

# DICTIONARY News

Number 12 | July 2004

## Microsoft and Dictionary Makers: Defining Partnerships

Julian Parish

I am happy to take up the invitation to reply to Joseph Esposito's article, *Dictionaries, another Netscape?* (KDN 10, 2002). Esposito's article is fascinating and provocative in many ways, and I share many of his views about the importance of dictionaries and how they will be used in the future. Where we differ is in his interpretation of Microsoft's interest in lexical data and in what this means for established dictionary publishers.

### Commitment to lexical data

Microsoft, as rightly identified by Esposito, "views lexical databases as an aspect of strategic technology", and thus perceives such data in the broadest sense: the definitions and translation equivalents which are central to printed reference

dictionaries are only part of the lexical data that Microsoft utilises. Information about spelling, pronunciation and grammar is perhaps even more important in many of Microsoft's products. Whether in established businesses like Office or new areas such as Tablet PC or MS Reader, we make extensive use of wordlists – in look-up dictionaries, yes, but also in spellers, handwriting recognizers, and search and speech recognition engines, to give just some examples. Our requirements for lexical data will continue to develop, as we develop new products and add further localized languages to existing products. For Office 2003, for instance, we have added new localized versions for Catalan and Nynorsk.

Furthermore, I share Esposito's view that



Julian Parish is the Business Manager for Dictionary Central at Microsoft's European headquarters in Paris, responsible for coordinating the planning and acquisition of lexical resources across the company. He studied Modern Languages at Oxford, and taught at the Sorbonne.  
[julianp@microsoft.com](mailto:julianp@microsoft.com)

- 1 Microsoft and Dictionary Makers: Defining Partnerships | **Julian Parish**
  - 4 Some Lexicographic Concepts Stemming from a French Training in Lexicology (2-3) | **Jean Pruvost**
  - 6 Lexiculture and the EFL Dictionary | **Anthony P. Cowie**
  - 17 *Dictionnaire du français usuel*: an innovative french learner's dictionary
  - 18 Some Highlights of Contemporary Hebrew Dictionaries and Lexicography
  - 18 Modern Hebrew Dictionaries | **Ora (Rodrigue) Schwarzwald**
  - 23 *Milon ha-Hoveh* and *Milon Sapir* | **Mordechay Mishor**
  - 26 From *Milon ha-Hoveh* to *Milon Sapir* | **Yitzhak Shlesinger**
  - 27 *Rav-Milim* – a Modern Dictionary for an Ancient but Thriving Language | **Yaacov Choueka**
  - 31 *Milon-Kis Ariel* | **Maya Fruchtman**
  - 32 *Milon Even-Shoshan*, Revisited | **Moshe Azar**
  - 34 The Historical Dictionary of the Hebrew Language | **Doron Rubinstein**
  - 37 Glossary
  - 38 A Glance at Porto Editora | **Graciete Teixeira**
  - 39 Lexicography organization founded in Taiwan
  - 40 Kernerman Dictionary Research Grants
- K DICTIONARIES | **Recent Titles**



© 2004 all rights reserved.

Editor | **Ilan J. Kernerman**

Graphic Design | **Studio Orna Cohen**

Thanks to | **Anthony P. Cowie, Raphael Gefen, Doron Rubinstein, Merav Kernerman, Miriam Shlesinger, Nili Sadeh, Lionel Kernerman**

**K DICTIONARIES LTD**  
Nahum 10 Tel Aviv 63503 Israel |  
tel 972-3-5468102 |  
fax 972-3-5468103 |  
[kdn@kictionaries.com](mailto:kdn@kictionaries.com) |  
<http://kictionaries.com> |

Charles M. Levine's 'The Coming Boom in English Lexicography: Some Thoughts about the World Wide Web (Part One)' appeared in *Kernerman Dictionary News*, Number 9, July 2001 (<http://kdictionaries.com/newsletter/kdn9-1.html>).

Joseph J. Esposito's 'Dictionaries, another Netscape?' appeared in *Kernerman Dictionary News*, Number 10, July 2002 (<http://kdictionaries.com/newsletter/kdn10-1.html>).

Levine's 'The Coming Boom in English Lexicography – Reconsidered (Part Two)', appeared in *Kernerman Dictionary News*, Number 11, July 2003 (<http://kdictionaries.com/newsletter/kdn11-1.html>).

#### Comment from

#### Joseph J. Esposito:

"Mr. Parish does not respond to my piece at all. Microsoft's intentions are irrelevant; what is important are its effects, an ineluctable outgrowth of Microsoft's position in the marketplace. Mr. Parish's illustrations all serve to confirm my thesis."

lexical data will increasingly be accessed from within computer applications, whether in machine-readable form or directly by the end-user. During the 1990s many electronic dictionaries were published, often adapted from existing print dictionaries, and marketed as stand-alone consumer products (e.g. Microsoft Bookshelf). That period has passed, as Gilles-Maurice de Schryver notes: "If there is one single feature likely to be applicable to all [electronic dictionaries] of the future, it is that they will stop functioning as stand-alone products." ('Lexicographers' Dreams in the Electronic-Dictionary Age', in *International Journal of Lexicography* 16.2, 2003).

In future, we should expect to see an increasing range of applications which make use of lexical data. Already, these include spellers in products such as Office, Works, Outlook Web, Access and Hotmail, as well as reference dictionaries (both monolingual and bilingual) which may be accessed in the Encarta Reference Library, online on MSN or through the new Research Pane in Office 2003.

#### Partnership in development

The parallel that has been drawn between the development of lexical data and Netscape is, I feel, misleading: it is technology – whether for Internet browsing or search – and not data itself which is central to Microsoft's business. Developing our own lexical content (across more than 40 languages at that) is simply not a part of the company's core mission. Why should we seek to develop so many dictionaries ourselves when excellent resources already exist, developed over many years by a range of dictionary publishers? Those resources, moreover, are already available, whereas any new dictionary would require several years' work to create *ab initio*.

The example of the *Encarta World English Dictionary* may also be misleading, insofar as the development of this dictionary, whilst jointly funded by Microsoft, was entirely directed by Bloomsbury Publishing in London. In practice, it has been rare indeed for Microsoft to develop dictionary content itself, other than in the specific area of IT vocabulary, for which Microsoft Press publishes a specialized *Computer Dictionary* (<http://www.microsoft.com/mspress/books/5582.asp>).

In all other cases, Microsoft has developed its lexical tools with the help of third party specialists in lexical data. These may be publishers with an established background in print-based reference publishing or newer independent

software vendors (ISVs) who have built their businesses supplying lexical data to general software houses such as Microsoft.

In an earlier issue of *Kernerman Dictionary News*, Charles Levine questions what opportunity there is for businesses to make any money from developing new spellers: "Since spell checkers are bundled freely, there is no money to be made and no incentive in developing truly better, more intelligent spell-checking software." ('The Coming Boom in English Lexicography: Some Thoughts about the Worldwide Web (I)', in KDN 9, 2001).

That assertion over-simplifies the case of many Microsoft products: whilst the owner of the lexical data used may not be paid directly by the end-user of the software, that data can be monetized through the license fees ISVs charge to companies like Microsoft. And it is in our interest to continue improving the quality of linguistic tools in new product releases.

The parallel I would suggest is not with Internet browsers, but with the use of mapping data in computer software. Over the past decade this market has changed enormously, with paper maps increasingly giving way to electronic applications, first PC-based solutions such as Microsoft's Autoroute Express, then in-car GPS-based navigation solutions. Those cartographers who think of themselves only as book publishers will certainly see their businesses decline. For those companies, however, who see their value in providing high-quality cartographic data in the required (electronic) form, these applications create new business opportunities. Again, those companies may be existing publishers such as Rand McNally or Michelin, or new providers like NavTech.

#### Continuing our commitment to partnership

As we look to the future, Microsoft sees more, not fewer, opportunities for publishers to provide lexical data to work with our technologies:

First, we will continue to license lexical data for new or existing applications from partners who can offer us high quality resources.

Secondly, we are increasingly creating new opportunities for publishers to develop and market themselves additional products which integrate with our core applications. Examples of these already include the add-on spellchecker files for Office in areas such as law, medicine or economics, which exist today for Dutch, French and Italian. Or again the

Translation Dictionaries technical article (a form of software development kit) for the bilingual dictionaries in Office 2003: this article – which is available free-of-charge – enables publishers of bilingual dictionaries to adapt their existing content and sell it as a module which is fully integrated in Office (see <http://www.microsoft.com/downloads/>

[details.aspx?FamilyId=38934F90-FB06-4ABF-ABA5-94D16BF813BB&displaylang=en](http://www.microsoft.com/downloads/details.aspx?FamilyId=38934F90-FB06-4ABF-ABA5-94D16BF813BB&displaylang=en)).

Dictionaries and other lexical data remain a strategic investment for Microsoft, but one that we believe is based on partnership, not exclusion, an opportunity for dictionary publishers and not a threat.

## Dictionary tools in Microsoft products

### Microsoft Office 2003

- The speller, thesaurus and grammar checker are already an established part of the Proofing Tools in Office.
- It is possible to add support in more than 40 available languages.
- The Language Auto-Detect feature in Word automatically recognizes after a few words the language being used and will switch the speller, thesaurus and grammar checker to that language. Alternatively, the language itself can be specified.
- Words that are not included in the standard speller (e.g. specialized terms or company names) can be added to the user's custom dictionary in each individual Office configuration.
- For French, Dutch and Italian, additional spellchecker files covering specialist vocabulary for science, law, medicine, IT and economics can be downloaded and integrated into the existing speller
- Access to a range of research and reference information is offered without leaving the Office application. The dictionaries available include:
  - *Encarta World English Dictionary*, developed in association with Bloomsbury Publishing, with 100,000 headwords (US and UK versions);
  - *Encarta French Dictionary*, built specifically for Microsoft by a development team in France, 45,000 headwords;
  - a German monolingual dictionary, produced by a leading German dictionary publisher, with 57,000 headwords;
  - bilingual dictionaries for English to and from several languages including French, German, Italian and Spanish.

### More to come

- In 2004-2005 Microsoft will be adding localized versions of Office for many languages, with spellchecking support provided in many cases. These new versions will extend its coverage of the languages of the new member states of the European Union and beyond (for languages such as Macedonian and Afrikaans). Discussions are also under way to offer further specialized spellcheckers for other languages.

### Encarta Reference Library 2005

- It is possible to consult a dictionary without opening up Office. The same dictionary content is available in the latest versions of Encarta, featuring one-click access to definitions, synonyms and translations (the exact mix varies by language).
- Encarta Reference Library 2005 is available for English (in US and UK editions), French, German, Spanish, and Dutch. For English, French and German, Microsoft offers the same dictionaries as in the Microsoft Office 2003 Research Service; for Spanish, the prestigious dictionary of the Real Academia de la Lengua Espanola is included, while for Dutch the dictionary is provided by Het Spectrum.

### Dictionaries for Pocket PC

- The Microsoft dictionaries available in MS Reader format can be downloaded directly to Pocket PC using Active Sync, including a specially shortened version of the *Encarta World English Dictionary* (with concise definitions in English) and bilingual dictionaries for English to and from French, German, Italian and Spanish:  
<http://www.microsoft.com/reader/downloads/dictionaries.asp>.

### Online: MSN

- It is possible to access *Encarta World English Dictionary* on the Internet:  
<http://encarta.msn.com/encnet/features/dictionary/dictionaryhome.aspx>.

### Computer dictionaries in print

- *Microsoft Computer Dictionary* covers computer and IT terminology in English, with over 10,000 entries:  
<http://www.microsoft.com/mspress/southpacific/books/book19087.htm>.
- The same in German:  
<http://mspress.microsoft.de/mspress/product.asp?dept%5Fid=2000&sku=3%2D86063%2D896%2D3>.

## Some Lexicographic Concepts Stemming from a French Training in Lexicology (Parts 2 and 3)

Jean Pruvost



Jean Pruvost 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 lexicographers and lexicographers and lexicographers. 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 series *Que sais-je?* [*What Do I Know?*], one on French language dictionaries, the other on neologisms (with Jean-François Sablayrolles), as well as of numerous articles. Professor Pruvost co-directs, with Bernard Quemada, two collections at Éditions Honoré Champion (*Études de lexicologie, lexicographie et dictionnaire*; *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

### 2. Lexiculture: a forgotten but essential lexical component

To distinguish, as we have already done, *lexicographie* (theoretical lexicography) from *dictionnaire* (practical dictionary-making), is to ensure we do not confuse a lexicographical phase, devoted to research and focused on words, with a practical phase, devoted to developing the product, and focused on the user – the word having been well determined semantically. We are then located in the sphere of methodology with, as we have noted, many more consequences than would be expected if we did not take account of this useful distinction.

To be interested in *lexiculture* – a term and concept peculiar to Robert Galisson and referring to everyday culture shared by all and stored in words, over and above their semantic definition – is to take account of a particular and fundamental dimension of the vocabulary that, alas, is often lacking in *lexicographie* as well as in *dictionnaire*. Actually, within theoretical lexicography, *lexiculture* is either ignored altogether, or it is neglected, save by a few – too few – language-teaching specialists. And in practical dictionary-making, *lexiculture* is most often absent, for want of being the focus of studies in lexicography, cropping up only when an example contains in addition to the usage of the word some extralinguistic detail. And yet, as we must emphasize straightaway, the *lexiculture* that we shall attempt to define here forms in our opinion an integral part of the complete definition of a word. Indeed, it proves to be indispensable for the foreign learner, or for ensuring that the native speaker gains a perfect understanding of words, over a long period of time.

#### 2.1. A concept launched by Robert Galisson and originating in the learning of French as a foreign language

The specific educational backgrounds of certain leading lights in lexicology can lead to particularly original and rich conceptions for lexicography. Thus, Robert Galisson was first trained as a school-teacher, then quickly acquired a doctorate in lexicology, after a short period spent as a teacher of French literature in the technical sector. As a young academic, he then became interested in the training of foreign students of French. Having become

a renowned lexicologist and a professor at the Sorbonne, he was entrusted by Bernard Quemada with the editorship of the scholarly journal *Études de linguistique appliquée* (Didier érudition; Les Belles Lettres), which has appeared in 120 issues in thirty years, and whose very sub-title is enlightening: “Journal of the didactology of language-cultures”. Having supervised more than 600 theses, Galisson is today the holder of honorary doctorates from numerous foreign universities and enjoys a wide reputation in the worlds of lexicology and didactics.

His influence among researchers in the field known in France as FLE – that is, the teaching and learning of “le français langue étrangère” (French as a Foreign Language) – is considerable in Europe and in the Mediterranean world. His fame is based on the innovative nature of the ideas he has put forward and which I will review quickly here in so far as, in my view, they directly concern our lexicographic activities.

It is important to point out that the particular position adopted by Galisson, who describes himself – among other things – as a didactologist-lexicologist, is rooted in the difficulties encountered by non-native speakers of the French language when attempting to acquire it. Close observation of obstacles to the effective learning of vocabulary has effectively given a new dimension to Galisson’s approach to defining the word. He has ascertained, in fact, that for perfect comprehension of a conversation or a written text, the classical, semantic, definition of the word is inadequate. In order to be perfectly defined, indeed, a word should be presented in all its descriptive dimensions, with all its lexical weight, and so therefore not only with its lexical components but also with the cultural components appropriate to the country that has coloured its usage.

Robert Galisson has developed a number of concepts in didactics, but the founding concept that it is essential to preserve for lexicography is the one he first designated in 1987, in *Études de linguistique appliquée*, under the name of “*lexiculture*”, and then rebaptized in 1996, in the same journal, as “*lexicultural pragmatics*”. In fact, all the researchers have retained the first term, *lexiculture*, which I will try to present here in all its richness.

## 2.2. Beyond the semantic definition: the lexicultural definition

As a first approach, and to put it simply, if we are dealing with lexiculture it is important as regards the information given about a word, not to limit oneself to the meaning conveyed by the traditional type of definition. The sense given in the traditional definition corresponds only to a single aspect of the word, that which relates to its standard usage and its syntactic functions – to what Robert Galisson has called “culture savante” (learned culture).

In other words, one does not have a full knowledge of the word in all its vividness and cultural diffusion, if one contents oneself with the single meaning that the lexicographer has tried to capture within the dictionary definition, whatever formal shape it may take – classical definition, distributional definition, or whatever. Actually, one has not really grasped the popular flavour of a word, that is to say the echoes it evokes and conveys within the linguistic community, if one limits oneself simply to a narrow semantic analysis of its content. Of course, a semantic analysis will provide, for example, a precise list of all the “semes” (the smallest units of meaning) which define that content. But it will lack something at least as important, which is everything that speakers of the same language attach implicitly to this word, within the framework of the “common culture”, even though this does not form part of the semantic definition of the word in the strict sense. It is this implicit knowledge, shared by all adult speakers of a linguistic community, that Galisson calls lexiculture. Lexiculture actually represents for a given word everything contained in its “added value”, an added value that everybody is familiar with and which, however, is almost always absent from dictionaries that are assumed to reflect the complete significance of the word.

