use of dozens of special scripts of the world's languages).

Somewhat hidden to those outside of Europe is the growing role that English plays (and a controversial one at that) in the affairs of the European Union. In 1970, about 60 percent of all documents coming out of Brussels were written in French, few if any in English. By 1997, English (45 percent) had surpassed French (40 percent) as the most frequently used official language.[14] (Is it less CNN, MTV, and MacDonald's that so haunts the French than ELF, English as a Lingua Franca? As language buffs, we can be sympathetic about the potential decline of any vibrant language.) The Semantic Web should further accelerate the use, promulgation, and importance of ELF. And by providing even richer data for research, the Semantic Web should also accelerate the business of global linguistics and lexicography.

The key step in building the Semantic Web will be the addition of metadata URIs—Universal Resource Identifiers—that "define or specify an entity, not necessarily by naming its location on the Web."[15] Put simplistically, the Semantic Web will establish protocols to identify types of content on each Web page – in ways to make the content elements computer readable and useable. As Tim Berners-Lee, who laid the groundwork for the WWW, writes:

The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web. This structure will open up the knowledge and workings of humankind to meaningful analysis by software agents, providing a new class of tools by which we can live, work, and learn together.[16]

I am obviously skipping lightly over a number of important new works-in-progress that will profoundly affect linguistic and lexicographic research in the coming years. (Searching on Google for the phrase "Semantic Web" plus "lexicographer or lexicography," restricted to English pages cached in the past year, yielded about 850 results, with a number of fascinating leads to recent papers and conferences. Adding "linguist or linguistics" to the search criteria increases the results to 4000.[17]) Work is just gearing up, and while I bemoan not having more hard data and numbers, my instincts tell me that lexicography and linguistics are on the verge of a revolution as a result – though, sadly for me, much of this new linguistic and lexicographic innovation may take place outside of America, even ironically as American-English is the driving force behind the increasing global use of English.

Postlude
Some interwoven comments follow, from those whose help in writing this article has been most welcome and appreciated.

Rundell: There's definitely a big growth in corpus development worldwide (especially but by no means exclusively for use in dictionary making) – sometimes, it seems, almost anywhere but the US. The big Japanese publishers like Kenkyusha and Shogakukan are all partners in the ANC consortium, but also busy with corpus development of their own. There is, for example, a 100-million-word Corpus of Professional English under development in Japan.[18]

Levine: All this corpus work is immensely exciting, and it is going to be interesting to see how it will influence the look and feel of future native-speaker dictionaries of English as well.

Instant Yinglish – Google's Top Dozen
(adapted from the *Meshuggenary*)

Searching on google.com gives a good indication of which Yinglish words are most frequently used today. **Glitch** and **kosher** top the charts, way ahead of all the rest; while even such well-known Yinglish words as **nosh**, **knish**, **schnoz**, **schmuck** and **gonnif** seem to have fallen behind in popularity, if one accepts Google's results. With the following list, you'll be instantly up and running in Yinglish.

glitch Slip-up; bug in the system. [232,000 hits]

kosher Legit, on the up-and-up; ritually clean. [222,000]

bagel The doughnut-shaped bread of champions. [145,000]

maven Expert; pundit; smart aleck. [70,800]

yid Jew. (Pronounced <yeed>; (but use with care: in U.S. slang, pronounced with a short i, as in bid, it is very disparaging.) [62,800]

klezmer Lively, heart-tugging Yiddish folk music. [46,800]

mensch Decent, trustworthy person. [42,600]

tush Backside; rear end. [39,500]

schlock Cheap or shoddy goods; junk. [39,300]

klutz Clumsy, inept person; blockhead. [39,000]

schmooze To chat or gossip; by extension, to network. [38,100]

chutzpah Impudence; moxie; cojones. [32,700]

The above results were derived from searching about 80 Yiddish-origin words that are now accepted in standard American English and would appear in up-to-date larger or unabridged dictionaries. The search was restricted to English Web pages, searching on word clusters such as [glitch glitsh glitchy], to take into account alternative spellings and closely related uses.

Kernerman: The implications may be more far-reaching and actually concern other languages too, not just English. For example, expect a dramatically growing demand for bilingual dictionaries of "unorthodox language-pairs", including so-called non-major languages.

Levine: One of the key developments that I hope to see is the sharp increase in the quality of these dictionaries covering odd bilingual couples.

Esposito: I see a rhetorical error in the paper – rhetorical, not substantive. You are using my paper as a pushing-off point, which is fine: I have been a straw man before. But the contrast is imprecise: you are referring to lexicography, I to dictionary-makers. Lexicography is bound to grow. The current crop of dictionary companies can't grow. Apples and oranges.

Levine: There actually may be two lacunae in what I write: (1) The boom may be more in computational linguistics than in lexicography as such, especially when it comes to the Semantic Web. Separately, I learned, for example, that many students who major in linguistics go on to careers in software. (2) I am implying if not stating that because of ELF, you could grow the dictionary-making business; but I fudge about addressing the key question – whether you could make it a "growth" business. It's somewhat like talking about the U.S. economy. If it grows at only 1-2% a year, it would still be considered an investment crisis. You are saying you doubt dictionary making as a business can grow at all and will decline.

Esposito: I think you can grow any business that adds value beyond the default value of a bundled Microsoft dictionary, and learners' dictionaries add value.

O brave new worldictionaries

Ilan J. Kernerman

Miranda   O, wonder !
How many goodly creatures are there here !
How beauteous mankind is ! O brave new world,
That has such people in 't !
Prospero   'T is new to thee.
(W. Shakespeare, *The Tempest*, Act V. Scene I.)

When I began working for Kernerman Semi-Bilingual Dictionaries in the early 1990s, our prime concern was the advent of English as a lingua franca—ELF (also here by Levine)—and its forthcoming consequences, for everyone to now learn and use worldwide, in equilibrium with their native language. This trend has indeed evolved, with critical impacts and interactions of various sorts.

Meantime, while we were snoozing, within that arguable globalization process—disseminating communication and information, grinding all into ubiquitous uniformity and mediocrity—other languages have also been awakening to each other and to themselves. Although ELF is champion—simultaneously and complementarily—it has become necessary and easier to create dictionaries for unorthodox language pairs, as well as to reach and explore—and sometimes even safeguard and enhance—any language still spoken.

With growing direct contact between languages not involving English, there are more trilingual and multilingual persons who want bilingual dictionaries without English, or dictionaries with two or more languages and ELF as an underlying bridge. Their quality might be painstaking to start with, but improvement usually follows suit. One way or another, soon you will be able to get any kind of dictionary, and via modern magic—be it Wi-Fi, broadband, cellphone—virtually anyhow, anywhere, anytime.

Some forecasts—such as teasingly or skeptically by Esposito—warn against the ill effects of such bliss on the traditional business of dictionary making. Yet, life forever intermingles so-called good with bad, bad with good, counter-running contradictions in cohabitation, bringing all together and centralizing, whilst breaking farther apart into atoms, quarks and who knows what next. So, big feed on small, the small disappear, yet big transform too. Giants come and go, while little men and women bear on. Change is—as Prospero might say—"such stuff as dreams are made of". You can tickle the Establishment, but the Establishment never goes away; its characters may be replaced but the roles remain.

Dictionaries pertain to civilized society, with an aim to confine the lawless jungle into order and fairness. How sad when they succumb to this very same jungle law, such as the recent change in Random House, which meant a mortal blow and terrible loss. On the other hand, I cannot lament the fate awaiting—according to Esposito—whichever legacy dictionary publisher whose so-called quality is undermined by a big name. The defamed Microsoft dictionaries might—eventually, if not yet—not be worse than theirs.

Sadly, established brand names are often something to beware of (so no surprise if sacred-cow slaughter becomes a global sport). When their originality has evaporated ages ago, big names get fat and preoccupied with self-preservation and enforcing monopoly, then impede the advance of new spirits whom they copy. That seems not a trait of dictionaries, but of humankind.

Levine: But, I do not fully address the key point (because it is so hard to pin down), which is whether you could grow a dictionary business enough to attract serious investment money. The sad truth of the matter is that it may be difficult that one could. A little growth is not enough, and there hangs a tale (as you would say) – a tale of most of the corporate problems we have recently seen in American dictionary companies, including M-W. Even if you could grow a dictionary business, under ideal circumstances, the growth may not be interesting or attractive enough for investors, American-style ones at least.

Other publishing entrepreneurs with a more worldly view, might be willing to accept a little growth. What puzzled me most, for example, about the dismantling of Random House Webster was that the dictionaries could be of immense benefit to the global branding of the Random House name, which fits in with the Bertelsmann global view, although not necessarily with the one emanating from the Broadway headquarters. For example, I am told that the Random House name is well known in both Japan and China, largely because of the local translations of the *Random House* dictionaries. But the latter benefits were apparently not strong enough to overcome the issue of growth for Random House/Bertelsmann.

I do lament this development, not only out of self-interest, but selfless interest when it comes to many good American lexicographers whose careers are being turned upside down by the stranglehold that dummies seem to have on dictionary publishing in the States today. But, all this is a natural process, in which the mighty fall by the wayside—out of hubris, complacency, or too much past success— and room is created for new scrappy small guys who are willing to take risks and innovate.

Links, notes, references

1 http://kdictionaries.com/newsletter/kdn9-1.html

2 http://kdictionaries.com/newsletter/kdn10-1.html

3 I roughly estimate that the combined size of the American monolingual dictionary business—measured in total annual sales dollars and including dictionaries sold directly to schools, exported abroad, as well as electronic versions—falls somewhere in the range of US$80-100 million annually. This number does not appear to have grown much during the past decade. One could point out that sales of electronic dictionaries, like the *Microsoft Encarta World English Dictionary*, do not seem to have greatly added to the total sales dollars, because of bundling with other products and/or the purchase of an electronic product in substitution for an equivalent one in print. This is all, of course, educated guessing.

4 http://qpb.com

5 http://google.com

6 http://pcwebopedia.com

7 http://kdictionaries.com/newsletter/kdn11-02.html

8 www.fltrp.com.cn

9 *The Random House Webster's Easy English Dictionary*, New York: 2001, paperback US$12.95, was certainly the first entirely new American English dictionary written specifically for middle school learners, a rapidly growing and underserved market around the world. In China, for example, soon every student, from the earliest grades, must master three core subjects: Mandarin, mathematics, and English.

10 One should also mention the *Newbury House Dictionary of American English*, edited by Philip M. Rideout, and published by Heinle in a revised edition in 1999. There is also a basic edition. Although Heinle is not a major dictionary publisher, these two editions are, according to Nichols, giving Longman a run for its money in the intermediate ESL market. McGraw-Hill, like Houghton Mifflin, has a small ESL group as well that tries to reach the domestic U.S. market.

11 *Meshuggenary: Celebrating the World of Yiddish*, by Payson R. Stevens, Charles M. Levine, and Sol Steinmetz. Simon & Schuster, New York: 2002.

12 Macmillan, London: 2002.

13 *Volume 29, Issue 3*, September 2003, MIT Press. www.mitpress.mit.edu/coli

14 See "Debate: The European Lessons," in *The Guardian Weekly*, April 18, 2001, posted at www.onestopenglish.com/culture/global/debate.htm.

15 Tim Berners-Lee, James Hendler and Ora Lssila, "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of possibilities," *Scientific American*, April 2002.

16 Op. cit.

17 Also see the recent paper delivered by David Jost of Houghton-Mifflin and Win Carus of the Dictaphone Corporation, at the DSNA (Dictionary Society of North America) conference held this past May 2003, entitled "Is the Semantic Web Possible?" (www.duke.edu/web/linguistics/dsnaabstracts.htm-jost).

18 See www.perc21.org/cpe_project/cpe_project.html.

English Dictionary Making in America Today

Wendalyn Nichols

Wendalyn Nichols is the former Editorial Director of Random House Dictionaries, and prior to that the Editorial Manager of Longman Dictionaries. She is now a freelance consultant and lexicographer, and is on the advisory board of the American National Corpus, the board of the Dictionary Society of North America, and the advisory board of Montclair State University 's graduate linguistics program.
wendalyn@nyc.rr.com

Why the UK still leads the way in the development of monolingual learners' dictionaries

The pioneering work in lexicographic publications for non-native learners of English was done in the UK, and the US has never really caught up. There are many reasons for this; the main one, I think, is the large size of the native speaker US domestic market combined with an unwillingness to cater to the special needs of immigrant populations; the prevailing attitude until the 1960s was the "bootstrap" mentality: "I (or my forebears) pulled myself up by my own bootstraps, and you should too."

The isolationism that prevailed in the US until the Second World War meant that few publishers saw the need to serve international markets, and domestically the US is such a large market for school publishing that the local educational publishers found it more lucrative to concentrate on producing school dictionaries geared toward the specific grade levels in elementary school and high school (called "elhi" for short). In contrast, Britain had a large empire (gradually replaced by the Commonwealth) as a ready-made market of people who needed to learn English (as a foreign language) to get ahead.

Once US publishers woke up to the need for special dictionaries for learners of English as a second language, they concentrated mainly on their already-established customers in the US market, specializing in literacy programs and bilingual (Spanish-English) education. These programs did not stress dictionary skills; at the lower levels students relied heavily on their bilingual dictionaries, and at the higher levels students were encouraged to switch to a standard native speaker dictionary.

Enough teachers admired the British EFL dictionaries that the *Oxford Advanced Learner's Dictionary* sold well in the US, and then Longman established a foothold in the 1970s. The *Longman Dictionary of American English* (LDAE) became the best-selling title once it was published in 1981, even though it wasn't truly American, being patchily Americanized from the *Longman Active Study Dictionary*. American publishers stuck to their elhi dictionaries, and so the British and US publishers happily split the market.

Why US publishers have been slow to create corpus-based dictionaries

The reason to keep up with the latest scholarship—like corpus-based lexicography—is an economic one, and too often reactive: if your books stop selling, then you figure out why. In the UK, the rivalry between Oxford and Longman, and the entry into the market of the COBUILD dictionary, meant that to keep up, everybody had to jump on the corpus bandwagon. US publishers, who were content to let the UK publishers have this slice of the market, did nothing about the new trend. Heinle & Heinle was the first US publisher to attempt an all-American ESL dictionary (the *Newbury House Dictionary of American English*), distinct from the Americanized ones coming from Britain, but it was written by one man rather than a team, and had no corpus input. Random House made the same mistake with its first foray into the monolingual ESL market, *Random House Webster's Dictionary of American English*. Now, it has always surprised me that a high percentage of US teachers prefer the Newbury House dictionary with its made-up example sentences to the second edition of the Longman one that is corpus-based; they like the pedagogical nature of the former. They'd gotten used to the first edition of LDAE, which pre-dates corpora and has example sentences that use a limited vocabulary.

It takes a lot of money to develop proprietary corpus data, and there was no equivalent initiative in America to the British National Corpus (BNC), because the US government has never supported lexicographic scholarship in the way that the UK has, and it's my understanding that the BNC would not have been possible without a huge chunk of money from Whitehall. At that time—the late 1980s and early 1990s—the ESL publishing market was undergoing great upheaval, with mergers, buyouts, acquisitions and divestments happening with such dizzying speed that even those US publishers who were aware of the "corpus revolution" could not convince their management to approve a significant, long-term, capital investment. Houses like Random House that did not have a history of selling into the ESL market didn't have the mergers problem to deal with, but they had the problem of financial models that no longer allowed for long-term amortization.

So, the UK educational publishers who have the greatest penetration into the US ESL market—Longman, Oxford, and to a lesser extent Cambridge—already have dictionaries now, and the US educational publishers remain unable to get approval for the kind of funding it would take to produce a product line that would rival the UK titles. McGraw-Hill ought to have seized the day—they had the cash, the sales penetration, and the size—but they chose instead to strike deals with other publishers to present their products to this market. NTC, the National Textbook Company, produces a large line of dictionaries that are, in my view, second-rate, but which people buy because they're cheap.

There is now the American National Corpus (ANC) Consortium, which got investment from enough publishers to start work that is modeled after the BNC so that comparative studies can eventually be done. The first 10 million words are being released this summer (2003). The initial founder investors have exclusive access during the developmental period; other commercial houses that wish to invest may still join, but at a higher fee than was the case for initial investors. Non-commercial educational institutions and individual researchers also have access from the start. The texts are being gathered under the supervision of Randi Reppen at Northern Arizona University; they are being tagged at Vassar under Nancy Ide; and the resultant corpus will be housed on the servers at the Linguistic Data Consortium at the University of Pennsylvania, which is also administering the licenses. (See next page.)

At this point, I see the UK and Japanese publishers as being more likely to take advantage of the ANC than American publishers, and for the disparity between British and American products to continue. I wish it weren't so; Charles Levine and I had great plans for the application of corpus-based lexicography to the Random House line, but what can you do when the visionaries don't hold the purse strings, and the upper management changes so often that you don't have a track record with them you can point to so that they trust you with large investments? This is the problem in nearly every US dictionary house; the one healthy one, Merriam-Webster, has so far remained unconvinced about introducing corpus-based lexicography. American consumers, meanwhile, will continue to make Merriam-Webster native speaker dictionaries their number-one choice; ESL teachers and students will continue to buy Americanized UK products.

The health—or otherwise—of US dictionary publishers vis à vis UK publishers

The top management of the big publishing groups look at the bottom line: dictionary publishing does not make the margins they like to see, so they are perennially putting pressure on the dictionary units to cut costs.

Merriam-Webster is the only major American dictionary publisher that is not under financial threat or at least dealing with perennial uncertainty: the publishers of the American Heritage line at Houghton Mifflin are still settling down after being sold by Vivendi; Random House closed its division in 2001; between 1997 and 2002, Webster's New World had three different owners. Encarta, the corpus-based UK-US collaborative project that was supposed to mark a new breed of dictionary, was done so quickly and edited so poorly that it was a near-complete failure: you now see copies of it everywhere on bargain book tables and street vendors' stalls next to the cut-price brands, because it had unprecedented numbers of returns of unsold copies from booksellers.

The Random House line, especially the great Unabridged Dictionary, is in danger of the fate of declining without any revision, unless another publisher decides to buy the rights to the Random House dictionaries and revive them. The current managers have even moved all of the citation cards into a storage facility where they cannot be readily accessed by anyone! Corporate changes are definitely a threat to the revision schedules and the very existence of the larger US dictionary publishing units.

Outside the US, American products simply do not have enough sales success to make an impact. The few exceptions, I think, included the works that Random House had the foresight (in the old days) to license for translation in Japan, Korea, and China – the beautiful editions of the Unabridged and College dictionaries that made Random House a respected name in East Asia. The American lexicographic tradition for native speaker products is long and illustrious, but the commercial climate has taken such a toll that the most brilliant lexicography now happens in specialized areas: Jonathan Lighter's *Historical Dictionary of American Slang*; the *Dictionary of American Regional English* project under Joan Houston Hall; and the recently-completed *Middle English Dictionary* at the University of Michigan, are examples.

Britain, in contrast, still maintains a commitment to promoting the English

language that is lacking in the US, so the UK-based publishers are less eager to divest themselves of dictionary units. The only dictionary house in the UK to undergo significant restructuring in recent years is Collins (the company is now HarperCollins), and this may have much to do with the fact that it is now owned by Rupert Murdoch's NewsCorp. Its schools assets in the US were sold to Pearson (Longman's parent company) in the 1990s; the COBUILD project was closed in the late 1990s because the sales of the product were disappointing. Collins still owns COBUILD and keeps updating it, but the lexicographic unit that produced it is no longer in operation. The dictionary program now concentrates more on native speaker and bilingual titles, and is based in Glasgow.

Having said that, it is becoming increasingly difficult for any commercially-owned unit, such as Longman Dictionaries, to get approval for new innovative capital projects – they seem to be in the "let's revise what we've got for now" mode. As for the two university presses: Oxford is also penny-pinching in most areas (it's more focused on its biggest capital project, the third edition of the OED); its Americanization of the *Wordpower* dictionary is not selling well. Cambridge now has a New York office and recently produced an American dictionary to compete with the LDAE, but its sales penetration is also disappointing.

What it will take to be a lexicographer in the future

The quality of a lexicographer will still depend heavily on all the traditional skills, as well as talent. I've trained plenty of people who learned the basic concepts but never became truly good, instinctual lexicographers – and unfortunately there are too many people out there who've had lexicographic training whose work is really quite patchy. Anybody can be taught the basic principles in a university course or an in-house training program on lexicography, but it takes someone with an instinct, an ear for the language—a poet, I would argue—to find just the right genus and differentiae and commit those to paper (or electronic database!) within the restrictions of a particular style guide.

A lexicographer will still need to have something of the teacher in him or her: an ability to convey complexity in a clear, simple, consistent form. A lexicographer will still need an unerring knowledge of grammar and a curiosity about usage and new words that keeps him or her alert to changes in the language – new words, new uses, shifts in sociolinguistic register. He or she will still need to be able to interpret citations, which have their own role to play in an active reading and marking program alongside corpus data. He or she will still need a keen attention to detail.

The skills required of a lexicographer going forward are also going to include an ability to analyze corpus data quickly and judiciously, identifying and differentiating significant patterns from "rogue" uses of language, and making allowances for any bias the corpus may have. The lexicographer will have to understand data tagging and be able to work in an electronic medium, manipulating entries across databases.

Electronic applications and consumer (non-)awareness

There are some good CD-ROM products on the market from reputable companies, and then there are a lot of bad products with very old data sets being offered for license at bargain-basement rates. You get what you pay for. Electronic handhelds are still limited in their usefulness and helpfulness because of the limitation on memory; I think that wireless handhelds could solve that problem. That's where the future is, so whoever is first at successfully manipulating their data into a compelling, flexible, and useful format for wireless access, and can strike exclusive deals with the main manufacturers, is going to make a lot of money.

The perennial problem is that consumers the world over do not know how to tell a good dictionary from a bad one – it doesn't matter if it's print or electronic. They look at the number of definitions the product claims to have, and buy the one with the largest number. And the manufacturers of these devices often choose the cheapest licensing deal they can get rather than the best content. About the only defense against this is strong consumer awareness campaigns – if a manufacturer were to choose a high-quality licensing partner (or develop its own high-quality English content) and then hit the market with a very strong marketing campaign that focused on the quality of the product, educating the consumer in the process, then it might make a dent in this trend. That's how Longman beat out Oxford in many markets: they were quicker to exploit corpus resources and more innovative in their applications, and were able to demonstrate the difference in a global blitz of teacher-training workshops and conference presentations. Therefore, schools that teach English ought to be teaching the students how to choose a dictionary; you're not going to convince

manufacturers to reform their practices, so you've got to teach the consumer not to buy the inferior products.

The Internet also contributes to the confusion of quantity—or ease of access—with quality. Being mindful of the quality of the source matters, regardless of whether the delivery format is print or electronic. I think it was a mistake to offer online dictionaries for free – the newer works that are still under copyright and are the most up-to-date should have been set up with a subscription model from the beginning. Internet users now feel that they have the right to free information, no matter how much it cost the original publisher to produce it. Some publishers, like Columbia University Press, have been successful with encyclopedic works offered online by subscription, and I think people will start to accept this model, especially now that companies like Napster have been barred from allowing free music downloads of copyrighted material.