Some examples are called for. When words such as *accordéon*, *muguet* (lily of the valley) and *écureuil* (squirrel) are used in France, the image that one has of the word and of its functioning in the language conveys much more than its definition. For example, *accordéon* as it is presented in dictionaries is generally reduced to a quick definition, such as: “a musical instrument with a bellows and metallic reeds” (*Le Petit Robert*, 1967) or, in a fuller definition, “a portable musical instrument, with keys and buttons, whose metal reeds are set in vibration by a bellows” (*Le Petit Larousse*, 2004). These definitions are certainly precise, especially the second one, but

they nevertheless remain disembodied with regard to the connotations that every French speaker has in mind when the instrument is mentioned. In fact, another type of definition exists, one that we shall call lexicultural, implicit and pragmatic, and which is stored in the mind of each French speaker. This kind of definition, which is relevant to the living history of the country and to people’s experience of the accordion, comes as a supplement to that provided in dictionaries, limited by tradition to denotative meaning, which is supposedly objective. Yet no less objective are the implicit references of the word. The connotations are certainly there, where the accordion is concerned, in a more or less identical form, for every French speaker.

What in fact are the lexicultural features of *accordéon*? First of all, the word brings immediately to mind a popular instrument, the “poor man’s piano”, which is never or seldom taught in a conservatory. Every French person knows that traditionally it is not children from well-to-do families who learn to play the accordion. Then, it is an instrument whose sound is perceived as joyful, making one think immediately of the atmosphere of the “guinguettes” (small restaurants with music and dancing), and of what are known as “bals musettes” (popular dancehalls), those Saturday night dancehalls where to an accordion accompaniment one can dance tangos and waltzes, but not as a rule rock’n’roll. It is associated especially with dancing on July 14 (the French national day). Then, thirdly, as soon as *accordéon* is mentioned, a name immediately appears on all French lips: that of Yvette Horner. She is, in fact, the incarnation of the warm-hearted woman of the working classes who has become the queen of this instrument, with numerous well-known refrains and a simple and popular type of humour that make of her a symbolic figure. To such an extent that, in a certain sense, Yvette Horner forms part of the “natural” definition of *accordéon*, the “natural” definition being the one elicited from informants – who were not lexicographers – asked to provide the essential features of the word, linguistic as well as encyclopedic. Finally, within the framework of this same piece of research on the connotations of *accordéon* for a broad range of French speakers, another name was associated with the word for 75% of the persons consulted: that of President Giscard d’Estaing. Why such an association of ideas? It is actually enough to recall that during his first presidential election campaign, Valéry Giscard d’Estaing played the accordion in a village, an event which earned him

Part One of this contribution was published in *Kernerman Dictionary News*, Number 11, July 2003 (<http://kictionaries.com/newsletter/kdn11-3.html>).

This article was translated from French. The original French manuscript is available online: <http://kictionaries.com/newsletter/kdn11-03fr.html>.

I am most obliged to Anthony Cowie for his dedicated and competent assistance with the English language version. Inaccuracies that might concern any aspect of the translation are my own. IJK

## Lexiculture and the EFL Dictionary

Anthony P. Cowie

As Jean Pruvost has argued convincingly in this stimulating account of Robert Galisson's pioneering work on 'lexiculture', cultural aspects of meaning are a neglected element in standard dictionaries, and a much-needed one in dictionaries intended for foreign learners of a language. Less progress has admittedly been made among English-speaking than French-speaking scholars in elaborating a theory of lexiculture – an exception being Gabriele Stein's invaluable article on 'EFL dictionaries: meaning, culture and illustrations' in *Better Words* (Exeter, 2002). Yet, some noteworthy advances have been made since the early 1990s in this area, and specifically in the development of the so-called 'EFL cultural dictionary'. We now have the *Oxford Advanced Learner's Dictionary of Current English, Encyclopedic Edition* (1992) and the *Longman Dictionary of English Language and Culture* (1992, 2e 1998), each based on the immediately preceding edition of the standard EFL work.

In both dictionaries, there are notes on various aspects of English culture. For example, in the *Oxford Encyclopedic*, there are ninety-four special articles dealing with 'class', 'crime', 'food', 'gardens', 'the royal family', and so on. These topics are generally treated at some length, and with

countless amused comments in the press and many cartoons. His aristocratic style contrasted effectively with the "poor man's piano".

To sum up, one can provide a lexicultural definition of *accordéon* by recalling the essential lexicultural features of the word, defining in this way what Galisson called its "shared cultural content". Thus, the accordion is for all French people synonymous with a popular instrument, and dancing to accordion music. It is also associated with the image of Yvette Horner, and for 75% of the informants with Giscard d'Estaing. Yet, when examined more closely, these essential features of the lexicultural definition of the word *accordéon* are almost always absent from our dictionaries. The definition of the word therefore lacks all that gives it its deepest resonance.

From time to time, though rarely, one of these lexicultural features is, of course, recaptured in the example that follows the definition. But this is very far from being indicated systematically and Yvette Horner or Giscard d'Estaing are of course never cited, although their names are immediately associated with the accordion in the minds of French people. It goes without saying that limiting an entry to a semantic definition of the word is incomplete. Not to indicate in fact the lexicultural aspects can leave a terrible gap, especially for the non-native speaker, who needs to decipher the allusions, the implicit references of a word encountered in conversation, in a newspaper, in a novel, etc.

Not all the words used in a linguistic community for which the lexicographer is the legitimate analyst necessarily carry a common cultural load; and yet, when we look at them more closely, adding a lexicultural definition is essential for many. To give just a few more examples, a word such as *muguet* (lily of the valley) is in France bound to be associated with May 1, Labour Day, for which this flower actually represents the symbol. It is sold on this day and no other: to buy a lily of the valley on May 15 or April 15 makes no sense for a French person. In any case, it would not be on sale at the florist's... Also, to define the lily of the valley as "lilac with small white flowers giving off a sweet and pleasant smell" is indeed very interesting – here we are in the world of "learned culture" – but not to add in an example or in an encyclopedic expansion that we are concerned with a symbolic flower sold on May 1 in the streets, in all the shops, in the metro, etc, is to overlook the heart of the matter. The lexicultural component must

be mentioned here to avoid presenting the reader of the dictionary with a definition of the word that is very far from complete.

In the same way and to give a final example, the word *écureuil* (squirrel) is defined in France as in other countries as an "arboreal (= tree-dwelling) rodent with fur ... and a bushy tail, feeding mostly on seeds and fruit". But, it is quite right that most French lexicographers specify, in the manner of the *Petit Larousse illustré*, that its fur is "generally reddish (in France)", 95% of the French population ignoring in fact that a squirrel can have grey fur. But then come elements that are not found in our dictionaries but that also form part of the lexicultural component of the word for a French person. First, without being unduly anthropomorphic, it is important to say that the squirrel is the object of much affection among French people. We are always happy to catch sight of one in the garden; it is a symbol of liveliness and grace. However, for my friends in Québec, and in more and more countries, the image is reversed: it appears dangerous, and comparable almost to the rat, which causes so much damage in people's attics. This lexicultural feature should thus be specified. Finally, and above all, for the French, the squirrel represents the symbol of savings because it has been chosen as the extremely popular emblem of "la Caisse d'épargne" (the Savings Bank). There is hardly a young French person who has not received the gift of a savings account booklet bearing this image.

So, a politician taking part in an election campaign who was to declare today: "I am not the type to play the accordion; I would rather offer you a sprig of lily of the valley and talk to you about real work, and awaken in you the dormant squirrel and its piggy bank", will be understood by all Frenchmen. However, no dictionary would enable a foreigner to understand that message. And if in the twenty-second century the Savings Bank no longer exists, May 1 is no longer celebrated, and Giscard d'Estaing's accordion is forgotten, there will no longer be anyone able to translate this message, and no dictionary will be able to help.

### 2.3. Some lexicographic and dictionary perspectives

#### 2.3.1. The lexicultural anchorage points

Essentially, in the fields that interest us – *lexicographie* (theoretical lexicography) and *dictionnaire* (practical dictionary-making) – it is words listed as dictionary entries that are our primary concern in lexiculture. If we are aiming not to obscure the lexicultural dimension in dictionaries,

it is in fact those words that are to be treated first. Other lexical elements, often positioned in the body of entries, are nevertheless to be taken account of, as they too show themselves to be privileged bearers of the “shared cultural content”, the lexiculture.

Over and above the words which comprise the entry-list, one will note first of all those longer expressions that have been delexicalized, unfrozen and reshaped, called by Robert Galisson “verbocultural palimpsests”. They include, for example, the titles of films and novels, and famous pieces of poetry, which everyone in a given linguistic community knows and which, by changing a word, can be re-utilized to create an amusing or eloquent effect. Such is the case with “My kingdom for a horse!”, the famous exclamation of Richard III, which could be ironically transformed into “My kingdom for a good book!” Now here, few dictionaries give guidance: the expressions that serve as moulds are not really listed. In the domain of lexicography, there is a lack of research based on large corpora that would enable us to determine the frequencies of use – to identify for instance what are, over a decade, the lexicalized expressions that are most often taken up and reshaped to create a new effect.

“Verbocultural palimpsests” clearly belong to lexiculture and one can appreciate how difficult it is for the lexicographer cum dictionary-maker to find precise criteria for recording them. A paralyzing concern with clearly defined objectivity and with the permanence of what is recorded leads one to be very cautious in this area, which is nevertheless perfectly linguistic. The French in fact are constantly resorting to these devices: everybody is aware that a famous song or the title of a film that has been very successful in France can be memorized by an entire linguistic community in the space of a few decades and, by being “unfrozen”, serve as a model for other formulas. A French singer, Alain Souchon, has for example launched the expression “Allô, Maman, bobo” (Hey, Mum, it hurts), “bobo” being baby-talk for “it hurts”. Such a well-known expression has often served as a matrix for numerous captions in newspapers, articles, and so on. “Allô, Maman, canicule...” (Hey, Mum, it’s a real scorcher...) can appear in the press whenever the weather is scorching hot. In the same way, “The fabulous destiny of Amélie Poulain”, the title of a highly successful film, serves as a mould for numerous other expressions. Readers will have noticed that since September 2002 there have been dozens of titles

promoting this or that character, or this or that product, after the pattern of “the fabulous destiny of...x, y or z”.

The phenomenon is not new, all linguists have noted its development, and in nearly all languages this process of linguistic creativity is actually very active. One must admit that there is really no dictionary reflecting all this. However, for some of these expressions, it would be good if they were to appear in a “lexical” dictionary, since their lexicultural nature is shared by an entire linguistic community. Thus, the French expression “Métro, boulot, dodo” (Metro, work, sleep), illustrating one of the tiring and restrictive aspects of Parisian life for people who daily travel to work there, has undeniably served as a mould for over twenty years for numerous other expressions, e.g. “Métro, boulot, promo” as a headline in *Le Point* of 8 August 2003, p.15. Although it is generally not listed in dictionaries, the expression “Métro, boulot, dodo”, because of its frequency of repetition, surely deserves to appear there, as it has, so to speak, entered the language.

Also eminently lexicultural are proverbs, which, in different languages, do not always have equivalents, or convey different images. It is known for example that the English expression “if pigs had wings (they might fly)” corresponds in French to another amusing image “when hens have teeth”, and that here there are a number of images clearly susceptible to various reshapings: “when chickens have teeth”, “when hens have no cockerel”, etc. But it should be acknowledged here that, as a rule, dictionaries devote a good deal of space to proverbs. We notice for instance that, in the *Petit Larousse illustré*, they enjoy a special place in the pink pages that separate the part devoted to the language from that devoted to proper names. Very sensibly, too, since the beginning of the 21st century, *Larousse* have also added in the same place historical phrases such as “Rally around my white plumes”, or “Paris is well worth a mass”, uttered by Henry IV, or again “After us the flood” attributed to Louis XV, all historical phrases well known as means of saying, respectively, “follow me, in honour”, “one should know how to make concessions”, or also “let’s think about ourselves first”. This is taking effectively into account a part of the lexiculture.

Another domain is represented by the brand names that are increasingly in evidence in all the languages of countries where consumption is high. A certain number of brand names can become common nouns that dictionaries cannot

▶ a wealth of lexicultural detail, such as for example guidance on how to order a pint of bitter beer! Here and there, too, one finds very precise information about the use of various routine formulae. For example, under ‘conventions’, the reader is told when he or she should use ‘please’, ‘excuse me’, ‘how do you do?’ and ‘that’s all right’.

Other places in the *Oxford Encyclopedic* in which cultural detail appears are the ‘mini-notes’: “short extra paragraphs giving information on the special connotations these words have for native speakers of English.” Consider some of the detail for ‘tea’ – suggesting parallels with the small details of everyday life which clearly fascinate Galisson: ‘Tea also suggests comfort and warmth, and sitting down with “a nice cup of tea” is a common response to problems and worries.’

Corresponding, in the Longman work, to Oxford’s mini-notes are a large number of so-called ‘cultural notes’. These deal with a wide range of topics, including religion, popular superstitions and social stereotypes, and are well set out for quick reference and learning purposes.

A noteworthy feature of the Longman dictionary is the space given over to cultural illustrations. Several pictures (e.g. the one for ‘yuppie, or Young Upwardly-Mobile Professional’) reflect in an entertaining way the connotative details appearing in the definition, which include: ‘In Britain, yuppies are seen as young people who earn a lot of

▶ money without necessarily working very hard, usu. on the financial markets in the city.'

Less than a hundred mini-notes, and about the same number of special articles – and this is just to speak of the *Oxford Encyclopedic* – do not amount to a great deal in a dictionary of 93,000 entries. However, the two dictionaries represent a notable step forward, both in identifying words and phrases of cultural interest and in devising effective methods of presenting them to the advanced learner. None the less, English-language dictionaries still have much to learn from the type of systematic exploration of culturally-rich items to be found in the research of Robert Galisson.

● Anthony P. Cowie has vast experience and expertise in teaching English as a foreign language, particularly in Nigeria, as a specialist with the British Council, and as a Lecturer at Leeds University. He has been involved in six ELT dictionary projects, including editorship of the *Oxford Advanced Learner's Dictionary* (4e) and joint compilation of the *Oxford Dictionary of Current Idiomatic English* in two volumes, and was Director of the OUP Lexical Research Unit at Leeds. He was later appointed Reader in Lexicography at Leeds University, and now serves on the Editorial Board of the *New OED*. He was until recently Editor of the *International Journal of Lexicography* and is author of *English Dictionaries for Foreign Learners – a History* (Oxford 1999).  
a.p.cowie@btinternet.com

avoid treating, such as, in French, for example, *frigidaire* for a refrigerator, *mobylette* for a moped, *bottin* for a telephone directory, etc. Now, a good number of brands become associated with slogans that everyone knows, and they in a certain way pass into the language, being picked up with a wink by speakers who are used to hearing them. "Because I'm worth it", associated with a pretty actress and a brand of hair lotion is recognized by every French person, as is the expression, "It moves...", indicating that something is very strong, by reference to a brand of mustard for which it is the slogan. The addition of the word "grandmother" to "make good coffee", because of a pleasant advertisement that links the product to the reassuring grandma image, is also currently familiar in France. Now, these are facts about the language, with a lifespan exceeding a decade in some cases, which no dictionary takes note of – except for the one compiled by Robert Galisson, which is unfortunately difficult to obtain, the *Dictionnaire des noms de marque* (Dictionary of brand names), published by the CNRS (National Centre for Scientific Research, 1998). It goes without saying that we need to consider seriously whether certain of these items should be included in the general-purpose dictionary.

Indeed, everything that at the level of discourse arises from the common culture, and is integrated into it by the entire linguistic community – which does not hesitate to use it, whether by adapting it or employing it as is – deserves, in one way or another, to be included in the dictionary.

### 2.3.2. Dictionaries with a lexicultural dimension?

#### *Reflexes for the lexicographer to develop: investigation and oral corpus*

Here one enters the experimental domain, and it may well be the case that the first step should be to transform in part the attitudes and practices of the lexicographer. Actually, the lexicographer can be characterized in general by the linguistic and philological competence he or she has acquired in training and by the working experience accumulated year by year. He or she puts this knowledge and this experience at the service of the community in order to compile entries based upon a close observation of the language. To do this, he or she has recourse to a corpus which, most often, is written and consists of texts drawn from works of literature, from the general and specialized press, and most recently, from the Internet. This corpus serves above all to provide the lexical documentation which enables

one to pin-point good examples as well as possible new meanings and neologisms.

But if as lexicographers we wish to introduce a lexicultural dimension into our entries, we need to "listen" more than we do today to the radio, "watch and listen to" the television, by all means follow cultural developments, the learned culture, but also and especially popular, everyday culture. Thus, songs, films and advertisements should form an integral part of the corpora. To take just one domain that is eminently lexicultural, that of the popular song. In France we need to take account in our dictionaries of phrases that have become well-established in the collective memory for many decades: "Auprès de mon arbre (Near my tree)", "Une jolie fleur dans une peau de vache" (A pretty flower in the hide of a cow) for Brassens, "C'est un jardin extraordinaire" (It's an extraordinary garden) for Trenet, "Les portes du pénitencier" (The prison gates), "Qu'est-ce qu'elle a, ma gueule?" (What's wrong with my face?), "Allumez le feu" (Light the fire) for Hallyday, "Laisse béton (tomber)" (Drop it), "Mon beauf (beau-frère)" (My brother-in-law), "C'est la mer qui fait l'homme" (It's the sea that makes the man) for Renaud, etc.

It is important then to note down as one goes along, with a watchful eye, everything that happens by way of lexiculture establishing itself in the minds of a linguistic community. The impact of current affairs, of cultural life, of advertising, should then be assessed in terms of the deep impression it makes on each person; statistical investigations will be needed to evaluate this impact. And, just as neologisms of form and meaning are always difficult to record with certainty as to their lifespan in the language, so lexicultural features, once they are identified, should be followed attentively for as long as they survive. Some will disappear quite quickly, but others will gain cultural permanence: the lexicographer needs to be an attentive and eclectic observer.

The truth is that practically no lexicultural features are introduced today into our dictionaries; they are present only in a random, patchy and subjective manner. Precise investigations, with constant reference to the oral corpus and daily attentiveness to the common culture, such are the new attitudes that should be added to those of the observer of the language in action. Let us admit it: here is a new task that demands much effort and that, if it is to take concrete form in dictionaries, requires also new methods.

### ***The dictionary example: usage but also lexiculture***

Whether we are concerned with citations or invented examples, it is obvious that dictionary examples are most often chosen or made up to throw light on the usage of the word that they contextualize. Syntactic function generally prevails over the encyclopedic nature of the information. However, when the example has an encyclopedic dimension, the reference is almost always to the learned culture. For the tiger, one will always be informed about “the growling tiger” or “the tigress with her young”, but one will never find “to put a tiger in your tank”, which matches an advertising slogan which was so successful that it became established in people’s minds without them even remembering precisely what brand of engine oil it was designed to promote.

It can indeed appear difficult to integrate advertising material into a dictionary. Legal problems of course arise. It is nonetheless true that if all French people have this connotation in mind, it should appear in one form or another, and examples represent without doubt one possible means of entry to the dictionary for lexiculture. Examples drawn from the news media, from reviews, or from advertising are therefore not to be ruled out.

For the entry devoted to *cicada*, to give as an example – as is the case in the *Petit Robert* (1st edition, 1967) and the *Grand Robert* (1st edition, 1964) – “The cicada sucks the sap of plants: the male emits a piercing sound”, is certainly interesting on the encyclopedic level, but it would be appropriate to add another, lexicultural, example, which relates the cicada to the ant, in reference to the fable known by all French people, where the cicada is made to appear carefree and lacking in forethought whereas the ant is inclined to be thrifty. Everyone has in fact memorized the lines of La Fontaine: “The cicada, having sung all summer, found itself at a loss, when the north wind began to blow”. Not to include it, is to treat lexiculture with disdain and not properly meet the needs of the non-native speaker. Paul Robert certainly wanted to carry the work of Littré further, yet distinguish himself from it with citations drawn from the 19th and the 20th centuries, but sometimes, the lexiculture dates back to the 17th century!

### ***Lexicultural development: the insert, the hypertext, etc.***

It is not always possible to add a lexicultural example. For the word *apostrophe*, one will find for example

“the apostrophe of Ciceron to Catilina” that skilfully combines the use of the word with a scholarly reference, yet missing from the entry is a reference to a television programme that all French people know – “Apostrophe”. This weekly programme, which in fact hosted the writers of the most recently published books and symbolized discussion of books, left an impression in everyone’s memory. It forms part of everyday French culture. In reality, in order to enlighten a foreign reader or a reader of the 22nd century, a lexicultural type of comment should be added to the definition of the word *apostrophe*, to point out that a weekly televised discussion programme was so named, with more than a thousand broadcasts spread over twenty years. In this way it is possible to understand a comment made the other day on a café terrace, when, faced by an impassioned discussion launched by three customers about a novel, the waiter said to them: “You are playing Apostrophe!”

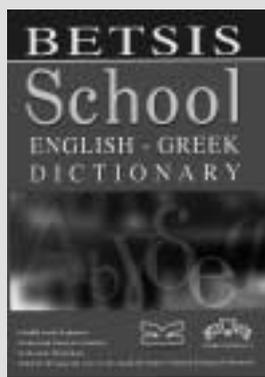
A dictionary format that is suitable for promoting lexiculture is in fact already in existence. In the language part of the *Petit Larousse illustré*, after the different meanings of the word have been listed, we are actually often offered a small encyclopedic expansion of the subject treated, an expansion that clarifies the word and the concept it represents. For example, for the word *engrais* (fertilizer), an encyclopedic comment follows about the nature of fertilizers, but also about their function. One appreciates how here, at the end of the expansion, is mentioned the “environmental damage” caused by fertilizers, “especially through the pollution of underground streams”. In fact, one is already in the domain of lexiculture, to the extent that the notion of fertilizers, highly favourable in the years 1960-1970, has little by little been devalued by negative connotations regarding pollution. The word *fertilizer* no longer carries the same “cultural charge” that it did in 1960.

So, then, the route is fully mapped out: if in these encyclopedic expansions more space is allotted to lexiculture, one can benefit from a complete explanation of the word, with all its resonances. Let us suggest then that we add a lexicultural expansion systematically to all the words that need it: *accordéon*, *pétanque* (a game of boules associated first with the south of France), *dauphin* (an untouchable fish, symbol of animal intelligence), *renard* (cunning, as in the wolf and the fox, etc), not forgetting words brought to life by advertising, by songs, by radio or television programmes: the *cow* (“that laughs”, one of our most known cream cheeses for



PASSWORD English-Hungarian Learner’s Dictionary Angol-Magyar Tanulói Szótár, published by NEMZETI TANKÖNYVKIADÓ, was awarded the Special Prize of the Jury of HUNDIDAC – the Hungarian Association of Manufacturers, Distributors and Consumers of Educational Media – Budapest, November 2003.

From the series  
KERNERMAN SEMI-BILINGUAL DICTIONARIES



### PASSPORT

Betsis School English-Greek Dictionary

ANDREW BETSIS ELT

Pireas, Greece

August 2003

Editing: Andrew Betsis,

Lawrence Mamas, Linda

Collins

Translation: Manos Betsis,

Aliki Lambea, Dimitra

Zaharopoulou, George

Dimitriades, Sophia

Papadimitriou

ISBN 960-7414-06-3

Paperback, 607 pages

206 x 139 x 31 mm.

CD-ROM: Windows XP/

2000 or higher

[http://kdictionaries.com/](http://kdictionaries.com/products/junior/junior-greek.html)

[products/junior/junior-](http://kdictionaries.com/products/junior/junior-greek.html)

[greek.html](http://kdictionaries.com/products/junior/junior-greek.html)

From the series

KERNERMAN SEMI-

BILINGUAL DICTIONARIES

over half a century), the *écureuil* (and the Savings Bank), *bobo* (Hey, Mum, it hurts), the *loft* (associated from now on with the televised programme devoted to those youngsters who are filmed continuously and eliminated day by day by viewers), the *weather* (hardly separable from the televised report corresponding to it: “watch the weather”), etc.

Clearly, it is just as easy to add a lexicultural expansion to a paper dictionary as to one in electronic form, where hypertext can be very welcome. With regard to paper dictionaries, the margins on the right and left can also accommodate this type of lexicultural commentary, as has been done in part, quite effectively, in the *Larousse Super Major* targeted at pupils between the ages of 9 and 12.

Lexiculture is introduced through examples and at the heart of a specific expansion, and it will certainly be appreciated by native speakers, who like to verify what they have rather confusingly memorized, and by non-native speakers, who need this information to penetrate the kind of complicity that surrounds the use of a word by speakers of the same foreign language.

#### ***Lexicographic and dictionary-making sieves, annexes: the antechamber of security***

It has been pointed out how, in the latest versions of the *Petit Larousse illustré*, the pink pages have been made to include “Historic words”, a device which is both extremely useful and also meaningful regarding general, everyday culture and a better understanding of the language, because, in order to express his thoughts, a French dictionary user will resort to this section. With these appendices, known under the name of “pink pages” and devoted first to Latin phrases, then to proverbs, a dictionary like the *Petit Larousse* benefits from a subtle tool that, in practical dictionary terms, is very efficient. We might add that, in the 2004 edition, the fact that fifteen pages of “Mots nouveaux” (New words) have been built in, combined with cartoons, and “an artist’s views”, really opens up the way to lexiculture. Very flexibly, it introduces, for example, a new meaning in French of the word *collègue*, designating a friend, a mate, engaged in the same enterprise, or another new word, *pêchu*, “someone who is in form”, in very good health, according to the familiar French expression, *il a la pêche* (he’s feeling great).

In the same spirit, we need to draw attention also to the blue pages in the *Dictionnaire Hachette* that, at the end of the book, are devoted to “New words of

living French”. To protect themselves from the very brief life-span of a certain number of new words, the lexicographers have listed here all the recent words that, if their usage is confirmed, will be integrated into the dictionary columns of the next edition. This practice, which is nearly ten years old, seems interesting because it allows more flexibility to the lexicographers, who no longer find themselves faced with the dilemma of whether to include or exclude. The lexicographers have with this appendix a compromise voice, a sort of sieve, an antechamber of security that makes less onerous their task of being well-informed observers, watchdogs posted at the gate of the dictionary to welcome or turn away the new arrivals.

This flexible practice of appendices outside the dictionary text, and this use of the sieve, of the antechamber, for certain concepts whose durability is not yet assured, seem to us very much to the point. As regards lexiculture, it is necessary in fact to distinguish between what is already very well established (the accordion and the “popular dancehall”) and what is more recent (the title of this or that novel or film, or indeed a recent expression from a politician, for example, the one coined by Jean-Pierre Rafarin when speaking of “France from above and France from below”), without forgetting the brand names and a certain number of advertising slogans, forming part of the lexiculture of the year, of which certain elements will enter the language and others will disappear. So by giving flexibility to the lexicographer, room for manoeuvre is created.

It should be added that it is in principle up to the lexicographer to take account objectively of everything that has an impact on the language of his contemporaries. Yet, even if he subscribes to the (disputable) tradition of giving priority to the written over the spoken language, lexicultural features operate strongly in everyone’s speech and the silent recognition of a shared language among persons of the same tongue and the same common culture is found constantly in literary texts as in the press. As has already been stated, a certain number of these features remain active for several decades, not to say permanently, others last only a few years. And yet, if the lexicographer must imperatively note down everything that is lasting, why should he not also feel compelled to offer to everyone this lexical memory that would ensure that nothing in the language could be lost?

It is in dictionaries that we should record such expressions as “France from above”

and “France from below” if we are to expect that in half a century from now all the numerous articles could be understood that took this expression more than a year ago, and adapted it to form: “the music from above, the music from below”, “the pollution from above, the pollution from below”, and so on.

I feel inclined to conclude this plea for the introduction of lexiculture into dictionaries by using this formula: let us avoid imagining, in fact thinking, that there will be “dictionaries from above”, for learned culture, and “dictionaries from below”, for everyday culture.

Our dictionaries are in fact designed for everyone and there can be no doubt that the immense talent of lexicographers as expressed in dictionaries can find in lexiculture both a linguistic theme and a new inspiration. In any case, we should not be worried: theoretical lexicography and practical dictionary-making belong to a useful, generous and triumphant genre. These pages offer an excellent forum for making that welcome expansion still more widely appreciated.

### 3. The triple dictionary investigation: the dictionary as a corpus

As regards theoretical *lexicographie* and practical *dictionnarique* (dictionary-making), it is Bernard Quemada to whom we must attribute the distinction. Recognizing him as at one and the same time a lexicologist, lexicographer and metalexicographer of the highest reputation in France, French researchers realize how indebted they are to him for the revival of our disciplines and for the extraordinary links that he has always been able to forge between, on one hand, tradition – history – and, on the other hand, modernism – the future. I have the good fortune of having studied under his guidance and of working at his side today.

As regards *lexiculture*, it is Robert Galisson whom the invention and elaboration of the concept should be attributed to, as well as its dissemination. It is his reputation and a shared passion for teaching vocabulary that first brought us together. I have also had the great privilege of working at his side within the framework of the journal *Études de linguistique appliquée*.

As for the *triple dictionary investigation*, it is an approach I implemented during research I was asked to carry out on the topic of “the norm”, an approach tested on one word and which seemed to me capable of benefiting from being more widely known about and tested on two counts.

#### 3.1. Two beneficiaries: the reader and the lexicographer

It seems in fact that, on the one hand, anyone seeking to exploit a dictionary as richly as possible for a word, for a given concept, may be surprised by the richness of the results obtained by the triple dictionary investigation, while on the other hand, for a lexicographer, this triple investigation seems capable of bringing about a marked improvement in the coherence of our dictionaries.

In fact, the initial research I was asked to do into the concept and definition of the word *norme* (= norm, standard) has proved in itself to be very rich in thought, because the norm represents, on one hand, what is unconsciously fitted in by everyone and, on the other hand, what is consciously circulated in the specific works that collect and diffuse it, notably dictionaries. The dictionary represents in fact both the place of memory, and thus the description of a language at a given moment, and the place of arbitration where readers come to be reassured as to the precise meaning of a word – that is the norm.

How do dictionaries define the norm, a concept that, when all is said and done, runs through the whole dictionary as soon as we are concerned with defining a word? And how is the word *norme* used in the dictionary, beyond the actual entry that is devoted to the word? It was while I was researching exhaustively all information about the word *norme* provided by a corpus of dictionaries, that the idea emerged of the triple dictionary investigation. What this really means is X-raying our dictionaries in such manner that they provide more information than the lexicographer believes he has obtained from them.

First of all, the prime objective of the triple dictionary investigation is to implement a method that allows us to bring to light from dictionaries information about the looked-up word that goes farther than simply the entry devoted to the word whose meanings and usages one is searching for. In this way the researcher who wishes to determine as exhaustively as possible a word or a notion, can make the most of the dictionary.

Then, the second perspective that is held out is perhaps that of offering a means of verifying the quality of the entry with regard to a consistency to be established in the dictionary, and thus possibly of improving the quality of the said entry and of all those related to it. No doubt lexicographers will be surprised to discover, besides, through the triple investigation the unconscious riches that



#### PASSPORT

Angleško-Angleško-  
Slovenski Splošni Šolski  
Multimedijški Slovar  
ZALOŽBA ROKUS  
Ljubljana, Slovenia  
September 2003

Editor: Boštjan Zupancic  
ISBN 961-209-347-4  
Hardcover, 799 pages  
167 x 120 x 37 mm.  
CD-ROM: Windows XP/  
2000 or higher  
[http://kictionaries.com/  
products/junior/junior-  
sloven.html](http://kictionaries.com/products/junior/junior-sloven.html)

From the series  
KERNERMAN SEMI-  
BILINGUAL DICTIONARIES



### PASSPORT

Dictionar Englez-Român  
EDITURA NICULESCU

Bucarest, Romania

October 2003

Editors: Anca Florea,  
Dragos Trifan

ISBN 973-568-505-1

Paperback, 632 pages

229 x 162 x 25 mm.

[http://kdictionary.com/  
products/junior/junior-  
roman.html](http://kdictionary.com/products/junior/junior-roman.html)

From the series  
KERNERMAN SEMI-  
BILINGUAL DICTIONARIES

they bring to the process and which merit full exploitation.

Whereas the first two approaches, *lexicographie* and *dictionnarique* on the one hand, *lexiculture* on the other, seem able to benefit from immediate applications, the third approach, the *triple dictionaric investigation*, depends rather more on experimentation being set up.

### 3.2. The first dictionaric investigation

The first approach inherent in this triple dictionaric investigation can appear childish as it is part of current practice among lexicographers, yet it is necessary to describe it for dictionary users and thus show how much the consultation of a single dictionary is restrictive. The first dictionaric investigation therefore naturally consists of reading and analyzing in several dictionaries the entries corresponding to the word whose different senses are being looked up. To be fully effective, this reading should bring together two dictionaries of a similar size.

Simple comparison in this respect of two reputable reference dictionaries for readers of the French language, in this case, the *Petit Larousse* (PL) and the *Petit Robert* (PR), helps us to understand that, for example, the entry *norme* represents already at this stage the obvious result of an interpretation of language and speech. An explicit interpretation of the word, differing from one dictionary to the other, is displayed through the structure of the entry, with its different senses and sub-senses, with its defining developments and exemplification. Here emerges in fact a first interpretation, all in all the visible side of the dictionary.

Thus a comparative analysis of the word *norme* in PR and PL, in the 1994 editions that we selected, turns out to be particularly enlightening. Not being the object of a homonymic grouping in either dictionary, the entry *norme* looks like a very interesting case of “polysemy”, if one takes into account the high number of its meanings, six in PR, four in PL. On closer inspection, the light thrown on the word and its interpretation, which we refuse to perceive as an entity cut up into four or six senses, is quite different from one dictionary to the other.

In PR, a point is made of a meaning that is peculiar to linguistics (“Ling.”). This is described in fourth position and represents “that which, in speech, in discourse, corresponds to general usage (opposed in the one case to system, in the other to discourse)”. Apart from the fact that neither in the entry for *system*, nor in the entry for *discourse*, do we find the notion

of *norme*, the brackets thus remaining mysterious for the non-linguist (and even for the linguist), the editorial team thus thought fit to determine a specific sense in the domain of linguistics, which is not the case in PL. In the latter, it is for the sense given at the head of the entry, the general meaning, to include the one pertaining to linguistics: “usual condition, conforming to the established rule”, a definition that is quite close to that given in second place by PR: “usual condition, conforming to the majority of cases”. In PR in fact comes first the sense marked (“Liter.”): “concrete type or abstract expression of what should be”, which does not appear, actually, as an archismeme.