Dictionaries and references

American Heritage Dictionary of the English Language, 4e, 2000. Boston, MA: Houghton Mifflin.

ANC: the American National Corpus, www.americannationlcorpus.org

BNC: the British National Corpus, www.hcu.ox.ac.uk/bnc

Cambridge Dictionary of American English, 2000. New York: Cambridge University Press.

Collins COBUILD English Dictionary for Advanced Learners, 1987, 1995, 2001. Glasgow: HarperCollins.

Dictionary of American Regional English, 1985, 1991, 1996, 2002. Cambridge, MA: Harvard University Press.

Encarta World English Dictionary, 1999. New York: St. Martin's Press.

Longman Active Study Dictionary, 1983, 1991, 1999/2000. Harlow: Longman.

Longman Dictionary of American English, 1983, 1997, 2002. New York: Longman.

Longman Dictionary of Contemporary English, 1978, 1987, 1995, 2003. Harlow: Longman.

Merriam-Webster's Collegiate Dictionary, 10e, 2002. Springfield, MA: Merriam-Webster.

Middle English Dictionary, 8e, 2001. Ann Arbor, MI: University of Michigan Press.

Napster, www.napster.com.

Newbury House Dictionary of American English, 1996, 2000. Boston, MA: Heinle & Heinle.

Oxford Advanced Learner's Dictionary of Current English, 1948, 1963, 1974, 1989, 1995, 2000. Oxford: Oxford University Press.

Oxford American Wordpower Dictionary, 1998. New York: Oxford University Press.

Random House Historical Dictionary of American Slang, 1994, 1997. New York: Random House.

Random House Webster's Dictionary of American English, 1997. New York: Random House.

Random House Webster's Unabridged Dictionary, 2001. New York: Random House.

Webster's New World Dictionary and Thesaurus, 2e, 2002. Hoboken, NJ: John Wiley & Sons.

The American National Corpus

The ANC Consortium members include publishers, software companies, and academic members. Members have exclusive access throughout the development period and for five years after the full corpus becomes available. The ANC is freely available for the purposes of academic research and education.

Acquired data includes, so far, about 2 million words of spoken data (the LDC Switchboard corpus and a portion of the CallHome corpus); 1.5 million words of previously unreleased newspaper data from the *New York Times*; a few hundred thousand words of "ephemera" (pamphlets, newletters, etc.); several novels published by Oxford University Press USA; Berlitz Travel Guides from Langensheidt; Verbatim magazine; government documents drawn from the web; about 5 million words from Slate magazine (Microsoft); and about 900,000 words of research papers from the Association for Computational Linguistics.

Commercial Members

• Pearson Education
• Langenscheidt Publishing Group
• HarperCollins Publishers
• Cambridge University Press
• LexiQuest
• Microsoft Corporation
• Shogakukan Inc.
• ACL Press Inc.
• Taishukan Publishing Company
• Oxford University Press
• Kenkyusha Ltd.
• IBM Corporation
• Obunsha Publishing Co. Ltd.
• Bloomsbury Publishing Plc
• Benesse Corporation
• Sanseido Co. Ltd.
• Sony Electronics Inc.
• Macmillan Publishers

Academic Members

• Vassar College
• Northern Arizona University
• New York University
• Linguistic Data Consortium, University of Pennsylvenia
• International Computer Science Institute, University of California, Berkeley
• University of Colorado at Boulder

Some Lexicographic Concepts Stemming from a
French Training in Lexicology (1)

Jean Pruvost

Jean Provost is the
Vice-President of Université
de Cergy-Pontoise, where he
teaches linguistics, lexicology
and lexicography, directs the
laboratory of the CNRS
[National Centre for Scientific
Research] (Metadif, UMR
8127) that is devoted to
dictionaries and their history,
and organizes the annual
international conference, *La
Journée des dictionnaires*, a
meeting point for lexicologists,
lexicographers and dictionary-
makers. He is the author of
*Dictionnaires et nouvelles
technologies* [*Dictionaries and
New Technologies*], for which
he received the international
linguistics prize *Logos* in 2000,
and of two titles in the *Que
sais-je?* [*What Do I Know?*]
series, one on French language
dictionaries, the other on
neologisms (with Jean-François
Sablayrolles), as well as of
numerous articles. Professor
Pruvost co-directs, along with
Bernard Quemada, two
collections at Éditions Honoré
Champion (*Études de
lexicologie, lexicographie et
dictionnairique*; *Lexica*), serves
on several editorial boards
(*Cahiers de lexicologie*; *Études
de linguistique appliquée*;
*International Journal of
Lexicography*), and presides the
Association des Sciences du
Langage.
pruvost.jean@wanadoo.fr

Introduction

During the Euralex Congress held in
Copenhagen in 2002, three key ideas in
particular occurred to me as a French
lexicologist and lexicographer making
contact with English-speaking colleagues:
• first, there was obviously a great deal
  for a French specialist to learn from
  his non-French-speaking colleagues
  concerning their specific approaches to
  our discipline;
• then, it seemed to me from the outset
  that to date there had not been much
  interaction between "Francophone"
  approaches and "Anglophone" ones;
• finally, I realized that while I was learning
  a lot from these colleagues and friends,
  I might also have a number of points of
  view and methods that were peculiar to
  my training, and which could effectively
  contribute to a meeting of minds.

Then, too, I was encouraged by Ilan
Kernerman, as well as by Tony Cowie,
to try to set out some of the ideas which
form part of my credo and my background
as a contribution to this newsletter. Three
perspectives in particular seemed to me to
be worth exploring.

The first corresponds to the distinction
to be drawn between "lexicographie"
(roughly, "theoretical lexicography") and
"dictionnairique" (or "practical dictionary-
making"). The notion of "dictionnairique"
has in fact been quite recently introduced
by Bernard Quemada, editor-in-chief of the
Trésor de la langue française (16 volumes:
1971-1994), and has been adopted fruitfully
by numerous French lexicologists. This
distinction seems fundamental to me.

The second perspective is the one
developed by Robert Galisson under the
heading of 'lexiculture'. Galisson is one
of our most original and productive
lexicologists in the field of French as a
foreign language. Yet, in fact, lexiculture
is probably one of the most neglected of
the components involved in the editing of
entries in French or English dictionaries,
and is sometimes even completely
overlooked.

The third perspective is what I call the
'triple investigation in dictionary-making'.
Various lectures I have given on the
subject have convinced me that this
particular approach could very likely play
an important part in the improvement of
our dictionaries.

1. The helpful distinction between *lexicographie* and *dictionnairique*

In order to understand fully the difference
between these concepts and to appreciate
their essential complementarity, it is
necessary to place them within the recent
history of French dictionaries which is,
more or less, not all that remote from the
history of lexicography in other Western
countries. One can actually distinguish four
successive stages during the second half of
the 20th century.

1.1 *Lexicology* disassociated from *lexicography*, in the traditional sense of the term

From 1950 to 1965, we can separate off
an initial period in which one needs to
draw a distinction between "lexicology",
the scientific study of words, and
"lexicography" in the traditional sense of
the term – that is, the actual compilation
of dictionaries. We know that, in fact,
lexicology as the study of words did
not really attain the level of a scientific
discipline until the second half of the 20th
century. In France, a particular date has
become symbolic in this respect: it was not
actually until 1959 that the first issue of
the *Cahiers de lexicologie* appeared; this
scientific journal, established and edited
by Bernard Quemada, entered the 21st
century having appeared in no less than 78
issues and with an unchallenged scientific
reputation.

During this first period, lexicology and
lexicography in the classical sense of the
term have each redefined themselves and
been redefined in relation to each other,
lexicology becoming in the fullest sense
a scientific discipline, and lexicography
clearly coming to meet at one and the
same time the requirements of a craft and a
science.

Lexicologists, while fully accepting the
continuity of their discipline with philology,
had then to take the measure of the
emerging theory of structuralism and of
the most recent technologies of the day,
technologies represented at the time by
punched-card machines. The study of the
lexicon and of specialized vocabularies by
reference to large corpora, and with the aid
of punched cards – such was the pioneering
issue of this period. It was notably at
Besançon, in its laboratories equipped with
punched-card machines, that lexicologists
from all over Europe were trained. Thus,

in June 1961, a symposium on the mechanization of lexicological research, which today has symbolic value in my eyes, was organized by Quemada at the University of Besançon. It was a meeting altogether in keeping with the new state of mind which had set in. What in fact does one of the participants, the Reverend Father Busa, director of the 'Centro per l'automazione dell'analisi letteraria', from Gallarate, near Torino, have to say? "One is aware that all of us taking part in this conference are pioneers in the automation of lexical analysis. We are playing a necessary part in the continuing evolution of the book [...]. Today, alongside printed journals and books, the *electronic book* is in the process of finding a place for itself." Such a statement, made in 1961, deserves to be recognized as visionary! It shows in any event that lexicology is acquiring a further dimension derived from the new technologies that are now emerging.

As for the lexicography of this period, it is distinguished in France by the awareness that dictionaries should be based on a closer analysis of the criteria by which they are defined. Dictionaries are no longer considered simply as working tools, but are examined in an altogether new light.

Interest was shown most particularly in the ideas underlying the *Dictionnaire du français fondamental* (1958), with a line of descent from "Basic English". This dictionary of basic French derived from a scientific programme with didactic perspectives, being based on an analysis of vocabulary frequency. Then, while we awaited a new large-scale dictionary of the French language, we witnessed the republication of a large dictionary that was symbolic of the 19th century, the *Dictionnaire de la langue française* by Littré (first edition, 1873; reprinted in 1956). During the same period, in the *Grand Larousse encyclopédique* in ten volumes (1960-1964), the first defining steps were being taken towards the use of the new technologies – those then developing and specifically the 400,000 punched cards assembled in advance of the compilation of this paper dictionary. An *encyclopedic dictionary*, but one that was effective also in its treatment of the language and of technical varieties, the *Grand Larousse encyclopédique* deserves to be remembered as one of the landmark dictionaries of the period. We had not yet reached the computer era in the precise sense of that term, but very rigorous methods, based on algorithmic analysis, were already in evidence.

However, for the moment, "lexicography" could still keep its traditional meaning: it approximated in fact to the compilation of dictionaries, with recourse as circumstances required to the most suitable technologies, and the use of editorial teams that were becoming more and more professionalized.

1.2 The birth of *metalexicography* and the new distinction between *lexicographie* and *dictionnairique*

The second period, broadly speaking, ran from 1965 to 1980, and demarcated a stage when the dictionary enjoyed a new status, being widely recognized as an object of scientific research. A French thesis entitled *Les dictionnaires du français moderne (1539-1863)* (Didier, 1968), submitted by Bernard Quemada, established itself as a benchmark for numerous studies that would come to fruition concerning one dictionary of the past or another. A new discipline was thus born: "metalexicography". Lexicography, which until then, had been mainly concerned with meeting everyday needs, with the dictionary regarded above all as a tool, henceforth became linked up with corpora that were studied in order to improve understanding of the history of the genre and the functioning of the language. In the process, dictionaries ceased to be simply the creation of philologists and skilled craftsmen; they became a matter for linguists as well.