As for PL, it begins by giving three meanings pertaining to the general language, the first sense already mentioned and, in second place, that corresponding to “criterion, principle to which are referred all judgements of a moral or aesthetic nature”. The third sense, although preceded by the sign “Techn.”, remains in fact multifunctional and is presented in general terms: “Rule setting the conditions for the fulfilment of a process, of the execution of an object or the development of a product in order to unify its use or assure its interchangeability”. The examples (“ISO norm”, “norm of productivity”) are present in order to locate this sense in the technical world. The last to appear, with a non-restrictive development, is *norme* in the algebraic sense of the term, the “norm in a vector space”. Contrary to the dichotomy – very French and in part arbitrary – that we recognize between a language dictionary and an encyclopedic dictionary, the definitions given by PL are in reality very close to those that could be given by a language dictionary, such as the *Dictionnaire de l'Académie*.

As for PR, in fact, a single sense, the second, is devoid of any label. The polysemy is deliberately treated as an integral part of the special-purpose vocabulary with its identifying labels: literary (“Liter.”), technological (“Technol.”), linguistic (“Ling.”), legal (“Leg.”), mathematical (“Math.”). The general look corresponding to that of PL gives way here to a look that subdivides the domains of usage, in an almost distributionalist manner with, in addition, examples of usage peculiar to a dictionary that situates description mainly on the level of the language as system, although the fourth and fifth senses are desperately short of examples.

One can go farther in the comparison and note from the different organization of its senses that PL is positioned more

on the side of the general language, with definitions that allow us to understand better the specialized senses, which are placed second, whereas PR does not give a general definition from the outset and takes account of the variable definitions from one special-purpose language to another, in an order that is unclear to the reader. Here, then, are two ways to pinpoint the norm regarding precisely the word *norme*.

This first investigation has definitely enabled us to ascertain that the polysemous treatment provided by the two dictionaries can, for example, swing between two perspectives, the one installing the description of the word first in the general language and the other mixing special language and general language in an order whose hierarchy does not seem important. This first type of investigation, based on the comparison of dictionaries, naturally gains by being extended to the largest number of dictionaries possible: beyond the different perspectives revealed by reading each one, we can equally bring out, of course, the common features, in a sense, common denominators. Likewise the information can also be combined.

### 3.3. The second dictionary investigation

A second approach takes the form of locating all the entries in the dictionary that contain the word *norme*. These entries thus make it necessary for the lexicographer to use the word whose definition one has just looked up in the relevant entry. Are any of us very clear here? Admittedly, this tracking is by no means easy unless we benefit from access to an electronic version, which is the case with PL and PR.

One can then launch an investigation of the entire dictionary as a corpus, noting down all the occurrences of the word, in this case the word *norme*, and thus identifying all the entries involved in its use for the theoretical or practical lexicographer. Over and above words of the same family, we find appearing in this way a network of entries that have in common the use of the word in the development of their meanings or in their exemplification.

One finds oneself, then, in a context which provides an implicit interpretation of one norm though it is actually called on to clarify another. One thus discovers a second network that forms part of the indirectly visible face of the looked-up word – a face which is however well represented in the body of the dictionary. The computerization of the dictionary that makes possible this complete and rapid

radiography opens the door to numerous entries that would have never been consulted, with the exception of a very small number that are the subject of an analogic cross-reference in the definition of the word. A comparison of PL and PR is equally revealing of a different hidden side. In this way, the characteristics of this or that work can be thrown into relief, quite apart from the illumination provided by the accumulation of two updated networks.

This is how one can bring to light a certain dictionary tomography (we prefer this form to the better known “scanning”) of a word. In X-raying the whole of PL and PR in order to shed light on entries that make use for example of the word *norme*, we cause an image to appear that allows us also to assess indirectly what the word represents and ipso facto the norm in general, once it is no longer under direct scrutiny within the entry that corresponds to it. Thus one finds in PR 77 occurrences of the word *norme(s)* divided almost equally between the singular (40) and the plural (37), against 31 in PL.

By tracking the use of the word *norme(s)* in the discourse of PR or PL, we are naturally also throwing much more light on the notion of norm as it exists confusedly in the minds of the linguistic community, such are the differences of approach between one dictionary and another. To throw into relief our interest in this second type of investigation, let us give here in broad outline the results obtained for each of the two dictionaries.

In PR, one notes for example six major themes with, on the one hand, four themes associated with a particular specialist domain (linguistics and writing; mathematics, science and economics; law and politics; special professional fields); and, on the other hand, two themes that correspond to values that will be qualified as divisible by non-specialists. These last two themes are defined, first of all, by an opposition between a norm to which one assimilates and one to which one is opposed and, then, by philosophy and sociology.

What then is the network of entries that accommodates in PR *norme* in the singular and in the plural? Roughly, in specialized domains, the singular is dominant: it is suitable for “linguistics” with *accent*, *difference*, *language*, and for “mathematics” with the entries devoted to *space*, *intensity*, *module*, etc. When, however, notions of measure or of the economy are involved, all the uses are plural: the antisymic, parasymic, dimensional norms of an object; the



#### PASSPORT

Anglisko-Slovenský Slovník

Junior

SLOVENSKÉ

PEDAGOGICKÉ

NAKLADATELSTVO

Bratislava, Slovakia

November 2003

Translation: Drahomíra

Ondrová, Mária Badinská,

Anna Šebestová, Alexandra

Jarošová

ISBN 80-10-00009-4

Softcover, 726 pages

208 x 148 x 32 mm.

CD-ROM: Windows XP/

2000 or higher

[http://kdictionaries.com/](http://kdictionaries.com/products/junior/junior-slovak.html)

[products/junior/junior-](http://kdictionaries.com/products/junior/junior-slovak.html)

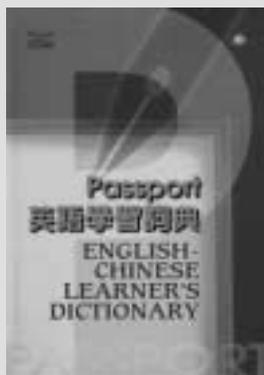
[slovak.html](http://kdictionaries.com/products/junior/junior-slovak.html)

●

Awarded 2nd place in the Dictionary of the Year 2004 competition of the Czech and Slovak Union of Interpreters and Translators (JTP), and the Special Jury award (3rd place) for the best vocabulary edited in Slovakia.

<http://www.jtpunion.org/SlovníkRoku/2004.html>

From the series  
KERNERMAN SEMI-BILINGUAL DICTIONARIES



## PASSPORT

English-Chinese Learner's  
Dictionary

SESAME PUBLICATION

Hong Kong, 2003

Softcover, 692 pages

216 x 155 x 28 mm.

ISBN 962-8811-70-3

[http://kdictionaries.com/](http://kdictionaries.com/products/junior/junior-chinesehk.html)

[products/junior/junior-chinesehk.html](http://kdictionaries.com/products/junior/junior-chinesehk.html)

From the series  
KERNERMAN SEMI-  
BILINGUAL DICTIONARIES

“norms of dimension” in the entry *cut*; the “norms of the group standard” in the entry *standard*, etc. But in the sector pertaining to the human sciences, between *philosophy*, *sociology* and *value*, the word *norme* is found only in the plural with the entries *depersonalize*, *model*, *chance*, *logic*, *voluntarism*, but also *value* and *taste*. Is it by chance that, once we are concerned with a reflection on values in the domain of the human sciences, “the” norm fades to the benefit of, in a certain way, its negative, that is the plural. The “norms” actually kill the “norm”... Without becoming involved in risky interpretation, it is still possible to note here a convergence that is, without exception, quite disturbing. The norm has in fact something of the instinctive: it crumbles away and is reduced when installed in the domain of thought. The dictionary thus hides without realizing it a distinction that the language instils and which a complete radiography scan reveals to us.

What are the entries affected by the idea of *norme* in PL? In the body of PL, the distinction between singular and plural is no longer insisted on: the use of the word *norme* is in fact always singular, which is somewhat in keeping with an entry which, for the word *norme*, does not offer homonymic grouping for a sense that is not clear cut as regards the use of the word in the singular or plural. If one puts aside utilitarianism, that could be linked to a praxis choice of the norm, one can distinguish in fact between two big thematic networks, each very consistent. One is defined by the specialist vocabularies, with the entries *VHS*, *DHR*, *standard (ISO)*, *orthonormal*, *yield*, *unit*, while the other is characterized by the same contrast between notions of assimilation to the norm, and on the other hand those of difference from, or opposition to it. Thus, the assimilation side involves the entries *conform*, *adjust*, *standard*, *standardize*, *purism*, *good*, *juice*, or more precisely one of its extensions, *pure juice*, to which adds the whole series of *normal*, *normalize*, *normality*, *normativity*, *norm*. As for the difference or opposition side, it is distinguished by the use of the word *norme* in the entries *slide*, *deviance*, *deviant*, *gap*, *error*, *excessive*, *anomalous*, *anomaly*, *abnormal*.

Such a scan is enlightening as regards the comparison that may be made with that of PR: the entries that call into play the word *norme* are in fact very different.

Which are the ones that are common to PL and PR, apart from words of the same family? One can only retain with certainty *deviant*, *deviance* and *gap*, on the one

hand, and *conform* (*conformism* in PR), on the other. That is really very few, and it is all the more interesting for broadening one's view of the *dictionnaire* field of the norm. On the one hand, the common concepts appear to be reinforced in their importance, on the other hand, the different entries take on, all the more strongly, distinct tones, which no doubt contribute to the specific tonality of each dictionary. One will note, among other details in PL, the presence of acronyms and, in so doing, the desire to offer some space to the new technologies.

One comment is called for: the second dictionary investigation gives an image of the word that, especially in the case of PR, seems to have escaped the notice of lexicographers, at the moment of editing an entry dedicated to a chosen word. Somehow, the lexicographer has associated, without realizing it, a group of entries with the word *norme*, thus establishing a network around the word – a network forgotten at the moment of compiling the entry devoted to the word.

### 3.4. The third dictionary investigation: a dictionary concordance radiography

The third approach is one which corresponds to the analysis of the different uses of the word *norme* throughout the dictionary: it is concerned with establishing a concordance of the occurrences of the word in a corpus consisting of all the entries in the dictionary where the looked-up word is to be found. Thus appears the dictionary uses of the word, as distinct from the entry that is devoted to it, revealing through the “cotexts” of this word (“cotext” meaning the syntagmatic or phrasal environment of the word and “context” the conceptual environment). The cotext indicates what precedes and what follows the word, the palette of usages, of uses, needed for a better definition of its semantic and syntactic nature. The agents of the norm that dictionaries act as thus deliver without knowing it a semantic and syntactic illustration of the word which nicely rounds off the entry devoted to the word.

Then too, a comparison between two dictionaries, in this case PR and PL, is particularly illuminating: the concordances are not in fact identical, revealing as they do both convergences and differentiated choices. If we continue here our comparison between the two dictionaries, it is in order to show clearly that the third type of dictionary investigation is just as revealing of perceptible differences between one dictionary and another. The accumulation of the two sets of information is bound to offer through its

complementarity an enriched image of the concept that one seeks to determine.

We will not present in full here the exhaustive analyses we have made of the dictionary concordances of the word *norme(s)* in PL and PR. Let us state simply that having established the full account of all the uses of the word throughout each dictionary, three distinct fields are clearly distinguished in the phrases in which the word is found – and we are thinking mainly of example sentences. First, we have what precedes the word, second, what follows the word, and finally, what is presented in a semantic relationship of synonymy, analogy or antonymy.

As regards what precedes the word *norme*, the concordance established for PR is specific. Either, in fact, one moves away from the norm (“deviate from the norm”, “stray from the norm”), or one moves closer to it (“conform to the norm”, “satisfy the norm”), or else one defines it (“specify the norms”, “constitute a norm”) or, finally, applies it (“implement a norm”). One point must be recognized: beyond the noting down of the concordance field of the word *norme*, with all that it will contribute to the framing of a complete definition, the norm is defined nearly as much by what is opposed to it than by what is comparable to it. It is in any case, throughout the language itself, an evident area of debate.

Examining what follows the word *norme(s)*, three sub-groups stand out. In the first sub-group, the norm is qualified by the field of specific application (the “norm of production” in the entry *standard*, the “judicial norms” in the entry *legislation*). In the second sub-group are gathered, by contrast, fields that concern society in general, and culture in general (the “norm of aesthetics, intellectual activity, morals, society”, the “norm of culture” in the entry *deculturation*). Then, a final sub-group is characterized by the taking up of a position, making a judgement of value, in relation to norms (“norms of good sense” in the entry *honest*, “norms of truth” in the entry *logic*, “norms of delicacy” in the entry *taste*).

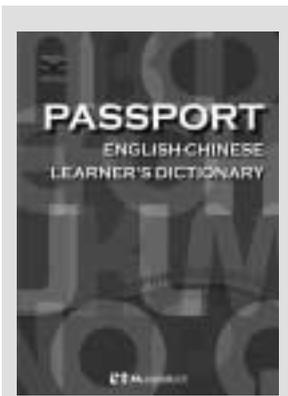
Finally, apart from those elements that occur on either side of it, and help to define the word syntagmatically in the body of the dictionary, we need to take account of the relations of synonymy, analogy and antonymy which the word *norme*, for instance, contracts throughout the different entries in which it is found. It is very interesting in this way to discover synonyms included in various entries which are not usually found in the entry for *norme* itself. Such is the case with

the synonym *majority* (entry *conform*), *habit* (entry *skid*), *real*, *ideal*, *ethical*, *law*, *precept*, *prescription*, *principle*, *standard*, etc. There we find so many interesting pieces of information about the word that are absent from the entry that defines it. As for antonyms, one notes the entry *deformity* indicating the word *norme* as its antonym.

With regard to PL, and to the context that precedes the word, it is possible to discern from the outset a first type of use characterized by the notion of distance (“departing from the norm”, “a breach of the norm”, “contrary to or different from the norm”), which corresponds to 38% of the contextual total situated to the left of the keyword in the concordance. A second type of occurrence corresponds clearly to the symmetrical attitude that consists of comparing something to the norm, with expressions like “according to the norm” (entry *abusive*), “meet the demands of a norm” (entry *conform*), which represent 41% of the total. A third type of occurrence is defined in relation to the norm in action, where one “establishes a norm” (*normative*) or else the language is “set up as a norm” (*purism*), or again where it concerns “production assessed in relation to a norm” (*yield*) or “a philosophical doctrine that makes of usefulness the principle and the norm” (*utilitarianism*), just as one can “bring something back to a norm” (*standardize*). This last group corresponds to about 21% of the total of contexts situated to the left of the word.

Upwards of the word *norme*, i.e. to its left one will discern first of all the norm determined by a technological domain, with for example the “norm of encoding a sign of telecommunication” (*standard*), the “norm for video material” (*VHS*), the “norm of production” (*standard*), the “broadcasting norm” (*DAB*), the “norm 180” (*ISO*), etc. Then one observes in a second sub-group, the norm that is “ideal and intangible” (*purism*), the norm “of every individual or social action” (*utilitarianism*). Finally, there remains the “given” norm (*deviant*).

If the “left” and “right” uses of the word are less numerous in PL, one has to agree that, by contrast, where the relations of synonymy, analogy and antonymy are concerned, the editors of PL assign to contextual synonymy and analogy a relatively more important place than do the authors of PR. The most frequent association remains the association of *norme* and *rule*, which is repeated six times (in the entries: *conform*, *deviant*, *error*, *anomalous*, *abnormal*, *norm*), of which one is in the plural (in the entry



### PASSPORT

English-Chinese Learner's Dictionary

Second Edition

BOOKMAN BOOKS

Taipei, Taiwan

May 2004

Softcover, 704 pages

215 x 155 x 30 mm.

ISBN 957-445-046-5

CD-ROM: Windows XP/

2000 or higher

[http://kdictionaries.com/](http://kdictionaries.com/products/junior/junior-chineset.html)

[products/junior/junior-chineset.html](http://kdictionaries.com/products/junior/junior-chineset.html)

From the series

KERNERMAN SEMI-BILINGUAL DICTIONARIES



## PASSPORT

Anglicko-Ceský  
Multimediální Slovník  
FRAGMENT  
Prague, Czech Republic  
May 2004  
Translation: Jarmila  
Jurecková  
CD-ROM: Windows XP/  
2000 or higher  
[http://kdictionaries.com/  
products/junior/junior-  
czech.html](http://kdictionaries.com/products/junior/junior-czech.html)

From the series  
KERNERMAN SEMI-  
BILINGUAL DICTIONARIES

*error*). It is then *model* that is cited twice (*anomaly*, *standardize*); with *standard* (*standardize*), *type* (*standard*), *fixed framework* (*derivative*), *unit of measure* (*yield*) benefiting from one occurrence. The norm is also compared with *ethics* (*good*). Finally, the norm benefits from only two qualifiers: *authentic* and *orthodox* (*juice*: sub-entry *pure juice*).

One notes here that, without realizing it, the lexicographers of PR treat the word *norme* differently according to its use in the singular or the plural, and that the authors of PL offer, beyond a similar synonymic link to *rule*, to *model*, to *standard*, new openings such as *fixed framework*, *type*, *unit of measure*, and also *ethics*.

The third investigation indeed reveals the linguistic richness of the dictionaries, a richness that is unsuspected when consulting the entry devoted to the word one is looking up. A comparable consultation of two dictionaries shows in addition at which point the uses of a word are both numerous and enlightening, with coherences that, paradoxically, escape the lexicographer at the moment that the given entry is written. Without his or her knowing it, through the uses he or she has already made of the word in other entries, a kind of consistency takes shape that may or may not be found in the entry itself. The third type of investigation is actually as useful for the lexicographer as for the person consulting the dictionary.

### 3.5 The triple investigation: offering the hidden dictionary

At the end of the third type of investigation, the consultation of the dictionary has become very operative: the visible side and the hidden side of the word are in fact then seen to be complementary. The explicit aspect, that is the entry that is open and offered for consultation, is completed by the implicit uses of the word throughout the whole body of the dictionary.

The second type of investigation enables us to observe the creation of a network of entries which lexicographers, with few exceptions, are not aware of. Even when the analogical network is indicated in the entry of the word being referred to, one notes that systematic investigation of the various entries where the word is used provides a network, a system, which is much larger and more clarifying. Whereas the second type of investigation is of a semantic and associative order, the third investigation adds to it the pertinent syntactic aspects and the uses of the word in context, while continuing to throw light on the semantic analysis of the word. The looked-up word is indeed all the more interesting to trace in that third way, since it is being used in

a situation, outside its own entry, at a stage when it enters the linguistic awareness of the lexicographer as he or she attempts to define this or that other notion. One can easily understand how a word thus forced in this way into different dictionary nooks and crannies, in uses made for it by the lexicographer outside the entry that directly concerns it, can reflect an entire semantic and semasiologic network which it is very useful to bring to light.

There is thus a hidden image of each word in the actual uses of this word outside the entry devoted to it. The dictionary is in reality a bearer of information formerly unexploitable in the paper dictionary, but which today becomes accessible in the dictionary in electronic form. The different researches that can be undertaken still remain however quite tedious for the reader, especially if one intends to preserve a structured trace of it. Also, if one might make a suggestion, in the computer era of the hypertext, why should there not be, for each entry of a dictionary in electronic form, a monograph, prepared by the lexicographer-dictionary-maker, about "the uses of the word in the dictionary"? No doubt there are tools and developments that will be particularly rewarding for both the reader and the author, who is subject to the constraints of increased consistency. The publisher will object that it is of course easy for the metalexicographer to advocate a policy of "always more", but is this concept not at the very essence of competition?

The three-fold investigation outlined here acts as a touchstone at the level of the dictionary but the effect is obviously reinforced at the level of the word if one proceeds in a comparative manner between two dictionaries. This operational setting allows one in effect to reveal in them the particular perspectives, the conscious and the unconscious ones.

To go on a journey of discovery of lost analogies, of usages nested at the heart of the entries that are different from the one devoted to the looked-up word, to find the hidden dictionary thanks to the successive radiographies, is in reality to offer an extremely rich image of the words, it's to rediscover the full message of the privileged interpreters who are the authors of dictionaries. Through a corpus so meticulously produced as is for example a dictionary such as PL or PR, to thus work out completely the usage of the word beyond the entry that is devoted to it, this is also to make lexicographical work out of a dictionary object.

Will the triple dictionary investigation not be a sort of open sesame, a magic

formula, aimed at opening the cavern onto the treasures concealed inside a dictionary? One can count on the dictionary authors to offer us the caverns that are ever richer; they will know how to provide all the keys and the treasures. I am immensely grateful to *Kernerman Dictionary News* for offering us the opportunity to summon up the existence of these keys.

## References

### Dictionaries

*Le Dictionnaire alphabétique et analogique de la langue française*. 1964. Paul Robert (ed.). Paris: Société du Nouveau Littéré. 5 volumes.

*Le Dictionnaire Hachette 2004, Noms communs et Noms propres*. 2003. Editor-in-Chief: Jean-Pierre Mevel. Paris: Librairie Hachette.

*Dictionnaire des noms de marques courants, Essais de lexiculture ordinaire*. 1998. Robert Galisson et Jean-Claude André. Paris: INA LF, CNRS, Didier érudition.

*Le Petit Larousse illustré 2004, Mots de la langue, Noms propres, Chronologie, Planches visuelles*. 2003. General Director: Philippe Merlet, Editorial Directors: Yves Garnier, Mady Vinciguerra. Paris: Librairie Larousse.

*Le Petit Robert, Dictionnaire et alphabétique de la langue française*. 1967. Paul Robert (ed.), revision directed by A. Rey and J. Rey-Debove. Paris: Société du Nouveau Littéré.

### Books and articles

Pruvost, J. 2000. *Dictionnaires et nouvelles technologies*. Collection 'Écritures électroniques'. Paris: Presses Universitaires de France.

Pruvost, J. 2002. 'À la recherche de la norme: sa représentation lexicographique et dictionnaire 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. and J.-F. Sablayrolles. 2003. *Les néologismes*. Paris: Collection 'Que sais-je?', n° 3674, Presses Universitaires de France.

Galisson, R. 1999. 'La pragmatique lexiculturelle pour accéder autrement, à une autre culture par une autre culture, par un autre lexique'. In *Études de linguistique appliquée* (ÉLA), Revue de didactologie des langues-cultures. Paris: Didier érudition. N° 116, octobre-décembre 1999, pp. 477-496.

## *Dictionnaire du français usuel:* an innovative French learner's dictionary

*Dictionnaire du français usuel* (*Dictionary of Common French*, DFU) was designed as an aid to vocabulary-related tasks, to help learners of French familiarize themselves with new words within a well-structured context.

DFU contains 442 highly structured articles, which are built around extremely common entry words. Within the articles are described and explained some 15,000 non-specialized words, meant to enable users to express themselves with ease, both in speech and in writing. Thus, DFU may be classed as a dictionary for production.

Developing this unique lexicographical structure was motivated by two main objectives: first, to develop the learner's vocabulary through the use of semantic networks, moving from what is already known towards new material, from highly frequent words to moderately frequent ones; second, to encourage the search for the right word by making comparisons with near-synonyms, antonyms, and the like.

The dictionary allows for words and their meanings to be explored in three major ways:

- (1) through semantic networks;
- (2) through themes;
- (3) through the word index.

The words are shown in simple examples that illustrate their various usages, following an order and structure that highlight and clarify each meaning as well as the lexical relationships to the entry words and to other words in the dictionary's lexical network.

One of the innovations of DFU is that various entries include two or three "star" words, for example *SAVOIR* and *CONNAITRE* [*to know*] or *FILS*, *FILLE* and *GARÇON* [*a son, a daughter, a girl, a boy*].

The dictionary articles offer topics for writing assignments and their necessary vocabulary. The structure of the articles also helps learners to prepare for text analysis by having them start with a main idea, choose a theme and then read the related articles.

This unique concept gives teachers of French all the latitude they need to help their students expand their vocabulary within a useful and well-designed framework. While it is based on strong linguistic foundations, DFU avoids the use of complex terminology whenever possible and steers clear of any linguistic jargon in order to maximize its user-friendliness. The vocabulary targeted literally opens the door to nineteenth and twentieth century French literature. However, the words are presented in present-day, easy-to-understand examples.

The result is a uniquely conceived language learning tool, featuring a well-developed structure and innovative design. The dictionary is available both in print and in personal and network CD-ROM versions.

A sample entry-page of DFU – **CONSEIL** p.188 – is available online:

<http://kdictionaries.com/kdn/kdn12/kdn12-2dfu-conseil.fdp>

### *Dictionnaire du français usuel*

*15000 mots utiles en 442 articles*

Jacqueline Picoche et Jean-Claude Rolland  
Bruxelles, Duculot-De Boeck. 2002. 1064 p.

Versions papier, cédérom (PC, Mac et en réseau)

ISBN livre 2-8011-1266-6 cédérom 2-8011-1295-X

<http://www.jacqueline-picoche.com>

(featuring the preface, methodological principles, sample entries, etc.)

<http://universite.deboeck.com/livre/>

?GCOI=28011100932940



# Some Highlights of Contemporary Hebrew Dictionaries and Lexicography

This feature highlights the main monolingual dictionaries of Hebrew in Israel today – *Milon Ariel* (Maya Fruchtman), *Milon Even-Shoshan* (Moshe Azar), *Milon ha-Hoveh* (Mordechai Mishor), *Milon Sapir* (Yitzhak Shlesinger), *Rav-Milim* (Yaacov Choueka) and its online version (Yoni Ne'eman, Rachel Finkel) – with an overall cross-review (Ora R. Schwarzwald), as well as the **Historical Dictionary of the Hebrew Language of the Academy of the Hebrew Language** along with a glossary and notes (Doron Rubinstein).

The roots of Hebrew lexicography are traced to **Rav Saadia Gaon**, who worked mostly in Babylonia in the early 10th century CE. His *Egron* (902 CE) contained nearly 1,000 Hebrew entries, and *Kitab al-sab'in lafZa al-Mufrada* had 70 (actually 90) entries translated into Arabic. In addition, he was the first to write an Arabic translation of the Bible.

The initiative for this focus has been generated by what is considered to be the first major monolingual Hebrew dictionary, the *Mahberet* by **Menahem ben-Saruq**, which appeared in Spain around 950 CE.

Issue Number 13, July 2005, will feature an article on the *Mahberet* by Aharon Maman, articles on modern Hebrew/Arabic dictionaries, and highlights of Arabic dictionaries and lexicography.



## Modern Hebrew Dictionaries

Ora (Rodrigue) Schwarzwald

Ora (Rodrigue) Schwarzwald teaches Hebrew at Bar Ilan University and part time at Talpiot Teachers College. She has a BA in Hebrew and Psychology and an MA in Hebrew from BIU, and a PhD in Linguistics from the University of Texas at Austin. Professor Schwarzwald has vast experience as a teacher and teacher trainer, is a member of several advisory committees on language policy of the Ministry of Education, and presides over the Israel Linguistic Society. Her publications include books on Hebrew and Judeo-Spanish (Ladino), *Dictionary of Linguistics and Philology* (with Michael Sokoloff, Reches, 1992), and numerous articles, particularly concerning Modern Hebrew. She was on the advisory board of *Rav-Milim* and a consultant for *Milon Sapir*.  
oschwarz@mail.biu.ac.il

The past decade has seen the appearance of a great number of monolingual modern Hebrew dictionaries, each with a different linguistic approach and editorial aims.

Besides these, there are dictionaries of proverbs and idioms (e.g. *Lashon Rishon* [First Tongue], 2000; *Nivon Ariel* [Ariel Dictionary of Idioms], 2001), clichés (*Medabrim bi-Klisha'ot* [Talking in Clichés], 2002), a thesaurus (*Mila be-Mila* [Word in/by Word], 2000), rhymes (*Haruzim le-Khol 'Et* [Rhymes for all Times], 2001), loan words (*Leksikon Lo'azi-Tvri* [Foreign-Hebrew Lexicon], 2000), lost words (*Milon ha-Milim ha-'Ovdot* [Dictionary of Lost Words], 1996), slang (*Sleng Tsva'i* [Military Slang], 1994; *Leksikon Shotrim ve-Ganavim* [Cops and Robbers Lexicon], 1997; *Sleng ve-Humor* [Slang and Humor], 2003) and professional dictionaries, all published within the past ten years.

The following nine dictionaries will be the focus of our discussion, listed here chronologically with their abbreviations – first the five general dictionaries, followed by the four junior ones.

### General Dictionaries

- *Milon ha-Hoveh* [Dictionary of Contemporary Hebrew, literally: Dictionary of the Present/Participle] by

Shoshana Bahat and Mordechai Moshor. Jerusalem: Maariv and Eitav, 1995. 815 pages; henceforth *H*.

- *Rav-Milim* [A Comprehensive Dictionary of Modern Hebrew, literally: Master-Words] by Yaacov Choueka and the Rav-Milim team. Tel Aviv: C.E.T, Miskal and Steimatzky, 1997. 6 volumes, 16+1955 pages; CD-ROM for Windows 98; online: <http://www.ravmilim.co.il>; *RM*.

- *Milon Sapir ha-Merukaz* [The Concise Sapphire Dictionary] by Eitan Avneyon. Tel Aviv: Hed Artzi and Eitav, 1997. 1126 pages; also CD-ROM; *SM*.

- *Milon Sapir* [Encyclopedic Sapphire Dictionary] by Eitan Avneyon. Tel Aviv: Hed Artzi and Eitav, 1998. 7 volumes, 2926 pages; 3 volumes, 2002; *S*.

- *Milon Even-Shoshan Mehudash u-Me'udkan li-Shnot ha-'Alpayim* [Even Shoshan's Dictionary – Renewed and Updated for the 2000s], edited by Moshe Azar and a team. Tel Aviv: Am Oved, Kineret, Zmora-Bitan, Dvir and Yediot Aharonot, 2003. 6 volumes, 2200 pages; *ES*.

### Junior Dictionaries

- *Rav-Milim ha-Tsa'ir* [Young Rav-Milim: An Annotated and Illustrated Dictionary of Contemporary Hebrew for the Young] by Yaacov Choueka and the Rav-Milim team.

Tel Aviv: C.E.T, Miskal and Steimatzky, 1996. 2 volumes, 8+1022 pages; also CD-ROM; *RMH*.

● *Milon ha-Hoveh: Milon Kis Shimushi la-Lomed* [*Practical Pocket Learner's Dictionary of Contemporary Hebrew*] by Haya Gil. Tel Aviv: Maariv and Eitav, 1996. 512 pages; *HK*.

● *Milon Sapir le-Bet ha-Sefer* [*Sapphire School Dictionary*] by Haya Gil and Orna Ben-Natan, chief editor Eitan Avneyon. Tel Aviv: Hed Arzi and Eitav, 1999. 624 pages; *SB*.

● *Milon-Kis Ariel* [*Ariel Pocket Dictionary*] by Maya Fruchtmann. Qiryat Gat: Korim, 2001. 1030 pages; *A*.

Except for *ES*, all these dictionaries are entirely new. *ES* is a revised version of *ha-Milon he-Hadash* [*The New Dictionary*], written by Avraham Even-Shoshan and first published in 1948, revised in the 70s by Even-Shoshan, updated in the 80s, and now revised by Moshe Azar and a team.

*H* and its followers – *SM*, *S*, *HK* and *SB* – form one group of dictionaries based on their verb representation system, whereas *RM*, *ES*, *RMH* and *A* form another. From here on *S* refers to *SM*, *S* and *SB*, and *H* to *H* and *HK*, unless otherwise stated. All the dictionaries attempt to present the most up-to-date list of the vocabulary of contemporary Hebrew; they vary, though, in a number of ways. In the following sections the special features of the dictionaries will be discussed. The idiosyncrasies of some of them will be described at the end.

### Entry Form

*S* is the only dictionary that presents each of the meanings on a new line; all the other dictionaries, including *SM* and *SB*, have all the meanings assembled together within one paragraph. *H* includes in this paragraph derived words, such as *dati* [*religious*] in *dat* [*religion*], as well as sub-entries; in the other dictionaries, derivatives appear as independent lexical items, except for *RMH* which lists only gerunds as a sub-entry of the verb; other sub-entries appear separately, following the explanations (see discussion below).

### Roots and Verbs

Until recently, there was a clear difference in Hebrew dictionaries between the representation of verbs and all other words. The verbs were listed as sub-entries of the (tri)consonantal root, whereas other words were listed in their alphabetical order. Among the new dictionaries, *ES* continues the old practice, whereas *H* started a revolution in listing all the

verb forms alphabetically like any other word, and all the other dictionaries (except *ES*) use this system. Thus, for example, *ES* has under the root *TBL*, in the letter *Tet* [*T*], the following verbs: *taval* [*immersed, dipped*], *nitbal* [*was dipped*], *tibel* [*dipped* – especially food in liquid (*Mishnaic* use)], *tubal* [*was dipped*], *hitbil* [*dipped; baptized*], *hutbal* [*was dipped; was baptized*], and also *tovley shaharit* [*Essenes; morning bathers (Talmudic use)*], and *tovel ve-sherets be-yado* [*religious hypocrite*]. *H* and the other dictionaries list the verb entries alphabetically, though there is a difference in their listing, as will be described in the Tenses section. Hence, the above verbs appear under the letter *Tet*, as *tovel* or *taval*, including *tovel ve-sherets be-yado* (but not *tovley shaharit*); *nitbal* under the letter *Nun* [*N*]; no *metabel*, nor *tibel* and *metubal*, nor *tubal* (because they represent archaic use); *matbil* and *mutbal* under the letter *Mem* [*M*], *hitbil* and *hutbal* under the letter *He* [*H*].