The second period coincided with a time of intense commercialization of dictionaries intended for the general public, and a veritable revolution in information technology—in advance of actual dictionary production—in the classification of data and its interpretation. Actually, the field of lexical research assumed a new breadth, while it became easier to produce dictionaries based on different computerized databases, adapting them for different categories of reader. At that stage, Quemada introduced a new dichotomy, between "lexicographie", to which he gave a new meaning in relation to its traditional sense, and "dictionnairique", both concepts forming a helpful dichotomy while being perfectly complementary.

In the new opposition established between "theoretical lexicography", i.e. in its new definition, and "practical dictionary-making", lexicography went well beyond the activity of editing a dictionary to include actual scientific research, carried out into words and their categorization, with all the related work of definition.

"Dictionnairique" represented, by contrast, everything that has to do with the concrete aspects of production, and presentation, for a given readership, with all the commercial imperatives that are

Praktyczny SŁOWNIK
angielso-polski
polsko-angielski
WYDAWNICTWO
NAUKOWE PWN,
Warszawa
August 2002
Translation: Jadwiga Fisiak,
Michał Jankowski, Tadeusz
W. Lange
Editor: Jacek Fisiak
ISBN 83-01-13777-0
Hardcover, 920 pp.
243 x 170 x 45 mm.
http://kdictionaries.com/
products/medium/
medium-polish.html

called for in order to satisfy that public.

With "lexicographie", one is situated squarely within the domain of research, without being concerned about highlighting details for the non-specialist reader, or worrying about adapting the content for users who are simply buying a product. One is indeed at some distance from the dictionary as something that is shaped in order to be sold; one is engaged in pure research. There can even be a form of *lexicography* that, in contrast to the common definition of lexicography as "the compilation of dictionaries", does not necessarily lead to a dictionary that can be sold. Corresponding to this or that piece of research into groups of words, or their definitions, it may very well not leave the laboratory but correspond, for example, to the computerized databases designed uniquely for researchers. Here one is not at all concerned with grading information for an attractive product to be marketed in a purchasable form.

As for "dictionnairique"—a word that Charles Nodier had already used in the 19th century, but which fell into oblivion until Quemada resurrected it—this denotes by contrast the activity of developing a dictionary as a product, which is then offered for sale—with all the constraints and problems relating to each edition—as a reference tool, a cultural product deliberately designed for a specific group of potential buyers. Thus, one must never forget that the dictionary represents a technical-commercial product whose content is defined as a function of the means granted to it for a defined clientele, within the context of a precise piece of market research.

Let us take, for example, two dictionaries that are very widely diffused in the French-speaking world and which are considered to be of high quality, in this case the *Petit Robert* or the *Petit Larousse* (and let us remind ourselves that on average 200,000 copies of the *Petit Robert* are sold each year, and 800,000 of the *Petit Larousse*, thus over one million copies in 2001). When a new edition is being considered (and this is every year, since they bear a date) and it is necessary to add a new word to a particular page, it is out of the question to reset the entire printed dictionary at the beginning of each school year. This or that example is simply removed from a neighbouring entry on the same page—or one or another shade of meaning—to free up the few lines which will allow us to insert this new word without touching the beginning of the page and its end, and thus without having to modify the preceding pages and those that

follow. Here one is plainly in the domain of "dictionnarique": these are the practical constraints that prevail over definition quality and precision.

One should emphasize, to illustrate yet more clearly the difference between "lexicographie" and "dictionnairique", that one may be an excellent lexicographer, that is to say carry out effective research on groups of words, and their definitions, and yet turn out be a dreadful dictionary-maker, that is one who fails to meet the production deadlines and the material restrictions that are inevitably imposed. One knows of major dictionaries that in their earlier volumes boast of enormous entries, making them almost illegible, and in which, as one goes along, because of a lack of space and because it has already been necessary to increase the number of volumes that were initially planned, the entries get thinner, and one can even find oneself at the end of the alphabet with very poor entries indeed.

Publishers should not be confused with researchers. Publishers must sell a product of a given size to a readership that can be won over at a particular price, and during a given period of sale: practical lexicography concerns them above all else. There is no mystery about the rules: if the product is inappropriate, excessively large, inconsistent in the density of information provided, the dictionary as a product will not be successful, it will not sell, and the publishing house will be in mortal danger.

Be that as it may, theoretical lexicography and practical dictionary-making are complementary: there is indeed no dictionary-making of interest that does not rest on a solid basis of lexicography, and lexicography is sometimes more effective if it has been able to take account of the practical constraints of time and space which, in a sense, contain it and lead it towards greater homogeneity in the description of a large body of words.

1.3 A revealing distinction between basic principles

It is possible to learn a number of lessons from the necessary distinction between *lexicographie* and *dictionnairique*.

First of all, it is important to separate the two perspectives, 'lexicographiques' and "dictionnairiques". A dictionary, that is to say, a product, in which the two perspectives are blurred, runs the risk of failing to meet the user's expectations. In general, the user wishes to have information that is precise, but not overwhelming. If he buys, for instance, a thousand-page book, he would rather have useful and clear information than information that tends towards the exhaustive and transforms each entry into an exercise in compression,

into a dreadful digest. To aim always at providing the maximum of information in the minimum of space, is to condemn the user to reading with a magnifying glass, to the "intellectualized" reading of the researcher. After all, one might reflect, the dictionary represents a genre in which the editor in principle avoids stylistic verbosity, which is considered out of place. Superfluous details are hunted down; the object is scientific and for this reason, one needs, it is assumed, to be austere.

Note, however, that the first monolingual French dictionaries, those of the 17th and 18th centuries, which possess great charm, do not seem restricted to this degree by scientific rules, of an almost monastic nature, governing the entire work. Constantly saving space, by condensing as much as possible in order to add new information, is not sound practice. Outside the "dictionary" genre, in works of a didactic nature, redundancy is indeed very much present, not to say essential, for the explanations. It allows one to space out the information, to make it accessible and digestible; it also affords one the possibility of putting forward different approaches. Too much dense information definitely frustrates the effective conveying of information, while taking away all the pleasure from consultation. Without a *practical* approach which allows for flexibility of editing from the start, and without a *theoretical* side that can keep the "lexicography" involved within certain limits, so as to allow for the "dictionary-making" that is required—the kind that will make the reading of the dictionary a pleasant experience—you will no doubt sacrifice one of the primary functions of the dictionary: to make the information clear but also pleasant and legible. It is easy to add a mass of information in the name of "lexicography"; it is difficult to curb oneself and to choose as a good "dictionary-maker" should the discourse which is best suited to the task.

Then, to compare the findings of research to the editing of an entry that should account for them in their entirety, is to confuse the two stages. There is a time to conduct research, to be engaged in *lexicography*, resulting in an entry intended for the researcher alone; there is another time to adapt the results to the user's needs, to move into *practical dictionary-making*. Here, one does not want necessarily to hand on everything that has been found at the research stage. The entry which is then compiled is aimed at a reader who is not a linguist, nor willing to read and reread definitions that are too dense. The information to be given to the reader should

not be confused with the straightforward recapture of the dry, scientific language expected by the linguist. Thus, the absolute meticulousness and the concern for exhaustiveness which are uppermost in research are no longer necessarily the primary criteria: one must adapt in order to explain better. The lexicographer-researcher can write for his peers when he is in the domain of research, but when he becomes a dictionary-maker, he no longer writes for his peers; he writes for all his readers and especially for those who are not linguists. The dictionary has a didactic function as a reference tool for everybody.

Finally, have we reflected sufficiently on the fact that though it is desirable that the researcher knows everything possible about the use of a word in the language, he should, as a dictionary-maker, not necessarily summarize it in as little space as possible, but on the contrary take account, as much as he can, of the particular questions that the reader asks concerning that word? Yet, the often systematic treatment of information that is characteristic of the way we proceed in linguistics, sometimes falls short of answering many of the specific questions that dictionary users ask themselves about such and such a word.

As a matter of fact, of course, the term "word" is used in various senses: "word as part of the language system", "word as part of discourse", "literary word", "word as referring to something (having a referent)", and so on. Now, let us not forget that the word as treated in the dictionary, which often corresponds to a more or less successful synthesis of some or all of these senses hidden in a single word, is not the dictionary word itself (i.e. the item in bold print that heads an entry and is often referred to in English as the 'headword' or 'entry word'). The word described in the entry, then, is not simply that which has been analysed with reference to language and discourse; it may first be perceived as the "consulted" word (the headword). And, in this sense—the word as something consulted—has to some extent its own difficulties which often escape the uniform rules of description, devised for the whole body of words in the dictionary.

I have drawn up a summary list of difficulties relating to certain French words, and it will be noted that each word is almost always consulted in the dictionary to resolve the same difficulty. It is a curious fact that (in France) no-one has brought together groups of linguistically naïve observers, or of dictionary users, noting systematically the questions that they turn to the dictionary to resolve.
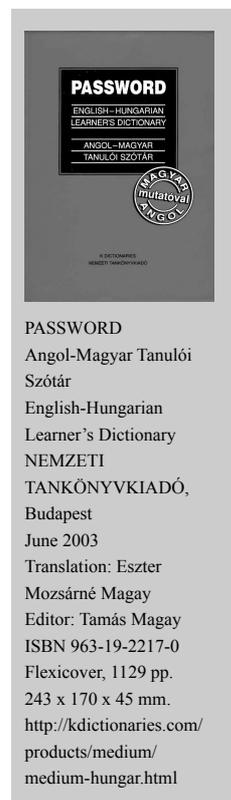
PASSPORT Mokomasis
Anglų-Lietuvių Kalbų
Žodinas Pradedantiesiems
English-Lithuanian
Learner's Dictionary for
Beginners
including CD-ROM
LEIDYKLA ALMA
LITTERA, Vilnius
September 2002
Translation: Rasa Katiliūtė
Editor: Rita Šepetytė
ISBN 9955-08-198-8
Hardcover, 639 pp.
234 x 150 x 35 mm.
http://kdictionaries.com/
products/junior/
junior-lithuan.html

A large-scale investigation of this type would be quite revealing. There has been such a proposal concerning online electronic dictionaries, connected with the organization of automatic logging of questions posed, but studies of the needs that were revealed are lacking.

Thus, how is the French verb *rejeter* (*reject*) written in the future tense: *rejetera*, or *rejettera*? Hardly any dictionary considers including this as an example, although 80% of look-ups of the entry focus on this question. For the commonly-used abbreviation *pro* (*a professional*), the plural is never given, and people hesitate – can they write *pros*? There too, those looking up this word are concerned for the most part with this question of spelling. Now, for the linguist, the problem does not arise. He has in fact thought in terms of very general rules which apply throughout the entire work, and he considers that, if he does not provide this or that piece of information, it is because in his eyes it goes without saying. If he does not provide a special note, it is because in this case one follows the general rule. Obsessed as we are, as lexicographers, with the need to save space, any saving of space is welcome, general rules enable us to gain typographic space: too bad for the reader who is not aware of our obsession and who looks up a word just because he is not aware of the general rule! This is to treat with disdain the concerns of the anxious dictionary user. The truth is that, statistically, the user never reads introductions and wishes for a direct reply to his queries. Thus, as regards the particular and complicated rules of pronominal agreement in French according to which "ils se sont développé" (they have developed (themselves)) does not take an 's', here too, it is quite rare that the dictionaries offer the illustration hoped for in the examples. The same is true regarding semantics, where one does not expect necessarily and systematically an exhaustive description of all the semantic components of the word, but sometimes examples clarifying the referent, as it was yesterday and is today. Hardly any French person looks up the entry *chaise* (*chair*) for its spelling or its use in the language: here, it is the referent that is crucial. It is therefore treatment of the referent that should be further expanded in sound *dictionnairique*, yet a French "lexical" dictionary in principle does not provide illustrations. It is therefore necessary to know beforehand the meaning of *chaise haute* (*highchair*), *chaise longue* (*deckchair*), *chaise à porteur* (*sedan chair*), *chaise percée* (*commode*), *chaise roulante* (*wheelchair*), to gain any advantage from

these words, presented in lists in lexical dictionaries, married (not always) to the briefest of definitions.