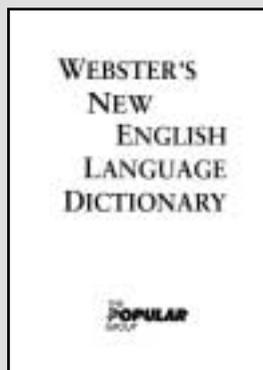
### Tenses

Contrary to the traditional way of listing verbs in their past tense forms, *H* was the pioneer in positing the present (participle) forms as the main lexical entry for all the verbs. Mordechai Mishor, one of its editors, presented the theoretical background for this system in an article published in *Hebrew Computational Linguistics* 24 (1987, see also in this issue). One of the arguments was that because every verb in its present tense form can potentially become a noun or an adjective, it is more economical for a dictionary to list the present form and catch both the nominal and the verbal functions at the same time. *S* follows the same system. *RM*, *ES*, *RMH* and *A* continue the old tradition of bringing the past tense singular form as the base form. The latter system has an advantage for the user, because it shows instantaneously the verb pattern in which the verb is conjugated (the *binyan*). Hence, if we pursue the previous example, *taval* is conjugated in *pa'al*, *nitbal* in *nif'al*, *hitbil* in *hif'il*, and *hutbal* in *huf'al*. Existing nouns and adjectives from the same roots are listed separately. For example, the verb *hidrich* [*guided; directed; instructed, coached*] is listed under *He*, whereas the noun *madrich* [*a guide; instructor, counselor; manual*] under *Mem*. *H* and its followers have one entry, *madrich*, for both functions. However, contrary to expectation, the phrase *bi-zman hoveh* [*in the present tense*] precedes the nominal meaning in *H* and *HK*.

The present tense system caused a change in the distribution of the letters: traditionally the letter *Mem* covered about

### A NOTE ON THE TRANSCRIPTION

According to the transcription rules of the Academy of the Hebrew Language, every *dagesh forte* has to be transliterated by doubling the letter; however, in this issue, this rule is applied only in the article and glossary by Doron Rubinstein. According to the above-mentioned rules, the symbol < > is used for the letter א (*Aleph*) and the symbol < > for the letter א (*Ayin*); however, in this issue, < > is used to indicate either letter. Moreover, this and other Academy rules are not applied systematically in this issue, especially concerning words and names that are used commonly in other languages, e.g. Avraham – not 'Avraham [*Abraham*], Gaon/Geonim – not Ga'on/Ge'onim [*Sage(s)*], etc.



## WEBSTER'S NEW

English Language  
Dictionary

THE POPULAR GROUP

New York, USA

June 2004

ISBN 1-59027-077-0

Paperback, 254 pages

175 x 120 mm.

[http://kdictionaries.com/  
products/us/enen.html](http://kdictionaries.com/products/us/enen.html)

Published by arrangement with  
K DICTIONARIES

13% of the headwords in a dictionary, whereas in *H* and *S* it covers about 30% of them.

### Vocabulary

1. Although most of the dictionaries claim to be contemporary, they do include obsolete items. Some of them give the old meaning rather than the contemporary one.

The example *tovley shaharit* above appears only in *ES* and the detailed *S*; the other dictionaries do not include it at all. The phrase *me'et le'et* appears in all the dictionaries; however, most of them give just the old meaning [*24 hours*], which is still used in orthodox circles for a religious *Halachic* term, while the current more widely-spread secular use [*from time to time*] is only represented in *RM*, *RMH*, *ES* and *A*.

*ES* describes all the sources used for compiling the dictionary, and it states explicitly that even rare words are included, for two reasons: anybody may encounter them in literature and should be able to find them in the dictionary; and, even if a word looks like dead wood, it can potentially become alive. The words *porfan* [*a tool to help buttoning or pinning*], *potahat* [(*Mishnaic*) *a key or lock*], *pizul* [*crossed-eyes*], *pazaz* [*was gold-like*], and many more, occur only in *ES* and the complete *S*, but not in the other dictionaries. *RM* and the junior dictionaries largely avoid ancient and too literary words, because they reflect no current use, literary or colloquial.

2. Not all the dictionaries include sub-standard, colloquial and slang words. *H* claims to be a normative dictionary, therefore such words are avoided; so are the junior dictionaries, for educational reasons. For instance, the word *tafran* [(*slang*) *down-and-out*] appears in the general dictionaries, but only in *A* of the junior ones. The popular expression *haval 'al ha-zman* (also pronounced as the acronym *havlaz*) [(*colloquial*) *a waste of time; (slang) extraordinary*] is explained on the web version of *RM*, but does not appear in any other dictionary.

3. Encyclopedic information appears in detail only in *S* and *SM*; they include names of countries (*Peru*), rivers (*Danuba*), cities (*Teheran*), etc, with a comprehensive explanation. Only wide-ranging information terms are brought in the other general dictionaries, e.g. *pahmeman* [*hydrocarbon*], *petunia*; however, this encyclopedic information is one of the causes for the wider volume of *S* and *SM*.

### Historical Information

Only *SM*, *S* and *ES* indicate for each word and meaning when it was first used: in the Bible (unmarked in *ES*, <*tm*> in *S*, *SM* and *SB*), in *Mishnaic* Hebrew (<*t*> in *ES*, <*tm*> or <*r*> in the others), in Medieval Hebrew (<*b*> in *ES*, <*yb*> in the others), and in Modern Hebrew (<*h*> in *ES*, <*'h*> in the others). New loan words are marked as well: <*t*> in *ES*, <*ms*> in the others, but not old loan words that became part of Hebrew, e.g. *pardes* [*orchard (modern); fruit tree ground (Biblical)*]; (cf. *paradise*). The information about the first appearance of a word or a meaning is totally unnecessary for the layman, but relevant for researchers and for people who are interested in the history of words.

### Etymology

Only *ES* marks the word etymology consistently. This information appears in parentheses next to the entry indicating: (1) parallel words in other Semitic languages, i.e. Acadian, Aramaic, Arabic; (2) a related root of the same word; and, (3) the source language of loan words and their original form. This is semi-scientific information that resembles *Webster* or *Oxford* dictionaries. *RM* and *RMH* indicate the language of origin only in loan words, e.g. *profil* [*profile*] from Italian. All the other dictionaries do not have it.

### Sub-entries

Sub-entries include phrases and commonly used idioms that the lexical item shares with other words. Except for *H*, all the other dictionaries list sub-entries following the basic meanings of the item. *H* has the sub-entries in the same paragraph with the meanings and derived words. All the sub-entries are listed consecutively in one new paragraph after the explanations in *SM*, *HK* and *A*. However, *RM*, *S*, *ES* and *RMH* have each sub-entry in a separate line. As stated above, *ES* has all the verb forms as sub-entries to the root.

### Examples

*H*, *SM* and *A* do not incorporate any examples after the definitions. *ES* brings citations from the listed sources after each meaning in the same font size as that of the meaning, and only the reference is marked by a smaller font. At times, especially with new words, examples are given with no indication of the source. *RM*, *RMH*, *S*, *SB* and *HK* bring invented examples in either a smaller font or italics, though sometimes citations and references are brought in *SB*. *RM* and *RMH* in particular have examples for the sub-entries as well, whereas the other dictionaries give them only occasionally.

## Illustrations

*H*, *S* and *ES* do not have illustrations. The junior dictionaries *HK*, *RMH* and *A* include pictures to exemplify some of the lexical items, and *RMH* has them in color. *RM* has a few illustrations as well, in black-and-white.

## Orthography and Vocalization

Two spelling systems are used in Hebrew, vocalized (*ktiv haser*) and 'plene' spelling (*ktiv male*, *ktiv hasar nikud*). In vocalized spelling it is not necessary to add the vowel letters *Yod* [Y] and *Vav* [V] in word medial position, whereas in plene spelling their addition is imperative. *ES* and *A* are the only dictionaries that use full vocalized spelling. However, *ES* brings the plene spelling after the vocalized word, whereas *A* lists plene spelling in an index at the end of the dictionary. All the other dictionaries use plene spelling. If there is no difference between the two spellings, the vowels are added to the entry; if they are different, the vocalized form is written immediately after the plene spelling. For instance, *tigun* [frying] is written in vocalized dictionaries as <tgwn> טגון, whereas in the others it is written <tygwn> טיגון followed by <tgwn> טגון; *netiya* – as <ntyh> נטייה or as <ntyya> נטייה followed by <ntyh> נטייה. *HK* and *SB* add the script writing of the word following the vocalized form, e.g. טגון, טיגון.

But not only are the lexical entries either vocalized or not – explanations follow suit. In *ES* and *A* the whole text is vocalized, even the originally unvocalized citations are vocalized in *ES*. The other dictionaries avoid systematic vocalization in the explanations and examples. Vowel signs are added sporadically to facilitate reading, e.g. a *dagesh* [= dot stressing a consonant] is inserted to indicate a stop or fricative p, b, k – such as פּ p or פּ f.

## Grammatical Information

1. Parts of speech are indicated in all the dictionaries. *H* and *HK* do not have an abbreviation for verbs (that appear in their present form), but the conjugation (feminine; past, future, infinitive) that immediately follows the main entry indicates it is a verb. Because only in nouns one finds gender distinctions, they are marked only for gender (<z> for masculine; <n> for feminine). Like verbs, adjectives are not marked at all, and if they do not belong to the participle forms, only the feminine form is inflected, e.g. *yahir* (*yehira*) [arrogant]. Other parts of speech are marked.

Parts of speech are marked in the other dictionaries. *S* and its followers distinguish in their dictionaries between verbal and

nominal uses of the participle forms, and the words are listed as separate lexical items, one as a verb, and the other as a noun or an adjective. *RM* and *RMH* mark the entries for Noun (<š>) and gender, Verb (<p>) and verb pattern (*binyan*), and other parts of speech. *ES* and *A* indicate a verb by <p> but nouns only by the gender; other parts of speech are marked, too.

2. Conjugations and inflections are brought immediately after the main entry in *H*, *S* and *A*, whereas they are placed after the definition (and examples) in *RM*, *RMH* and *ES*. *A* demonstrates verb conjugation and noun inflection, but not adjectives, before listing the meanings.

3. All the dictionaries except *A* bring the consonantal roots of the word (if there is one). *ES* brings it as part of the etymological information at the beginning of the lexical entry, whereas the others bring it at the end. *RM* and *RMH* are unique in that they bring the root with a list of other words derived from it, in *RMH* in the margins next to the entry, which provides important information about word families.

4. Prepositions required by the verbs are given in *RM*, *S*, *SM* and *ES* – in the first three at the beginning of the entry next to the relevant meaning, and in *ES* at the end of the entry before the inflections. They are implied from the examples in the other dictionaries.

5. Phonetic information is given through the vocalized words. Ultimate stress is unmarked; other stresses are marked in *RM*, *RMH*, *S* and *ES*.

## Register Information

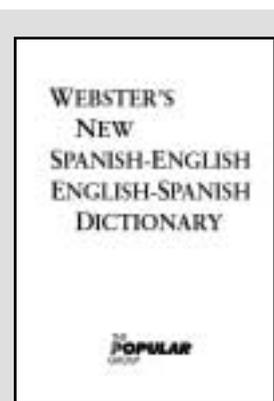
The general dictionaries and *RMH* indicate registers by assignment of phrases like *bi-lshon ha-dibur* [colloquial], *slang*, 'aga [jargon], *sifrutit* (literary), *mahshevim* [computers], *refu'a* [medicine]. This information is missing in the other junior dictionaries, which provide just normative practical vocabulary.

## Sources

Only *ES* and *SB* record in the beginning of the dictionary all the literary sources that served in compiling the entries, with the abbreviations. Most of their examples are based on these sources.

## Homonyms

All the dictionaries list homonyms as independent lexical entries. However, *H* and *HK* are the only ones that include verbs and nominal forms in the same present tense form, while all the other dictionaries use parts of speech as well as sense variations to distinguish between



## WEBSTER'S NEW

Spanish-English English-Spanish Dictionary

THE POPULAR GROUP

New York, USA

June 2004

Editor: Juan José García Pañero

ISBN 1-59027-079-7

Paperback, 254 pages

175 x 120 mm.

<http://kdictionaries.com/products/us/spen.html>

Published by arrangement with  
K DICTIONARIES



## GLOBALDIX

Multilingual Dictionary

Version 2.0

(MOT Dictionaries)

KIELIKONE

Helsinki, 2004

DVD and Online

ISBN 952-9865-09-0

- Arabic · Chinese · Danish
- Dutch · English
- Finnish (WSOY)
- French · German
- Hungarian (NTK)
- Icelandic (EDDA)
- Indonesian (Kesaint Blanc)
- Italian · Japanese
- Korean (YBM)
- Latvian (Zvaigzne)
- Lithuanian (Alma Littera)
- Norwegian (Aschehoug)
- Polish · Portuguese,
- Brazil (Martins Fontes)
- Portuguese, Portugal
- Russian (Rusky Yazyk)
- Slovak (SPN) · Spanish
- Swedish (Studentlitteratur)
- Turkish

<http://kdictionaries.com/products/multi/gd2.html>

Published by arrangement with  
K DICTIONARIES

homonyms, e.g. *tsedek*<sup>1</sup> [justice; rightness], *tsedek*<sup>2</sup> [Jupiter (the planet)]; *moreh*<sup>1</sup> [a teacher], *moreh*<sup>2</sup> [teaches, instructs] (the latter is one entry in *H* and *HK*).

### Appendices

*H* has three appendices: 1. rules of plene spelling; 2. past, future and infinitive verb forms, referring the reader to the lexical entries; 3. a list of roots, and the words that are derived from them.

*S* has five appendices: 1. rules of plene spelling; 2. new rules of punctuation; 3. Hebrew letters as numbers, e.g. <b>=2, <k>=20, <r>=200; 4. Roman letters and their values; 5. Greek letters and their names.

*ES* has the largest number of appendices: 1. proper names; 2. a list of roots and their derivatives; 3. *hapax legomena* – single words that are not new loans, yet they cannot be analyzed as derived from a root and a pattern, e.g. *shuliya* [apprentice]; 4. fused words, e.g. *mashehu* [something]; 5. a brief summary of Hebrew grammar; 6. a table of patterns and their conjugations (verbs and nouns); 7. new rules of punctuation; 8. rules of plene spelling.

*SM* and *HK* have only the rules of plene spelling at the end. *RM* has a list of English and Greek letters, vowel and punctuation signs, Biblical cantillation and music notes, mathematical signs and geometric shapes, all with their Hebrew names.

*RMH* lists infinitive forms in the appendix, e.g. *ladug* > *dag* [fish], while *A* lists the plene spelling of the words in the dictionary, e.g. <bwhn> > <BHN> [examine], both intended to enable users to find the proper words in the dictionary. *SB* has no appendix.

### Cross References

*S*, *SM*, *SB* and *HB* refer the user from the past tense form to the present form. All the dictionaries in groups *H* and *S*, as well as *RM* and *RMH*, refer the user from vocalized spelling to plene spelling, and from either loan or non-normative words to the Hebrew normative one. *RM* refers vocalized spelling to plene spelling by the use of smaller letters in the lower margin of the page. *RMH* refers gerund forms to the verb in which they are listed as sub-entries. *ES* has cross references either for weak verbs, where the root is unclear, e.g. *hipil* [dropped] <NPL>, and for *hitpa'el* verbs where the first radical precedes the *t*, e.g. *histalek* [gone] <SLK>, or when there is an equivalent proper Hebrew word for a sub-standard or loan word. *A* has no cross references at all.

### Computerized Devices

As mentioned above, *RM* and *RMH* have

CD-ROMs that include all the information presented in the hard copy and much more: there are games and puzzles in the CD version of *RMH*, as well as audiovisual devices. The CD version of *RM*, which operates only on Windows 98, includes synonyms, inflections, phrases, lists of words derived in the same syllabic structure, ways to analyze each word, etc. The online version enables the user to access the most updated information about the words, including their translation into English, in addition to all the features mentioned for the CD.

The CD-ROM that accompanies *SM* offers no more information than the book whatsoever.

### Other Idiosyncrasies

The junior dictionaries aim at school children, and all except *RMH* are published in a relatively small format to enable users to easily carry the books to school. *RMH* is the most detailed dictionary, which gives exhaustive grammatical information on roots, word families, pattern formation, meaning of patterns, and some general information in boxes outside the regular listing, e.g. the source of *demokratya* [democracy], or *ha-giben mi-noterdam* [the Hunchback of Notre Dame]. As mentioned above, the CD has auditory options, games, and a lot of additional information about the history of words and their use.

### In Sum

*H* is the most compact, handy and practical dictionary, in spite of its awkward use of present tense entries. Its normative system and definitions are good. *RMH* is the best junior dictionary, though for school purposes *A* could be recommended due to its handiness. *RM* is the most updated and thorough dictionary of contemporary Hebrew, especially because of its computerized devices. *ES* remains the most reliable dictionary for scientific purposes, particularly for those who need historical information or those who like to rely on the authentic sources from which the word derives.

## Milon ha-Hoveh and Milon Sapir

Mordechay Mishor

In this short article two dictionaries are reviewed: *Milon ha-Hoveh* [*Dictionary of the Present* (MH)] and *Milon Sapir* [*Sapphire Dictionary* (MS)]. My aim is to present these dictionaries in the light of the lexicographic principles that guided their editing, and not to criticize the end product for minor faults or random slips.

### *Milon ha-Hoveh*

The distinctiveness of MH and those following in its suit is in representing the verb in the present tense form (the present participle). The idea was conceived in the course of my work in the Historical Dictionary Project of the Academy of the Hebrew Language, and was made public in 1985 at the Hebrew University, to an audience of linguists, in a programmatic lecture where I dwelled upon the qualities needed from a practical dictionary.<sup>1</sup>

Representation of the verb in the present tense form is based on the cognizance that the present participle in Hebrew fulfills a double function, of a noun (as a substantive, an adjective or an adverb) and of a verb. In traditional dictionaries the participle in its nominal meaning will appear according to the first letter, and in its verbal meaning it will be represented by the verb, and appear according to the first letter of its root. Thus, for example, the word *menahel* would appear under the letter *Nun* [N] as a verb in the present tense (the root *NHL* [*to manage, direct*]), and under the letter *Mem* [M] as the name of a professional (the entry *menahel* [*a director, manager*]). However, this recognition demands consideration, and the distinction between the two usages is not always sharp enough, such as in the phrase “so-and-so *menahel* a factory” [*manages / is the manager of*]. Listing the verb in the dictionary in the participle form exempts the user not only from the vacillation towards making a decision, but from the very awareness of the problem. In this aspect, MH is particularly friendly to those whose linguistic knowledge is not professional, or is not professional enough, but is practical-functional. That is, the less aware the user is of the double function of the participle, the more suitable MH is for him. The constant reminder to the user, to look for the verb in the participle form (identical to the present tense), is implied in the dictionary’s name – *Milon ha-Hoveh*.

In that programmatic lecture, besides

making the suggestion about the place of the verb in a practical dictionary, there were a number of recommendations. The last paragraph in the lecture sums up the standpoint that was subsequently taken in MH: “I will end with some recommendations, considered self-evident: A practical dictionary should be written with the usual spelling, which is the ‘plene’ spelling (including the headwords). In a practical dictionary there is no need for etymology, nor attribution to historical layers; on the other hand, one should be generous with stylistic and normative evaluations. Excerpts are superfluous. Entries that are not used in our present-day contexts are superfluous. If we add to this the cancellation of redundancies that are created by distinguishing the nominal participle from the verb – we would gain another merit: that the dictionary will be short.

The work on MH began in autumn 1988. During the editing many problems rose or became acute, and the solutions were directed along one guiding principle: to help the user find the requested word quickly, to define it in short and clearly, and to avoid excessive information that might distract his mind from the text in front of him. (The search speed in this dictionary as compared to others was proved in a survey carried out by the publishing house in the early editorial stages in order to assess its commercial worthwhileness.<sup>2</sup>)

As recommended in the above citation from my lecture, the entire dictionary is written in plene spelling (the “standard non-vocalized orthography”), with the addition of auxiliary diacritical vowel marks when needed, but the vocalized (‘defective’) spelling, also called ‘grammatical’ spelling, is attached to each headword. There are references from the grammatical to the plene spelling, in accordance with which the dictionary was compiled.

A radical morphemic principle was adopted in editing the dictionary for the separation of homonyms: these were separated into different entries only if their root or inflection varied. For example, according to this principle the word *musar* was divided to two entries, one from the root *YSR* (*hatafat musar* [*moralizing*], *musar klayot* [*remorse*]) and the other from the root *SWR* (passive of *mesir* [*remove*]); along this principle the two



Mordechay Mishor was born in Argentina in 1939. He graduated from the Hebrew (Shazar) College in Buenos Aires, immigrated to Israel in 1963, and studied Bible and Hebrew at the Hebrew University of Jerusalem. His thesis “The Tense System in Tannaitic Hebrew” was completed in 1983, under the supervision of Ze’ev Ben-Hayyim. In 1967 he began to work in the department of ancient literature in the Historical Dictionary Project of the Academy of the Hebrew Language, and in 1989 became the department head. Since 1985 he has been coordinating the work on the Dead Sea Scrolls, the epigraphic documents, and the *Talmud* and *Midrash* literature, and since 1998 he has been a member of the Academy’s Grammar Committee. He has taught at the Negev (Ben Gurion) University and the Hebrew University, and cooperates with universities in Spain. Dr Mishor has written and lectured extensively in Israel and abroad. His main research interests include Mishnaic Hebrew in all its linguistic and philological aspects, as well as the living reading traditions of Jewish communities, and the relationship between the Hebrew traditions in the Middle Ages and the language revealed by the epigraphic findings. acad2u@vms.huji.ac.il



### Milon ha-Hoveh

a Practical Dictionary for  
Standard Hebrew  
S. Bahat and M. Mishor  
Eitav Publishers – Sifriyat  
Maariv, Tel Aviv, 1995  
684 pages + introduction  
and appendices

meanings of the verb *no'el* [putting (shoes) on and locking (a door)] were given in one entry, and *qeren* was divided to different entries according its plural: *qranot* [horn musical instruments, funds], *qarnayim* [rays of light, animal's horns]. Participle forms with two usages, nominal and verbal, were put under a single entry, and the different usages were indicated in the definition according to their grammatical behaviour, also for nouns used in adjectival or adverbial functions. Adverbs that are constructed on the base of a noun with a preposition (*le-'olam* [for ever], *bi-frat* [in particular]) were given as sub-entries. Adjectives that are derived automatically from substantives by adding the suffix *-y*, as well as abstract nouns likewise derived with the addition *-ut*, were given as sub-entries. This approach broadens the definition of the actual entries, but reduces considerably the number of entries and accords the dictionary a “compact” character.

The dictionary consists of the information that is indispensable for the reader of present-day texts. In this respect the dictionary is singled out precisely by what was decided not to include in it. The planned vocabulary was based on 20,000 main entries, said to represent the most frequent words, in their common meaning in Modern Hebrew. As compared with other dictionaries, which allocated to the nominal participle its own entry, which separate the noun usages according to their functions (a substantive, an adjective, an adverb), which tend to separate the homonyms according to a semantic or etymological principle and not a formal one, and which bring the adverbial function of prepositional phrases as entries on their own – compared with these dictionaries the figure 20,000 in MH is much more comprehensive. In the end we were unable to limit ourselves to this number and it was extended to 21,000 (30.7 main entries per page on average [21,000:684]).

A substantial concession was made in the grammatical information accompanying each entry. Here the principle of predictability was applied, so that only what cannot be known by the actual grammar mechanism was specified. For example, for nouns the gender was indicated and the plural noted (from the form alone it is not possible to know, for example, the gender of the word *'eretz* [land, country, ground], or its plural); the feminine form of the adjectives was also indicated, but not their plural form. The verb had an indication of the government (its depending preposition), enabling the different senses to be differentiated. In

contrast, the part of speech of the entries was not indicated, nor was the conjugation stem (*binyan*) of verbs; and these are bold innovations. Nor was the historical layer that is attributed to the word indicated, since this is professional information that does not contribute to the word's meaning or stylistic status. On the other hand, the root of each entry was added (as far as was known or possible), this being the semantic foundation relating a word to its “family” members. The data was carefully filtered, to release the user from being overburdened. Incidentally, the fact that the dictionary is short is an end result of this principle, not an aim in itself, because when the editors found it necessary they did not hesitate to expand (see below re the verb conjugation key).

Examples of usage were given only when the definition alone was not enough to make the usage of the word clear. In principle, this dictionary was not meant to teach how words are to be used (“how do you say ...?”), but to provide their interpretation (“what does ... mean?”), after the user has come across them in their natural context in conversation or text.

A lot of effort was invested in grammatical help for locating the entry being sought. A key was appended for this purpose including all the past, future and infinitive forms of the verbs in the dictionary, with referral to the dictionary entry (e.g. *leishev* > *yoshev* [to sit > sitting]). This appendix contains 78 pages. Another appendix, a key of the roots and their attributed entries (37 pages), is meant to help those interested in revealing the meaning of a word that for some reason was not included in the dictionary, according to its “family ascription”.

The official standard was set as the point of reference for marking the stylistic and normative status of the entries, that is, the standard of the Academy of the Hebrew Language. This is expressed not by censoring non-standard entries, but by “grading”: literary, popular, vulgar, slang. The fidelity of MH to the Academy settings in all domains – the spelling, the formation, the usage, the relation to foreign words, etc. – has made this dictionary an authoritative aid, which has found its place on the desks of writers and editors.

My work on MH was done while working on the Historical Dictionary Project. Shoshana Bahat was then the scientific secretary of the Academy, until her retirement in 1990. MH appeared on 1 February 1995. The last stage of our work was very intensive. During that time Shoshana Bahat fell ill, and the final crafting was cast on my shoulders.

Shoshana was not fortunate enough to see the dictionary published (she died in November 1994). While still active she managed to accompany the first steps of a large-scope dictionary, edited in the method of MH (subsequently *Milon Sapir*), as linguistic advisor, and confided in me her quandries concerning her new interest. I completed the work on MH physically and mentally exhausted, and my sole desire was to return and invest my full vigour in the Academy's Historical Dictionary, which had naturally shifted for a while from the centre of my activity.

### *Milon Sapir*

The success pronosticated by the publisher for *Milon ha-Hoveh* long before it appeared gave way in 1991 to the idea of publishing a series of dictionaries of varied scope along this dictionary's editing principles ("ha-Hoveh method"). The dictionaries compiled according to this method are called *Milonei Sapir* [*Sapphire Dictionaries*]<sup>3</sup>. Thanks to their marketing momentum, ha-Hoveh method has had a wide dissemination. I will discuss here only the dictionary called simply *Milon Sapir* (MS), or *Milon Sapir ha-Merukaz* in its full name [*The Concise Sapphire Dictionary*].

The substantive difference between MS and MH is in the extent. Yet that is not the only difference, and I will devote the rest of my words to this.

The full title of MS announces that it was edited in ha-Hoveh method. In fact, the dictionary is built on the basis of MH itself, but the entries were expanded, with the addition of, naturally, many main entries and sub-entries, and other changes introduced in the setting of entries.

The editing principles are not different from those of MH: MS has plene spelling, and the entries include the vocalized form in the grammatical spelling. The auxiliary diacritical marks were added throughout systematically (unlike in MH). The verb is represented, of course, in the participle, and the government was indicated. The different senses of the entry are numbered (in MH they are separated by a semi-colon). The roots with their derivatives are integrated in the dictionary core (in MH – in an appendix). As with MH, the headwords have no examples of usage.

Among the bold innovations of MH, the renunciation of the verb conjugation stem was adopted. The formal principle in dividing homonyms was also kept. In contrast, an indication of the part-of-speech was added to each entry (verb, noun, etc), as well as the historical layer ascribed to each meaning. The most

striking concession, and most deplorable in my view, is dropping the verb forms key that supplemented MH.

The main contribution of MS as compared with MH is, as above-mentioned, in its number of entries. As it says on the cover, the number of entries is 90,000 (81.2 on average per page [90,000:1,108]; actually, the dictionary has not a single page that contains such a vast number of entries). However, it is not the precise number of entries in the dictionary that interests us in this review, but the principle practised in expanding the vocabulary. So, in addition to new entries and idioms that were not in MH, whose due place in a dictionary like MS is in no doubt, one domain was expanded to encyclopedic scope, that of geography: "... all the countries, capitals, biggest lakes and seas in the world, tallest falls and mountains, and special sites, such as tunnels. With regard to Israel all the settlements with a population of five thousand and over have been included."<sup>4</sup> Selecting this domain for broadening a dictionary which is not encyclopedic seems somewhat strange.

I have mentioned that the part of speech was added to each headword in MS. Following it came the words' various functions, according to the parts of speech, as independent entries. For example, the entry *qashe* [*hard, difficult*] was divided to three: 1. verb, 2. adjective, 3. adverb (but the adverb *qashot* [*severely*] was included for some reason in the adjective). The same applies to names that are not in the participle, for example, the entry *yaqran* [*person charging exorbitant prices*] was split in two: 1. adjective, 2. substantive. Unlike in MH, adjectives with the suffix *-y* and abstract nouns with the suffix *-ut* were all given as entries on their own. Thus, for example, instead of the main entry *mu'amad* [*a candidate / erected*] in MH, which includes also the abstract noun *mu'amadut* [*candidacy*] as a sub-entry, MS has four entries: 1. *mu'amad* verb, 2. *mu'amad* adjective, 3. *mu'amad* substantive, 4. *mu'amadut*. In MH the entry *mu'amad* began with the verbal meaning (passive of *erects*), and then stated that in the present form there is another meaning (*candidate*); there was no note about the distinction between the substantive and the adjective, since both have identical grammatical categories: feminine and plural. In comparison with MH, MS has a simulated broadening of entries, which stems from a different editorial policy.

As in MH, there are references from the entry in grammatical (plene) spelling to the entry in non-vocalized



### **Milon Sapir**

Concise Hebrew-Hebrew  
Dictionary according to ha-Hoveh Method  
Chief Editor: Eitan Avneyon  
Linguistic Advisor:  
Shoshana Bahat  
Scientific Advisor: Raphael Nir  
Scientific Proofreading:  
Yitzhak Shlesinger  
Hed Artzi Publishers and  
Eitav Publishers, Tel Aviv,  
1997  
1108 pages (including  
roots of associated entries) +  
introduction

This article was translated from Hebrew by IJK, translation edited by Raphael Gefen. General cases of using he/him/his refer to both masculine and feminine. The Hebrew version is available online: <http://kictionaries.com/kdn/kdn12-3-2-heb.html>

(defective) spelling. MS has the addition, systematically throughout the dictionary, of references from the past form of the verb to their present form. This is conspicuous in the letter *He* [H], particularly with the beginning *hu-*, where some pages contain almost only references. In fact, the subtitle of the dictionary – “according to *ha-Hoveh* method” – makes these references superfluous.

As against the compact MH, MS looks “respectable”, impressive in its dimensions, and more generous with grammatical and historical information. The reduction in this information, which was perceived by the editors of MH as a constructive breakthrough – if such an oxymoron is permitted – of the fences of Hebrew lexicographic tradition, seemed a too far-reaching dare to the editors of MS, who toed the line with “conventional” lexicography. Anyway, MS proves that *ha-Hoveh* method is applicable in dictionaries

with varied scopes and/or of a different character, as is clear from the range of *Sapphire Dictionaries* published since then.

#### Notes

1. The lecture ‘What Distinguishes a Scientific Dictionary from a Practical Dictionary?’ was given on the anniversary of the Department of the Hebrew Language at the Hebrew University (1985), and published in *Balshanut Hofshit*, 24.69-74, 1986.
2. The essence of the survey results appeared in a letter by Ilana Shkedi from Sifriyat Maariv, published in *Leshonenu La'am*, 47.43, 1996.
3. Including *Sapphire Encyclopedic Dictionary* (7 vol. 1998; 3 vol. 2002), *Sapphire School Dictionary* (1999), *Word by Word* (2000), *First Tongue* (2000).
4. From the Preface.

## From *Milon ha-Hoveh* to *Milon Sapir*

Yitzhak Shlesinger



Yitzhak Shlesinger was born in 1938 in Czechoslovakia, and immigrated to Israel in 1949. He has a BA, an MA and a PhD from Bar-Ilan University, where he is a senior lecturer. His field of research and teaching is Modern Hebrew, especially the language of newspapers. He is the chairman of SCRIPT (Israeli Association of Literacy) and is on the board of IAAL (Israeli Association of Applied Linguistics). Dr Shlesinger is editor of the journal *Hebrew Linguistics* and on the editorial board of the journal *Chelkat Lashon*, and has written and published extensively. [yitzhak@shlezinger.org](mailto:yitzhak@shlezinger.org)

The main innovation of *Milon ha-Hoveh* was in presenting the verbs in the present participle form, as opposed to other dictionaries that inscribe the verbs according to their past form. It can be assumed that, with the name *Dictionary of the Present*, the editors wanted to point to this quality of their dictionary and perhaps also to hint it is up to date in accord with the publication date.

Six years later a new dictionary appeared, *Milon Sapir*, whose chief editor was the publisher Eitan Avnion, and the scientific editorial team included Professor Raphael Nir, Shoshana Bahat (who edited MH with Mordechai Mishor) and Dr Yitzhak Shlesinger. This dictionary had a similar pattern to its predecessor, namely, edited in *ha-Hoveh* method for the lexicographic entries of the verbs.

On the one hand, there are a number of similarities in these two dictionaries, but on the other hand there are a number of differences.

The most prominent innovation in the dictionary of Bahat and Mishor is, then, editing the verbs according to the present tense form. The editors gave in the preface several reasons for this method, some pragmatic – for ease of use, and some

editorial – considerations stemming from the ambiguity of the Hebrew present tense form, which often appears both as a verb and as a noun or an adjective.

The editors adopted this editing method in MS. The contribution of this dictionary to those involved with Hebrew language research and to anyone interested in using a dictionary from time to time is primarily its scope: MH has 21,000 entries whereas MS contains over 100,000 main and sub-entries (the MH editors were sparing with sub-entries while the editors of MS treated sub-entries at length).

However, the increase in the number of entries in MS stems also from a grammatical-linguistic decision concerning the division into grammatical categories. Thus, for example, two entries for *mukpa* [frozen]: first, the verb, including tense inflections (*hukpa*, *yukpa* [was/will be frozen]), then the adjective, including the gender and number inflections (*mukpa*, *mukpet*, *mukpa'im*, *mukpa'ot* [is/are frozen]).

This division of the present tense form into two entries according to their grammatical category reflects the grammatical system of modern Hebrew, which is indeed the main aim of MS.

## *Rav-Milim* – a Modern Dictionary for an Ancient but Thriving Language

Yaacov Choueka

1. Hebrew is probably one of the oldest languages in current usage, and its dictionary-making history goes back more than a thousand years. It is not our intention here to trace the whole history of Hebrew dictionaries or lexicographic compendia, but rather – as a contrastive background to this brief presentation of the *Rav-Milim* dictionary of Modern Hebrew (hereafter MH) – to mention the modern ones, i.e. those in vogue in the twentieth century before the *Rav-Milim* publication in 1997.