The distinction which to us seems fundamental between the notions *lexicographie* and *dictionnairique*, the latter being concerned with providing a pleasant and efficient tool for the user and not for the researcher, does not disappear during the two subsequent periods of modern dictionary history. On the contrary, it has become all the more important. The second period we have called to mind saw the birth, in fact, of very great dictionaries such as the *Dictionnaire du français contemporain* (1966), of a distributionalist nature; the *Trésor de la langue française* (1971-1994), of a philological nature, based on a computerized textual documentation, unrivalled for the French language; the *Grand Larousse de la langue française* (1971-1978), likewise of a distributionalist approach; the *Dictionnaire alphabétique et analogique de la langue française* (1964; *Supplément* in 1970) by Paul Robert, in a renewed continuation of Littré. In truth, then the third period, which extends from 1980 till about 1995, only further reinforces the useful distinction we have drawn between "lexicographie" and "dictionnairique".

We are witnessing then a connection between, on the one hand, the domains appropriate to dictionaries designed for human consultation, and, on the other hand, computational lexicography, a discipline reserved until now for computer scientists. This last discipline brings together all that constitutes the basis of lexical knowledge and all that has to do with machine-dictionaries for the computerized treatment of languages and the language industries. There has been a meteoric rise in research in this field, and computerized resources are making possible very broad-ranging studies. *Lexicography* in the Quémadien sense of the term is in full swing.

Elsewhere, information technology, even before the birth of the first CD-ROMs and Internet, made possible the gearing-down of dictionaries intended for the general public by working from well-nourished databases. Many small dictionaries thus appeared, diversified according to the ages of their potential users; and *dictionnairique* could henceforth even dispense with the data provided by research, and sometimes enjoy an autonomy outside of *lexicographie*. Here was no longer research, but the adaptation of data, with as many "cocktails" of the data as there were potential readerships. Diversify to achieve better sales! And this is sometimes made possible by adapting, with talent and

efficiency, existing knowledge. This was the case with the *Dictionnaire historique de la langue française* (Le Robert, 1992), for example, which set out in a pleasant way etymological information provided by researchers, those of the CNRS, most particularly. Sometimes, by contrast, we are provided with nothing more than a fairly dull mixture of information, selected and targeted so as to secure a well-calculated commercial profit, in the manner of a well-packaged product.

If we wish to ensure that *lexicographie* does not turn in on itself, and that *dictionnairique* does not self-reproduce, these are the goals that should not be lost sight of. Both perspectives should remain united and complementary. Without research, there is in fact no interesting future for dictionaries. And without sound *dictionnairique*, lexicography might shrivel and be of benefit to very few, without really attracting new competence and skill.

As for the last period, the very end of the 20th century and the beginning of the 21st, this has been marked by the development of the Internet, which distinguishes itself primarily by the revival of editorial strategies, extended and adapted to the new virtual spaces, infinite spaces of information accessible in real time. It is defined also by a profound metamorphosis of our look-up habits.

A problem remains: for the time being, we are confronted chiefly by electronic adaptations of products, published not long ago in paper form, which are in the process of being redeveloped or offered on the market. This is "redictionarization", moving from paper on to computer, adding to it all the proper tags for the richest and most reciprocal look-up possible, and matching to it internet links. We are nevertheless left with the task of developing dictionaries conceived from the outset for the computerized medium, with no doubt real openings-up for the hypertextual means between the encyclopedia and the language, between the synchronic and the diachronic, between the general vocabulary and the specialized vocabulary, between textual examples and visual and sound examples, all in all synesthesic. Added to all this is *lexiculture*, which we will introduce later.

New approaches to lexicography and practical dictionary-making remain to be developed: the potential field of activity is immense. Many users start off by preferring disorganized searching on the Internet, which is no doubt rich but also unpredictable, to the consultation of a real "dictionary" based on this breaking down of barriers. Already, some works are taking up the challenge, especially in the field of learners' dictionaries. Working together, on a global scale, we will not be cautious.

References

Galisson R. et J. Pruvost (dr.), 1999. *Vocabulaire et dictionnaires en français langue maternelle et en français langue étrangère*. Étude de linguistique appliquée, n° 116 (octobre-décembre). Paris: Didier Érudition.

Pruvost J., 2002. 'À la recherche de la norme: sa représentation lexicographique et dictionnairique chez Larousse et Robert et la triple investigation'. In *La représentation de la norme dans les pratiques terminologiques et lexicographiques*, Langues et sociétés n° 39. Montréal: Office de la langue française, Québec, pp. 139-170.

Pruvost J., 2002. *Les dictionnaires de langue française*. Collection 'Que sais-je?' n° 3622. Paris: Presses Universitaires de France.

Quemada B., 1961. 'Actes du Colloque sur la mécanisation des recherches lexicologiques'. In *Cahiers de lexicologie*, n° 3. Paris: Didier Érudition.

Quemada B., 1987. 'Notes sur lexicographie et dictionnairique'. In *Cahiers de lexicologie*, n° 51. Paris: Didier Érudition, pp. 235-245.

Dictionnaire alphabétique et analogique de la langue française, 5 vol. et Supplément, 1964-1970. Robert P. Paris: Société du Nouveau Littré.

Dictionnaire de la langue française, 4 vol., 1863-1873. Littré É. Paris : Librairie Hachette.

Dictionnaire du français contemporain, 1966. Dubois J. (dir.). Paris: Larousse.

Dictionnaire du français fondamental, 1958. Gougenheim G. Paris: Didier.

Dictionnaire historique de la langue française, 2 vol., 1992. Rey A. (dir.). Paris: Le Robert.

Grand Larousse encyclopédique, 10 vol., 1960-1964. Dubois Cl. (dir.). Paris: Larousse.

Le Petit Robert, Dictionnaire alphabétique et analogique de la langue française, since 1967, revised every year. Paris: Société du Nouveau Littré, Société Le Robert.

Le Petit Larousse illustré, since 1905, revised every year. Paris: Larousse.

Trésor de la langue française, Dictionnaire de la langue du XIXe et du XXe siècle (1789-1960), 16 vol., 1971-1994. Paris: Klincksieck, vol. 1-10. Paris: Gallimard, vol. 11-16. Dir.: Imbs P. (vol. 1-7), Quemada B. (vol. 8-16).

The Symposium on Bilingual Lexicography, Paris

Thierry Fontenelle

Thierry Fontenelle is a computational linguist and lexicographer with Microsoft's Natural Language Group in Redmond (USA), where he is responsible for the French lexical database used in a variety of NL applications. He is also currently President of the European Association for Lexicography (EURALEX) and an Associate Editor of the *International Journal of Lexicography*.
thierryf@microsoft.com

Thomas Szende is co-director of INALCO's Central and Eastern Europe Department and of the Research Centre. He has initiated and organizes the international conferences on bilingual lexicography, and coordinates various lexicographic, translation, pedagogic and scientific projects, as well as cultural exchanges worldwide. He has written, edited and published extensively, particularly in the areas of dictionaries, lexicology, translation, and language teaching and learning.
thomas.szende@inalco.fr

The *4èmes Journées d'études sur la lexicographie bilingue* were held in Paris from 22 through 24 May 2003. Beautifully organized by Thomas Szende, who teaches at the Institut National des Langues et Civilisations Orientales (INALCO), the symposium focused on French in bilingual dictionaries and offered over 40 papers covering a wide range of topics and languages. Given the theme of the conference, presentations were in French, but the variety of bilingual dictionaries discussed was impressive, with languages ranging from English to Chinese, Arabic, Dutch, Hungarian, Italian, German, Danish, Norwegian, Swedish, Luxemburgish, Portuguese, Yiddish, Slovak, Romanian, Ladino, Turkish, Gbaya, Swahili, Bulgarian, Serbo-Croatian and Malagasy.

For space reasons, it is not possible to list the titles of all the papers, which means that highlighting some of them is somewhat subjective. The organizer managed to attract a number of established figures in the field of bilingual lexicography, such as Paul Bogaards, who discussed L2 production with a bilingual dictionary, Alain Duval, who convincingly and elegantly analyzed the expression "appeler un chat un chat" (call a spade a spade), or Marie-Hélène Corréard, who discussed French as a target language. Other topics were also dealt with, such as lexical-semantic relations in specialized dictionaries (Jeanne Dancette), cultural aspects in specialized bilingual dictionaries (John Humbley), or the difficulty of defining scientific terms in rapidly-changing disciplines (Yves Gentilhomme). Surprisingly, few papers alluded to the use of corpora or computational techniques for compiling dictionaries. This theme was not entirely absent, however, as was shown by Serge Verlinde, Thierry Selva and Jean Binon's presentation of their work on a computerized, corpus-based French business dictionary featuring an impressive amount of lexical-semantic, syntactic and collocational information accessible via a large number of paths.

The relationship between computers, lexicography and cyberspace was also the topic of a panel discussion. This round table, moderated by Jean Pruvost (University of Cergy-Pontoise), made it possible for the panelists to explore this relationship from the point of view of the dictionary publishing world (Laurent Catach, Le Robert; Ralf Brockmeier, Larousse), of the software world (Thierry Fontenelle, Microsoft Natural Language Group), or of the academic world (Christine Jacquet-Pfau, Collège de France; André Le Meur, University of Rennes 2; Thierry Selva, University of Leuven). The topics discussed ranged from the contents of electronic dictionaries for natural language processing to the dichotomy between electronic dictionaries for people and those for machines, as well as the increasing interest in standardization issues faced by publishers who wish to exchange data with partners. Such issues cannot be ignored, since lexical data providers who are competitors today might be partners tomorrow, and some kind of standardization is certainly desirable, as was energetically pointed out by Marie-Jeanne Derouin (Langenscheidt). Only time will tell to what extent such emerging standards for the representation of lexicographical data will gain acceptance, of course.

In a nutshell, the *4èmes Journées d'études sur la lexicographie bilingue* gave an interesting overview of bilingual French lexicography today. This was an excellent opportunity to meet representatives from the publishing world and from the academic world, and to explore some of today's hot topics. Dr Szende deserves a special note of praise for his successful organization of this event hosted by the Institut Hongrois. I personally look forward to seeing the proceedings of these "Journées", which will be published by Honoré Champion, like the preceding editions. I also definitely look forward to going back to France next year and hope the participants of this symposium will be eager to share their experience and expertise with a larger audience at the next Euralex congress in Lorient (Brittany, France) in July 2004.

The Seminar on Computer-Mediated Lexicography, Castelló

Maria Carmen Campoy Cubillo

The seminar on Computer-Mediated Lexicography was held on 19-21 May 2003 at Universitat Jaume I (Castelló, Spain). Its aim was to offer a forum for discussion on how the latest developments in computerised dictionaries may challenge and change current practices in the language learning context, both in terms of content and technological advances.

Among the issues discussed, criteria to describe computer-mediated dictionaries according to the features that differentiate them from paper dictionaries was a major concern. An important aspect is that of typology, which is undergoing significant changes, particularly for online dictionaries. Dictionary typology is in need of a major revision that takes into account new computer-mediated products, and these may in turn be studied according to agreed criteria, such as genre boundaries, degrees of customisability, and functionality.

The issue of customisability was examined in relation to the teacher's and the student's roles (by Krajka). On the one hand, students may decide which dictionary and look-up modes are suitable for their purposes and styles, thus promoting learner autonomy. On the other hand, the possibility for the teacher to decide on which customisation to use for a particular learner group could be of interest within syllabus design. These perspectives are not excluding, in both cases computer-mediated dictionaries were seen as flexible tools. Although training in dictionary skills still largely depends on the teacher, now it also involves the program designer (who may be a teacher, a computer expert, or both), and on how new features are implemented in order to be taught and learned.

It was also argued (by Tono) that the electronic dictionary interface may reduce difficulties posed by paper dictionary macrostructure, which seems to be the case for hand-held and CD dictionaries. Also, computers allow for a much easier tracking of users' look-up behaviour as well as of note-taking possibilities that appear mostly with CD dictionaries.

However, this is not the case for online dictionaries, some of which provide very complex designs that add new features to the traditional macrostructure, such as links to various databases, educational pages, new terms, simultaneous search in several reference works, links to topic-related pages, etc. Likewise, as pointed out by Luzón, online dictionaries have developed functionality features such as informational and interpersonal interactivity, which traditional dictionaries could not provide and which may turn dictionary consultation into a very complicated task.

In any case, it was clear that there is a need for both longitudinal and contrastive studies. Nesi's proposal (2000:108) "to investigate the use, not of purpose-built value glosses for a few selected texts, but of flexible support for all texts in the form of independent electronic dictionaries," could be extended to all kinds of computer-mediated dictionaries both for value and signification glosses as described by Roby (1999). Research is needed, I believe, not only in the use of actual dictionaries (as opposed to partial ones designed for specific tasks) but also in the promotion of dictionary-use skills as part of metacognitive learning strategies. In so doing, dictionaries should be considered as part of the learning process rather than as a learning tool for solving problems or a compensatory strategy.

Research on computerised dictionaries should also be concerned with aspects that are typical in the new genres. For example, taking into account that sound files are included in most products, and that the majority of learners find the sound option attractive, research should be carried out on how audio files are used – not only to know how a word is pronounced, but to understand spoken texts.

For most (young) learners, the integration of dictionaries into the world of computers has endowed them with a new attraction. This engaging new look appears to be leading to a more frequent enjoyable use and greater familiarity with the dictionary, while helping learners improve their vocabulary and language knowledge. Learners' motivation should be taken advantage of on the side of the teachers, while encouraging them to examine their students' needs and the new ways to present them with lexical information.

References

Nesi, H. 2000. 'On screen or in print: Students' Use of a Learner's Dictionary on CD-ROM and in Book Form.' In P. Howarth and R. Herington (eds.), *EAP Learning Technologies*. Leeds University Press. pp, 106-114.

Roby, W.B. 1999. 'What's in a gloss?' *Language Learning & Technology*, 2.2. http://llt.msu.edu/vol2num2/roby

Maria Carmen Campoy Cubillo teaches English and ESP at Universitat Jaume I. She initiated and coordinated the seminar. campoy@fil.uji.es

Lectures at the seminar:
• Computer-mediated dictionaries: an insight into online dictionary features. Mª Carmen Campoy • Digital genres: transforming and adding value to the dictionary genre. Mª José Luzón • Making ends meet: dictionaries, learning and the information age. Ilan J. Kernerman • Research on the use of electronic dictionaries for language learning: methodological considerations. Yukio Tono • Pragmatics and cognitive linguistics: practical applications in LSP lexicography. Santiago Posteguillo • Computer-mediated dictionaries as teaching and learning tools. Jarek Krajka • ADELEX: using computer-mediated dictionaries online to enhance vocabulary acquisition. Mª del Mar Sánchez Ramos and Carmen Pérez Basanta • Phonetic keywords in EFL dictionaries revisited. Wlodzimierz Sowkowiak • The acquisition of English in a multilingual and multicultural world: the use of multi- and bilingual dictionaries by bilingual learners. Pilar Safont and Mª José Esteve

The International Schools on Lexicography, Ivanovo

Olga Karpova

Olga Karpova is the Head of the English Philology Department at Ivanovo State University and the university's Vice-Rector for Public Relations. She has written extensivley on lexicography and dictionaries, received foreign and local awards and grants, participated in many international conferences, and since 1995 she organizes the lexicographical schools at Ivanovo and edits their proceedings. Professor Karpova specializes in Shakespeare dictionaries and translation, and her main interests include authors' lexicography, English stylistics and discourse analysis.

omk@ivanovo.ac.ru

The idea of holding International Schools on Lexicography at Ivanovo State University goes back to 1992, after the EURALEX Congress in Tampere. In spite of the fact that the Euralex congresses are held every two years, to our regret, only few Russian scientists attend them due to financial reasons. That is why we decided to organize International Schools on Lexicography regularly every two years, between Euralex congresses, in Russia and to host them at our university, where we have stable lexicographic traditions.

The topic Contemporary Problems of Theoretic and Applied Lexicography at the English Language Department is one of the leading themes of research at Ivanovo State University. Since 1996 ten postgraduate thesis and fifty final projects have been defended on general and special purpose dictionaries: authors', learners', several LSP (architecture, stock market and finance) and new words dictionaries, etc.

Special courses in modern and historical lexicography are delivered at the department, with about 25 students annually attending and writing course projects in lexicography. There is a student scientific society in lexicography, LEXICOGRAPHICA, that works quite successfully. Being a collective member of Euralex, Ivanovo State University regularly receives the *International Journal of Lexicography*, as well as the latest news about conferences, workshops and publications.

The aims of the Ivanovo Schools in Lexicography (which have acquired in the course of time the status of an international conference) are to give young scholars a good opportunity to learn current tendencies in foreign and Russian lexicography through plenary lectures of invited prominent scholars and to present their own research results and dictionary projects, surveys, etc.

Our University has already held four Schools on Lexicography, with about 160 participants on average, hosting world famous lexicographers and scholars:

1995: **Theoretical and Practical Aspects of Lexicography**, guest lecturer: R.R.K. Hartmann.

1997: **Problems of Bilingual Lexicography**, guest lecturer: Krista Varantola.

1999: **The Dictionary in the Contemporary World**, guest-lecturer: Ruth Vatvedt Fjeld.

2001: **Language. Culture. Dictionaries**, guest lecturers: Svetlana G. Ter-Minasova and Kenneth A. Haseley. (http://lcd2001.narod.ru)

The results of the Schools' work were published in the form of proceedings and conference materials.

The last conference aroused great interest in Russia and abroad. Professor Ter-Minasova, Dean of the Faculty of Foreign Languages at the Moscow State University, delivered three lectures: The Life of a Word in Speech and Dictionaries; Teaching Foreign Languages – Science, Art or Politics?; and, Linguistic and Extralinguistic Problems of Intercultural Communication. Professor Emeritus of Ivanovo State University, Kenneth A. Haseley, presented two reports: The New Communicators: What Research Can Teach Us About Effective Communication; and Handling Questions and Answers: Guidelines for Comfort and Success. There were also plenary lectures by prominent Russian scholars and guest-lecturers from Croatia. Besides the outstanding representatives of Russian lexicography there were senior university teachers and post-graduates who got an excellent opportunity to share the results of their scientific work, get them valued and approbated. Above all, the lectures were attended by the senior and junior students of regional universities, who also took great interest in the conference. Altogether, 26 plenary lectures and 80 session reports were delivered during the three days of the conference.

The 5th International School on Lexicography, with the subject, **Theoretical Lexicography: Modern Tendencies of Development**, will be held on September 8-10, 2003. Guest lecturers include Dmitry Dobrovolsky, Moscow State University, and Bertha Marie Toft, University of Southern Denmark.

The conference will include sessions on Historical Lexicography, LSP Lexicography, Computer Dictionaries, and New Dictionary Projects. The working languages are English and Russian. The Organizing Committee is planning to publish the conference proceedings.

The Third ASIALEX International Congress, Tokyo

Shigeru Yamada

It is our great pleasure to host the Third ASIALEX International Congress in Japan this summer. Soon after Professor Minoru Murata was elected President at the Seoul conference two years ago, the organizing committee was set up. The 29 members have been working hard toward a successful conference under his leadership. Professor Murata retired from Chiba University at the end of March, but will be denied the promised luxury of free time until the end of the conference.

The conference takes place at Meikai University in Urayasu, located between Narita Airport and Tokyo, close to the site of Tokyo Disneyland. The Asialex Secretary, Yukio Tono, teaches at this university, and holds much responsibility for this event on his shoulders. There are expected to be no financial difficulties at all with Professor Murata on our side. He has never been in the red throughout his long career, in which he served as treasurer for many societies and conferences, including AILA 1999 Tokyo.

We are proud of the program of the conference – of its content, quality, and variety. There will be six symposiums, 60 papers and 13 poster presentations, as well as exhibitions by dictionary publishers and software companies. The conference will feature plenary lectures by Henri Béjoint, Adam Kilgarriff, Tokihisa Kurashima, Kosei Minamide, and Anne Pakir. In addition, we are glad that Reinhard Hartmann and Tom McArthur have agreed to join in the line-up with the following lectures:
• Hartmann: Why lexicography needs a strong academic foundation
• McArthur: What on earth is English – a world, an international, or a global language?

The opening ceremony will include an award ceremony for the Kernerman Dictionary Research Grants, which was won by the project "A glossary of essential academic vocabulary" by Dr. Jacqueline Lam Kam-mei and her team, from Hong Kong University of Science and Technology. Prior to the conference, the Brighton intensive computer lexicography course will be offered by Adam Kilgarriff and Michael Rundell. This course is held at Senshu University, Tokyo, on 25-26 August, at a discount rate for the conference participants.

Asia is a vast area with diverse peoples, languages, and ancient traditions of dictionaries. Asialex should function as an institution and forum which brings people together to discuss, study, and promote lexicography within and beyond Asia. To serve its purpose and people, the latest addition to the "lex" family must grow as an academic association. Such issues as membership, fees, and the publication of a journal should be discussed in the business meeting at the conference.

All in all, this is going to be a busy and exciting week. Preparation is well underway, and we look forward to seeing many of you at Meikai.

The ASIALEX Landmarks
• The Asian Association for Lexicography was established in March 1997, during the Dictionaries in Asia conference that was held at the Language Centre of Hong Kong University of Science and Technology, at the initiation of Gregory James and Ami Chi.
• The major objective of Asialex is to foster scholarly and professional activities in the field of lexicography in Asia.
• Several papers from the inauguration conference were published under the title *Lexicography in Asia* (eds. McArthur T. and I. Kernerman, Password Publishers, Tel Aviv, 1998).
• The first Asialex international conference was held at Guangdong University of Foreign Studies, Guangzhou, China, in 1999. It was organized by the first Asialex President, Professor Huang Jianhua, and had two major themes: 'National experience in lexicography or dictionary compilation', and 'Bilingual lexicography'.
• The second Asialex international conference was held at Yonsei University, Seoul, Korea, in 2001. It was organized by the second Asialex Presdient, Professor Sangsup Lee, and had the major theme of 'Asian Bilingualism and the Dictionary'.
• The proceedings of each Asialex conference are available from its organizers.

Shigeru Yamada studied lexicography at the graduate schools of Waseda University, Tokyo, and the University of Exeter, UK. His interests include pedagogical and bilingual lexicography. He contributed as a lexicographer to *Luminous Japanese-English Dictionary* (2001) and *Kenkyusha New English-Japanese Dictionary* (6e, 2002). He is currently an associate professor at Waseda University, and the Asialex Treasurer.
shyamada@waseda.co.jp

•

The Third Biennial Conference of the Asian Association for Lexicography (ASIALEX) is entitled '**Dictionaries and Language Learning: How can dictionaries help human and machine learning?**'. It will be held at Meikai University, Urayasu, Chiba, Japan, 27-29 August 2003. For registration and details, please visit the conference website: http://www.asialex.org.

A Lifestory in Dictionaries

Peter H. Collin

Peter Hodgson Collin served in the Royal Navy, then studied modern languages in Oxford. He worked as a lecturer in Canada and Hong Kong, edited dictionaries and directed dictionary publishing with Harrap and Macmillan in London, and in 1985 founded his own dictionary publishing house that was sold last year to Bloomsbury. Altogether, Mr Collin has been involved in the making of scores of dictionaries in the UK and in many languages worldwide. At present he is the Chairman and Editorial Director of Modern English Publishing.
peter@modernenglishpublishing.com

Dictionaries written by P.H. Collin

Published by Harrap

Easy English Dictionary
Easy Learning English Dictionary
Standard Learner's Dictionary
English Vocabulary
English Mini Dictionary
English Pocket Dictionary
English School Dictionary
Mini Pocket English Dictionary
2000 Word English Dictionary
2000 Word Engels-Nederland Dictionary
2000 Word English German Dictionary
Dictionnaire de 2000 mots
Dictionnaire de l'anglais elementaire
English-Arabic Basic Learner's Dictionary

As often happens, I got into what turned out to be my life's work—working on dictionaries—by accident.

When I left Oxford with a degree in modern languages, I wanted to work with books, and so wrote to several publishers. I got a job almost immediately with Harrap, at that time the leading publisher in school textbooks for language learners, and also, incidentally, an important bilingual dictionary publisher[1].

In 1958 I started work in the Harrap offices, at the top of the old building in High Holborn, up flights of back stairs (the editorial staff weren't supposed to use the main stairs or lift) in a room about 6x6m, with two small windows looking out over rooftops, where four or five of us schoolbook editors and a secretary worked in Dickensian squalor. The first chore every morning in winter was to light the small and very smelly gas fire. Desk lights were on all day, as the windows were small. Although I was employed to work on school textbooks (my first jobs were to edit a new Russian course for beginners and a university edition of a French medieval text), very quickly I found myself helping out with the new Supplement to the big Mansion French/English dictionary, adding new entries, looking for translations for new words, making up examples, and eventually proofreading. After a few more months I was working on it more than half-time, and the schoolbooks receded into the background.

After the third supplement came out, the decision was taken to produce a whole new edition of the Mansion dictionary, and this was organized partly in the office, with several editors (one of whom, Françoise, I married), and by freelancers all over the place: there was a specialist who dealt only with names of birds; there was one who supplied the phonetics for the French side; there was a lady in Pau who sent in masses of handwritten notes of things she had read in French newspapers; there was an elderly Polish refugee whose very fiddly job it was to cut up a copy of the existing dictionary and stick each word entry down on a filing card, so that new entries could be added on new cards. The cards were stored in drawers in wooden racks. In those days (the early 1960s) there were no computers, although we had demonstrations of card-based inventories which we thought might be useful. In the end, we kept to the old system, and sent the cards off periodically

to be typeset (using metal setting, of course); our printers employed typesetters who worked only on that dictionary.

After a while my wife and I decided that we wanted to travel, so I got a job teaching in Canada (Dalhousie University, Halifax, Nova Scotia), and after five years there, went to teach at the University of Hong Kong for another five years. But we still kept in contact with Harrap, and proofread dictionaries all the time (1963-1972).

Then I was asked if I wanted to go back to Harrap, firstly to run the schoolbooks department, and then to take over as director of the dictionary publishing as well.

Although we were now in the mid-1970s, there were still no computers involved in the compilation processes, but some very advanced printers were typesetting from punched tapes. I remember discovering that the paper tapes of one of our dictionaries had been thrown away by the printers once the pages had been typeset, and I went down to their offices and searched through the dustbins to retrieve this mass of paper ribbons, which I felt somehow could be used (and indeed they were converted to very early computerized data).

During this stage of my work at Harrap, not only did we produce new editions of all the major bilingual dictionaries, but I also launched a new series of smaller bilingual dictionaries in Spanish and German, plus the first of several monolingual English learner dictionaries. These spawned a small series of semi-bilingual dictionaries in French, German, Dutch, Afrikaans, Arabic[2] and other languages.

In 1983, I was approached by Macmillan Inc (New York) to see if I would be willing to set up a whole new dictionary division in London to produce a series of bilingual dictionaries in several languages, simultaneously. The languages were to be English, French, Spanish, and German, with the possibility of adding Portuguese and Italian later. The dictionaries were intended to replace the old Cassell dictionaries (owned by Macmillan at that time). The whole project was to be computerized, and Macmillan were prepared to put a large amount of money into it. So I moved from Harrap, and started all over again.

Firstly we needed to find offices, and then staff, and we were lucky to find main language editors who all had dictionary experience; we then recruited numbers of junior editors and opened our main office

in Golden Square, with little offices in various parts of the world; compilation started on source files for each language and within two years we employed around 90 full-time and part-time staff. Although the different languages had different grammatical requirements, we were able to draw up standardized entry forms and compilation was done on paper while the decision on the computer system still had to be made. After two years' compilation most of the source files were more or less complete and were ready to be transferred to computer.

At that stage, I felt that my role was becoming less relevant, and I also wanted to start up something on my own, which I would actually own. So I left Macmillan and formed a new company, Peter Collin Publishing Ltd (or PCP for short) in 1985. I knew exactly what dictionaries to produce, because I could see that there was a gap in the market for English learners' dictionaries in the specialist subjects which were becoming widely taught as part of ELT courses, in particular Business, Legal and Medical English.

Because I wanted the material to be computerized from the start (as opposed to Macmillan where the choice of a computer system took about two years) and I didn't want to use any of the existing compilation systems which were both expensive and derived from other large dictionary projects, I devised my own internal coding system for each part of an entry (headword, phonetics, part of speech, definition, example, quotation, etc.) and bought my first small computer (an Apricot, which in those days had no memory, so that everything had to be stored on diskettes). The great day was when the first draft text was sent to a typesetter for testing and came back exactly as we wanted: this showed that our internal codes could generate typesetting in multiple fonts, different point sizes and with boxes to highlight parts of the text. It also gave us the flexibility to change the appearance of the dictionary page at will.

Compilation of the first two dictionaries (Business and Law) took about six months, so that for the Frankfurt Book Fair in 1986 my wife and I actually had two books on our little stand. Shortly afterwards, we signed agreements with Klett in Germany and Norstedts in Sweden, to use our data to make our first bilingual dictionaries. Again, the simple coding system proved invaluable in compiling bilingual versions, and diskettes of our texts were sent all over the world to have translations added: it proved surprisingly simple to take the original English text and add translations

to make semi-bilingual dictionaries. This process continued over several years, with dictionaries appearing in Chinese, Slovenian, Polish, Czech and other languages, all compiled locally using our material. It was interesting to go to one of the Chinese publishing houses (the China Financial Press, Beijing) and see the text which had been imported from our diskettes appearing on screen, and the Chinese translations being added as we watched.

We also decided to make our own big bilingual dictionaries in French, German and Spanish, for which I recruited specialist translators in various parts of the world.

All this required more and more computer skills, and my son Simon left his job as the technical editor of a computer magazine to join the family company and help run the business. By our 17th birthday in 2002 we had over 100 titles, mainly dictionaries, in fifteen different languages, and sales in almost every country of the world. However, it was becoming more of a strain to fund the new projects, and without new projects I felt that the company could not continue to develop. We had the distinction of being the only privately owned English dictionary publisher, and this made the company attractive to others. It was not unexpected, but a surprise nevertheless, when Bloomsbury, the publisher of the Encarta dictionaries (and Harry Potter!), suddenly made us an offer out of the blue. The PCP dictionaries have now moved to Bloomsbury and form the basis of the new Bloomsbury Reference list. They are being actively updated, and I am still involved to a certain extent in this process.

Because I didn't want to retire, we decided to move into magazine publishing, and set up a new company, Modern English Publishing Ltd[3]. We were lucky to acquire three ELT magazines, alongside which we are publishing a series of practical guides for teachers of English as a Foreign Language.

There is life after lexicography, after all!

Notes

1 The company was called at the time George G Harrap Ltd, then Harrap Ltd. Now it is part of Chambers Harrap Publishers Ltd, based in Edinburgh, owned by Larousse PLC.
2 *Harrap's Standard English Learner's Dictionary* served as a base for Kernerman's semi-bilingual Arabic (1987), Italian (1987), Chinese (1990), and Spanish (1991) editions.
3 www.modernenglishpublishing.com

Peter Collin Publishing

Dictionary of Business • English, French, Swedish, German, Spanish, Slovene, Chinese, Polish
Dictionary of American Business
German Business Dictionary
French Business Dictionary
Business Spanish Dictionary
Le Bilingue des Affaires
Dictionary of Marketing • English-German
Dictionary of Personnel Management
Dictionary of Human Resources
Dictionary of Banking & Finance • English, German, Chinese
Dictionary of Accounting
Dictionary of Hotels & Tourism • English, German, Chinese
English Law Dictionary • English, Swedish, German
Hungarian Law Dictionary
Chinese Law Dictionary
Spanish Law Dictionary
Dictionary of Government and Politics
Dictionary of Printing & Publishing • English, German
Dictionary of Medicine • English, Swedish, German
Dictionary of Ecology and the Environment • English, German, French
Dictionary of Agriculture • English-German
Dictionary of Horticulture
English Dictionary for Students
Crossword Key Dictionary
English Study Dictionary • Italian, Spanish
Children's English Dictionary

Published worldwide

• Difusion, Spain • DZS, Slovenia • Universal Dalsi, Romania • Klett, Germany • Financial Press, China • FLTRP, China • Kel, Argentina • Larousse, France • Norstedts (Esselte), Sweden • Panem Grafo, Hungary • P&R Centrum, Czech Republic • Southern Publishers, South Africa • Wilga, Poland • World Publishing Corp, China

The Kernerman Dictionary Research Grants -

Three applications for the Kernerman Dictionary Research Grants have been accepted for 2002. The Assessment Committee of AFRILEX awarded a grant to Karen Hendriks, from the University of Stellenbosch, for her research on the particular needs of a multilingual society and their effects over the structure and nature of bilingual dictionaries, and a grant to P.M. Lubisi, from the University of Pretoria, for his project to build a major corpus for siSwati, one of the official Bantu languages of South Africa. The ASIALEX Committee awarded a grant to Jacqueline Lam Kam-mei and her team, from the Hong Kong University of Science and Technology, for their project to create a glossary of essential academic vocabulary. The summaries of these applications are presented herewith.

A Glossary of Essential Academic Vocabulary

Jacqueline Lam Kam-mei with Sue Chang and Gregory James

Research has shown that word-lists have an important role to play in acquiring a language other than the mother tongue. At the university level, students need to be equipped with a sufficient vocabulary to follow lectures, comprehend academic texts, participate in seminar discussions, and write coherent research papers. Learners need to be familiar with at least 95% of the running words in a text if they hope to achieve these ends. Failure to acquire the core academic vocabulary for their field will adversely affect their work. A well-selected word-list based on academic texts used in tertiary institutions is therefore essential to help new entrants to acquire as many vocabulary items as possible in a relatively short time, so that they can cope with the heavy university student load once they have enrolled.

Based on the afore-mentioned rationale, a team of researchers, directed by Gregory James, at the Language Centre of the Hong Kong University of Science & Technology, have compiled five properly sampled electronic datasets of texts extracted from first-year textbooks used by university students in Hong Kong since 1991. These corpora, which have now been completed, comprise c.1,000,000 words in each of the following disciplines: (i) Computer Science, (ii) Business Studies & Economics, (iii) Biology, Biochemistry & Chemistry, (iv) Engineering, and (v) Humanities & Social Sciences.

The proposed project represents a coherent and long-term strategy aimed at providing undergraduates and potential university students with an academic word-list which will enhance their knowledge of academic vocabulary, thus improving their ability to cope with a three-year university programme.

The project aims, firstly, to identify a list of frequently occurring academic words appearing across the five corpora mentioned. This list (hereafter the HKUST Wordlist) will contain those content words most useful for academic reading comprehension, discussions and paper writing. The second aim is to develop a glossary, based on the HKUST wordlist, that will include key features such as definitions, examples in context, and examples of usage informed by relevant corpora and pronunciation. Other features will be considered and included subsequently as a result of feedback from user-perspective surveys.

All tertiary-level students, but first-year undergraduates in particular, in Hong Kong would benefit from this project through a systematically planned vocabulary learning scheme based on the HKUST Wordlist. Another group benefiting from this project would be students in Secondary Forms 6 and 7 intending to pursue further studies at tertiary level in the Science and Technology streams. Through using the proposed HKUST Wordlist, students could better equip themselves to make the often difficult transition from general or secondary academic reading/writing to tertiary academic reading/writing. More importantly, the proposed HKUST Wordlist will serve as useful reference material for language materials designers, when seeking to produce appropriate language learning and teaching materials to enhance their students' language proficiency.

The project has three main phases:

Phase One:

1. Identifying and generating word-lists including single words, compound words and distinctive phrases (bi-grams, tri-grams, etc.) in each corpus;

2. Identifying words and n-grams that appear in all five corpora and generating a unified word-list across five disciplines;

3. Devising and implementing measures to confirm the reliability of the word-list compiled.

Phase Two:

4. Designing and compiling the prototype HKUST Wordlist in both paper and electronic form by

4.1 determining headwords;

4.2 writing definitions;

4.3 illustrating definitions through appropriate phrases and sentences from the corpora;

4.4 exemplifying word usage;

4.5 including phonetic symbols for pronunciation help;

5. Designing and engaging in user-perspective surveys;

6. Collating and analysing survey feedback.

Phase Three:

7. Designing and compiling the HKUST Wordlist;

8. Developing training materials and vocabulary acquisition workshops and seminars, based on the HKUST Wordlist, to ensure students learn the words.

Once the framework for the HKUST Wordlist has been developed, it is envisaged that word-lists for learners at different levels and disciplines, for example, more discipline-specific lists for undergraduates, could also be compiled for use.

s - Accepted Applications for 2002

Certain Aspects of Bilingual Dictionaries in Multicultural and Multilingual Societies

Karen Hendriks

South Africans live in a culturally and linguistically diverse society with eleven official languages. The country has a complex political history, which cannot be severed from sensitive issues concerning language and communication. The process of acknowledging the truth and working towards reconciliation has been on its way for several years, yet the people of South Africa still have a long way to go in order to be able to genuinely celebrate their diversity rather than be threatened by it. I believe that lexicography, and specifically, good bilingual dictionaries, can make a vital contribution to the process of reconciliation, of affirmative action, and of building a nation out of the diverse speech communities and cultures in South Africa.

Bilingual dictionaries play an extremely important role in a multilingual society; they can be perceived as the key instruments in the communication between different communities. Efficient and active communication between different cultural groups is essential in the South African environment. Culture-specific lexical items, and the way they are treated in bilingual dictionaries, can have a great influence on this process. In the past, South African dictionaries have displayed a strong cultural bias in the treatment of these items. Today, lexicographers have to face the reality of the lexicographic wrongs of the past, and correct this in dictionaries for the present-day user.

My study will focus on the way the particular needs of a multilingual society imply certain adaptations to the structure and nature of a bilingual dictionary. In a multilingual and multicultural context the lexicographer may not rely on the intuition of the user, and the needs of different user groups have to be accommodated. The lexicographer has to keep in mind that the user of a general-purpose bilingual dictionary may well also be a learner of the target language.

Furthermore, I plan to examine the admission and treatment of culture-specific lexical items in bilingual dictionaries intended for a multilingual environment. Considering the vitality of enhancing clear and effective communication between speakers of different languages, the misrepresentation of culture-specific lexical items in bilingual dictionaries could seriously impede communication rather than enhance it. It is of great importance that lexicographers have adequate guidance in the treatment of these items.

South Africa is not the only country with a multilingual and culturally diverse society. Users worldwide could benefit from a comprehensive study of the way dictionaries can contribute to and enhance communication between different cultural groups. I intend to examine the situation concerning this issue in South Africa and then to make suggestions and provide guidelines for the treatment of culture-specific lexical items in bilingual dictionaries that would apply to any multilingual environment.

Corpus Building for the SiSwati Dictionary Unit

P.M. Lubisi

SiSwati is one of the 11 official languages in the Republic of South Africa, spoken by about one and a half million people. It is, however, one of the least developed and underprivileged of these languages. There is not a single monolingual dictionary in siSwati, and the terminology development of this language is in its infancy. The first matriculants in siSwati were in 1987, and there is still only a small number of published materials, most of which are based on the school curriculum. SiSwati is offered at only two universities in South Africa, though not at any teacher training college.

Although the government, through the Pan South African Language Board, is in the process of establishing lexicography units, languages like siSwati will still be disadvantaged because of the lack of human capacity and resources. The siSwati lexicography unit is housed in a technicon that does not offer the language as a course, and this unit will receive little academic assistance. Lexicography is foreign in many technicons, including the Pretoria Technicon, let alone corpus building. SiSwati has no professional lexicographers, and there are hardly any students who hold senior degrees in this language. Hence, it is developing at a snail's pace. Assistance is afoot in the collection of printed data, yet no attempt has been made to collect oral data from the siSwati radio stations, so a grant of this nature can play a pivotal role in the building of such a corpus. Moreover, it will allow enough time to do thorough research in this regard.

The aim of this project is to carry out a thorough research on corpus building for siSwati, by way of collecting data, especially oral data and information stored on tapes at the only two radio stations, namely Ligwalala FM in South Africa, and the Swaziland Broadcasting Corporation. There is a mammoth task to be done to retrieve this information as the unit does not at present have the necessary software to deal with this matter. The first stage of the project will concentrate on the oral literature, and will be conducted in the rural part of Mpumalanga lowveld, and in the four districts in Swaziland. A good tape recorder will be needed to record the oral data. The researcher will be assisted by veterans, like Professor Daan Prinsloo from the University of Pretoria, who have already made inroads in this field. The university's department of African languages has accepted this project, and is actively participating in the building process and offers full access to all its resources and expertise. In addition, a major publishing house (Macmillan) has agreed to permit the of use their siSwati material. The collection of oral data is expected to gain full momentum by the receipt of the grant. The study is invaluable for the development and compilation of siSwati dictionaries in that it will be the first of its kind in this language. The gained know-how will be used also to assist other units that are still struggling to build their corpus.

The Oxford Student's Dictionary for Hebrew Speakers, Third Edition
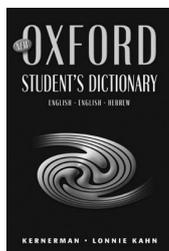
**Oxford Student's Dictionary
English-English-Hebrew**

Kernerman Publishing and
Lonnie Kahn Publishing

ISBN 965-307-034-7
Paperback, 984 pp.
150x230x48mm.

Based on *Oxford Student's
Dictionary*, OUP, 2001

Hebrew edition: July 2003
Traslation: Ya'acov Levy
Editor: Raphael Gefen

OSDEEH includes:
Entries consisting of
• headwords
• definitions, in English
• parts of speech of the
headwords, together with
other grammatical
information (*countable/
uncountable nouns,
transitive/intransitive verbs,*
etc.)
• idioms
• example phrases and
sentences
• the Hebrew translation of
the headwords (and where
necessary also of idiomatic
expressions found in the
example sentences)
• usage notes
• opposites and synonyms
• register and style (*formal,
informal, technical*)
Charts, diagrams and
pictures
Appendices (irregular
verbs, affixes, geographical
names, expressions using
numbers)

The new edition of *Oxford Student's Dictionary English-English-Hebrew* is based on the third edition of *Oxford Student's Dictionary* (OSD, Oxford University Press, 2001), and as such is a most up-to-date dictionary for learners of English as a foreign language at the upper-intermediate level, suitable for the upper grades of secondary (high) schools, universities, the professions and the general public.

The OSD has been adapted for the use of Hebrew speakers, not only by translating all the entries and subentries, but also by making it particularly relevant to Israeli users, e.g. by adding a large number of entries and subentries (especially words commonly used in Israeli educational institutions and local textbooks).

The former editions of this dictionary have been in use for over 25 years as auxiliary material in class and in examinations. When the semi-bilingual dictionary first appeared, it was soon recognized to be far more useful than the monolingual OSD used in class previously, which had been hardly referred to by examinees or indeed by students in their regular lessons. Hitherto, students had tended to use a bilingual dictionary, but the advantages of the semi-bilingual type were immediately evident, with the explanations in simple English, i.e. the language of instruction in the communicative framework of the English curriculum and methodology, and the numerous examples (after all, "meaning lies in the context")—features of the monolingual dictionary—but now also with the satisfying and confirming existence of the Hebrew translation of the headword.

This edition has been totally rewritten, taking into account advances in linguistics, teaching methodology and lexicography, and covering areas such as hi-tech in general and computers in particular, as well as business studies, the sciences and professions, education, literature and the arts, tourism and communication needs in general. This dictionary has over 47,000 references. The definitions and examples are in British spelling, but the American equivalents are also given. Thus, *faucet* is given as the American equivalent of *tap*, both as part of the entry for *tap* and as a separate entry.

Raphael Gefen, Editor

Published on January 1, 1986, *Oxford Student's Dictionary for Hebrew Speakers* (OSDHS) was not the first of its kind. Several semi-bilingual dictionaries (SBD) for learners of English at the elementary-to-lower-intermediate level were published before that:

• In 1982, Longman co-published with AO Livro Technico, in Brasil, a 10,000-word SBD, entitled *Longman English Dictionary for Portuguese Speakers*, by Rosa W. Konder.

• In the same year, Harrap published an SBD written by P.H. Collin, called *Harrap's Dictionaire de 2000 Mots, Anglais-Francais*.

• In Beirut, Librairie du Liban published a 20,000-word SBD in 1985: *Al-Mufid, A Learner's English-Arabic Dictionary*, by Nasr and Al-Khatib.

• OUP published that year a 2500-word SBD called *First Dictionary for French-Speaking Africa*.

It is possible that other SBDs appeared before, that have not come to my attention. Other dictionaries I have occasionally heard about turned out to be not semi-bilingual but rather, to use R.R.K. Hartmann's term, "bilingualized", in which not only the headword is translated, but the definitions and/or examples are translated as well.

In Israel, the appearance of OSDHS and, in 1987, *Harrap's English Dictionary for Speakers of Arabic*, marked the first official recognition of the use of the mother tongue in the foreign language classroom. The acceptance was so universal, that it has eventually become the only type of dictionary permitted in the English matriculation examinations, which are themselves geared to this dictionary. Monolingual dictionaries are not permitted in the school system at all, and bilingual dictionaries may be used by students at the lower level of matriculation.

OSDHS has had over 50 reprintings of the two editions (2e published in 1993). Its impact on pedagogical lexicography and English learners' dictionaries goes well beyond Israel, in many countries and languages worldwide.

Lionel Kernerman, Publisher