By universal opinion, the scene for the Hebrew dictionaries in the twentieth century was dominated by three major and influential works. First and foremost is the Eliezer Ben-Yehuda 26-volume dictionary, a monumental work of erudition and scholarship by “the reviver of the Hebrew language”, the publication of which started in 1908 but ended only in 1959. This is an OED-type of historical dictionary, whose glossary included not only (though mostly) Biblical and Rabbinical terms, but also whatever MH ones were available then, especially those coined by Ben-Yehuda himself. The Gur dictionary (1934-36) was really the first general dictionary of MH, quite popular in the late thirties and forties. Finally, the enormously popular Even-Shoshan one – first published in 1947-52, then reprinted countless times in various formats and numbers of volumes (with only one major revision in 1970) – completely dominated the scene in Israel for almost 50 years, being present, virtually, in most local households.

To complete this picture, one should mention three other dictionaries whose impact was rather negligible: Cnaani’s 18-volume dictionary (1960-82), Alcalay (1969-71) and Medan (1954).

All in all, then, just six dictionaries in a whole century, one of them – updated and revised in only minor ways – exclusively dominating the scene for most of the second half of that century, and none of them with any computerized components. Thus, during a period when not only the State of Israel and the Hebrew language were undergoing extraordinary dynamic cycles of changes and expansion, but the whole world was – and still is – exploring new frontiers (and devising new terms and semantic fields to describe them) in technology and science, and in intellectual

and social life, dictionary making in MH was in practice frozen for some fifty years.

This was the state of affairs in late 1992, when it was decided to compile and publish – both in print and in electronic form – a new and up-to-date illustrated dictionary of MH, *Rav-Milim* [*Master-Words*], with a shorter companion – richly annotated and copiously illustrated in color, specially adapted to young children and teenagers in elementary and secondary schools – *Junior Rav-Milim*. Here we restrict ourselves to the description of the unabridged *Rav-Milim* printed version, its underlying philosophy and some of its salient features.

2. Although not a purely corpus-based dictionary, the *Rav-Milim* design was deeply influenced by computerized methodologies and techniques of natural language processing developed since the mid-1980s, not only in its production and in its extensive cross-checking algorithms, but also in its very structure and editing method. Indeed, since the late eighties, computers have altered the way we view dictionaries, their functionality, their aims, and the degree of thoroughness, coverage, accuracy, precision and methodical writing we have come to expect from them. These influences were masterly described by Krishnamurthy in a previous issue of this newsletter (2002). True enough, the Krishnamurthy paper was about EFL dictionaries; taking into account, however, that for a great majority of the population of Israel, immigrants from all over the world, Hebrew is indeed, to a certain extent at least, a “foreign language”, such insights are highly relevant to a general dictionary of MH as well.

From its very inception, it was decided that *Rav-Milim* (RM) will be developed along completely different – in fact, radically different – lines than previously published dictionaries of Hebrew (PPDH in short), constituting an “anti-thesis” – so to speak – to them on almost each and every methodological issue of dictionary designing and editing. It differs from PPDH in the list of entries, in the entry’s structure, in the entry’s “explanation”, in the detailed and fine analysis of the various meanings of the entry and their order, in the usage examples, in the usage directives,



Yaacov Choueka was born in Cairo in 1936, immigrated to Israel in 1957, and got a PhD in Mathematics from the Hebrew University in Jerusalem in 1971. His computational work on the Hebrew language has pioneered innovations on a global scale, including the development of a computerized lemmatizer and morphological analyzer (1964), leading the Responsa corpus project (1974-86) for a full-text system for Rabbinical literature, and the *Rav-Milim* project (1989-97) for laying the infrastructure of advanced processing of Hebrew, producing, among others, *Rav-Milim* dictionary and Nakdan computerized program for automatic vocalization of texts. He was twice recipient of the annual award of the Israel Association for Information Technology, and received the Prime Minister Prize for computing in 1997. Since 1964, Professor Choueka has been associated with the Department of Mathematics and Computer Science at Bar Ilan University, where, since 1974, he heads the Institute for Information Retrieval and Computational Linguistics. ycsarah@netvision.net.il



### Rav-Milim

a Comprehensive Dictionary  
for Modern Hebrew  
Yaacov Choueka and the  
Rav-Milim team  
C.E.T., Miskal and  
Steimatzky, Tel Aviv, 1997  
6 vol. 16+1955 pages  
ISBN 965-448-323-8

### Rav-Milim Online

The online version of *Rav-Milim* ([www.ravmilim.co.il](http://www.ravmilim.co.il)), developed and maintained by Melingo Ltd, is the only full Hebrew dictionary on the web. It offers several features that are not included in the printed version.

### Morphological analysis

This is a unique feature of the online version, which is able to identify the correct lexical entry of any word, even if it has multiple inflections. For example, the user can look for the meaning of *vayashkuhu* [and they will water him], without knowing the root or basic form of the word, and find the correct lexical entry – *hishqa* [watered] – with its explanation, translation and a full grammatical analysis of the inflected form. For Hebrew this is a critical feature, because prepositional proclitics

in the registers' annotations, in the "etymological" notes, and in the thorough and detailed processing of collocations and when, where and how to include them. At the risk of being somewhat simplistic, we can state schematically that RM is intended to be synchronic and not diachronic, descriptive and not normative, explanatory and not definitional, contemporary and not archival, illustrative and not quotation-minded. Furthermore, a maximum of uniformity and consistency in the dictionary compilation was assured (and continuously checked by the computer) by having all editorial questions discussed, decided and recorded formally by the editorial committee, which counted among its members five prominent professors of Hebrew.

In the following we shall briefly present the main features of RM, most of which were "firsts" in Hebrew lexicography, and some of which have since been adapted in a few Hebrew dictionaries that were published after it.

3. The written form of Hebrew – as that of other Semitic languages – is an essentially unvocalized one, vocalization being marked by diacritical points that may appear below, above, or inside the word's letters. Such a vocalization is however rarely used in everyday writing, except for Biblical texts, poetry or (more recently) for children's books. To alleviate some of the annoying ambiguity that would thereby result in many different "readings" of a given word, it has been customary to add in appropriate positions of the word some mater lectionis: *Vav* for the vowels O and U, *Yod* for E and I, and *Aleph* for A, thus producing the so-called plene script. Still, most PPDH were edited in the formally vocalized grammatical script, and the entries were also given – and therefore sorted – in this form. We thought that such a vocalized script would seem totally out of context to any reader who never encounters such texts elsewhere, not to mention its childish (on one hand) and somewhat paternalistic (on the other hand) projection. RM is therefore edited in the plene spelling, and the headwords are given in that script, since this is exactly how the user will usually see it in a publication and look for it in the dictionary. Following the plene headword, its grammatical vocalized form is given, so as to assist the user in pronouncing it correctly and recognizing its pattern. Additionally, a pointer is given from that form, in its alphabetical position, to the plene one, just in case the user encounters that form or is extrapolating from the given plene

one and looking for it in RM. Incidentally, the number of spelling variants in Hebrew is rather large, also because of different ways of transcribing loan words from many languages over the ages, whether from Aramaic in ancient times or mainly from English most recently; since having pointers from these variants in the main dictionary page would have hopelessly encumbered it (indeed many pages in PPHD consist mostly of such pointers!), all pointers pertinent to a given page were collected and printed in a separate section at the bottom of that page.

The list of entries in RM is distinguished both by what it contains and by what it omits. Besides listing virtually every (Hebrew) Biblical word and most terms from early Rabbinical sources (except *Hapax Legomena*, whose meaning is not well understood and is inferred only from the context), the list contains every word in current usage, from all registers – from the highest literary ones to the most colloquial and vulgar ones. The only criterion for inclusion was whether such an utterance can be read or heard somewhere; if so, then we must help the user understand it, by including it and its meanings in the dictionary (this was indeed the first time ever that such terms were included in a general dictionary). On the other hand, the word's register is always clearly marked; from the highly literary (to warn the reader against using such a word in – say – asking directions) to the colloquial, vulgar or obscene, as well as corrupt form of, etc.

We included as entries also utterances that are not, linguistically, "words" of the language, but are used in certain ways specific to Hebrew, such as *tsvits tsvits* for denoting a bird song, *miaou* for a cat call, *koukourikou* for the rooster call, *sha* for requesting silence, etc.

Special consideration had to be given to the inclusion of "encyclopedic" terms and knowledge, and of terms from various scientific and technological domains. A dictionary is neither an encyclopedia nor a complete guide to the fauna and flora of the world or even of a certain region of it. As a rule-of-the-thumb, any term that may potentially occur in a general publication was included, and any term that occurs only in the relevant professional publications was excluded.

For various types of "non-linguistic" terms, the decision on whether to include them as entries in RM was made by the editorial committee, and rigorously implemented. Following are some examples of such decisions:

- No proper name of anyone (living or dead) is to be included; literary or

mythological figures are mentioned to the extent that they are used metaphorically (Samson, Venus, Casanova) or in collocations (Richter's scale, Columbus egg).

- Country names are included, along with the language(s), capital and up to three cities, and two denominations of currency – the minimal one and the main one (cent and dollar, penny and pound).

- Places in Israel are included if they have more than 5000 inhabitants as per the last Israel census.

- No specific “creations” (books, theater, arts, etc) are included, with the exception of the 24 books of the Bible and the canonical early Rabbinical sources.

- All elements of the cyclical table are included, with a uniformly designed explanation.

On the other hand, we omitted from RM thousands of obsolete entries that appeared in PPHD: words coined from the late nineteenth century and loan words from other languages that were almost never used, even words officially coined by the Academy of the Hebrew Language that did not enjoy wide acceptance, etc. Our policy was that not every word used once or twice by a writer, as great as he or she may be, should be automatically recorded in the dictionary. Delicate editorial considerations sometimes have to be applied in such cases.

Another issue that well illustrates the spirit of RM is the following. Because of the peculiar history of the Hebrew language, many words have persisted and are in current usage in certain conjugated or derived forms, while the original variant is – and was – never in use (*hav* [give], only in the imperative; *be'etyo* [because of him/it/that], only with the preposition and the pronoun). PPHD used, in such cases, to “extrapolate” and invent the presumed original form and list it as a dictionary entry. We refrained from inventing words, and such terms were given as entries “as are”, which is anyway the form in which the user will encounter such words and look for them in the dictionary.

4. “The principal reason for the existence of a general monolingual dictionary is its definitions. All the art and all the scholarship and all the scientific methods that the editors can command are required to study meanings and write definitions” (Gove, 1961).

Contrary to Gove's wise dictum, one cannot but notice that in most PPHD this aspect of dictionary compilation has been quite neglected, usually with the justification of offering one or more

synonyms of the entry. In RM, however, we fully endorsed this statement, with all its consequences and ramifications, except, maybe, for replacing “definitions” by “explanations”, since our aim was not to give an Aristotelian definition of an entry, but to explain it completely and precisely. According to the RM concept, the ultimate test of a good explanation is whether a user who has never encountered the word before can now understand it as fully and precisely as possible. On the one hand, we painstakingly analyzed and checked every word in the explanation to assure its appropriateness and pertinent coverage. On the other hand, we aimed at detailing explicitly all the nuances and shades of the basic meaning of the entry, as manifested in the different contexts in which it actually occurs. Indeed, as stated by Firth (1957), “you recognize a word by the company it keeps”.

One example should suffice to clarify this approach. The adjective *ham* [hot, warm] is defined in Even-Shoshan only as “having a more or less high temperature”. In RM, this entry details some 11 different meanings or usages in various contexts (that may well translate into different words in other languages), which an innocent reader would not be able to guess on her/his own. Thus, besides the basic meaning as in “hot soup” (vs. “cold soup”), we have “hot news” (but not “cold news”), “hot temper”, “warm heart” (the former with a negative connotation, the later with a positive one), “warm voice” (specific voice texture), “warm clothes” (the clothes themselves are not warm, they warm the body), “he is hot” (which doesn't mean he has “a more or less high temperature”, he is not sick, he just feels hot and would like to open the window), etc. Even the “Hot! Hot!” call in the hide-and-seek children's game deserves and gets its own numbered meaning. Indeed, the fine analysis of the extremely rich spectrum of the nuances of almost every word, according to the contexts in which it appears, is one of the greatest achievements and benefits of the application of computers to the processing of large corpora, and the lexicographer's efforts for collecting, classifying, sorting and adequately explaining these nuances is probably the most exciting and satisfying part of the dictionary making process.

When the meanings of an entry have changed throughout its history, they were always ordered in PPHD, traditionally, chronologically. In RM, which has always had the user in mind, meanings are ordered by decreasing frequency; the most frequent sense given first, and adequate period labels attached when necessary.

▶ and pronominal enclitics are attached to the word, resulting in many different forms for the same word. In addition, the lexical entries appear with diacritical vowels (*niqud* [pointing]), so in case of an ambiguous word like *SFR*, it is easy to find out whether we are looking for *sefer* [a book], or *safar* [counted], or *sapar* [a barber], or *sfar* [borderland], etc. These features are particularly helpful for children, new immigrants and other learners, who do not know where to look for a word in a printed dictionary, especially when it is an inflected form.

### Thesaurus

The thesaurus provides a rich selection of synonyms. Each synonym is linked to its own entry, making it very easy to ‘click’ one's way around the dictionary, travelling from a word to its synonyms, on to their definitions, and onwards through a wealth of linguistic information.

### Bi-directional Hebrew/English translation

The addition of the Hebrew-English and English-Hebrew translation provides users with the ability to translate any word. This, together with the fact that each Hebrew entry appears with vowels, makes the dictionary a useful translation resource.

### Regular update

A great advantage of any online dictionary, and certainly of *Rav-Milim*, is the ability to constantly update it. Melingo continues to maintain and enhance

the website, and regularly adds new words that are released by the Academy of the Hebrew Language, such as *hetpes* [stereotype] and *dvekanut* [perseveration], or neologisms, including the latest slang new technological terms, such as *netiqqa* [netiquette], or the latest slang, such as 'and 'en matsav [No way!]. Not less important is the interactive nature of the online dictionary, i.e., it allows users to suggest new words, or to ask for clarifications about certain examples and illustrations, etc.

### Phrases and idioms

The online version has two ways of finding a phrase in the dictionary: by looking at the list of phrases that appears under each component of the phrase, or by typing the whole phrase, a part of it or an inflected form of it. For example, the idiom *naga le-libo* [touched one's heart] can be found under *naga* [to touch] and *lev* [heart], as well as by typing any of the forms or inflections *nag'a le-libi* [(she) touched my heart], *nog'im le-libchem* [(they) touch your hearts], *yig'u le-libah* [(they) will touch her heart], etc.

### Additional features

The website includes an **automated rhyming system** that presents all rhyming words for each lexical entry and sorts them according to the rhyme quality. A **crossword solving** feature enables the user to insert the known letters and number of letters in the word, and to receive a list of possible words from which the correct answer can be chosen.

Finally, an explanation is almost always followed in RM by one or more examples of usage, which only rarely are quotations from canonical writing. In nearly all cases, examples were carefully crafted to add interesting and useful details to the explanation.

5. One of the impacts of large corpora processing on linguistic studies in general, and on dictionary making in particular, since the mid-eighties, has been the recognition of the critical importance of collocations in defining the language elements and structure. If this is true for European languages, how much more so for Hebrew! Indeed, with the world dynamically revolving around us, the Hebrew language has constantly had to acquire and absorb numerous new words from the various domains of modern life activities. Although some new terms are adapted as loan-words "as is" and easily become part of current Hebrew, in many cases, however, Hebrew – being a Semitic language with a structure of 3- (or 4-) letter roots and derivation patterns – is quite resistant to such assimilation. A common productive solution is to have a two- (or three-) word Hebrew sequence to represent a new concept. A large number of single-word nouns in English, for example, such as *school*, *hospital*, *lawyer*, *accountant*, are represented in Hebrew by a two-word sequence.

In spite of that, the treatment of collocations in PPHD has been rather poor, to say the least. Very few collocations found their way into these dictionaries; phrasal collocations, idioms and even proverbs (!) were all mixed up; no clear guidelines were respected in terms of where and how to have the collocation's main entry (in fact, in an extreme example, a 4-word collocation actually appeared in 4 different entries with 4 different explanations!), or in terms of how to deal with, and uniformly represent, the "empty places" in some of these collocations, etc.

Having researched the problem of collocations already in the eighties (see 1983, 1988), I was strongly biased in favor of a comprehensive, systematic, rigorous and consistent treatment of the collocational part of RM. A small sample of the new features introduced in this endeavor now follows.

- To the question of when does a sequence of two or more words deserve its own entry in the dictionary as a collocation, a common answer is: when the meaning of the sequence is not the total sum of its components' meanings, and cannot be guessed from it. This is indeed an

important criterion, but it is far from being unique. We delineated 12 different criteria that can justify such an inclusion, and every potential collocation was tested accordingly.

- Almost 10,000 new collocations were added in RM that never appeared before in PPHD. This is an extremely high figure when taking into account that the total number of (single-word) entries in PPHD is of the order of 35,000 entries only.

- Proverbs (e.g. 'not all that glitters is gold') were completely banned from the dictionary; phrasal collocations and idioms were sharply separated.

- Strict rules were set up and followed on where to introduce the main entry of a collocation and its explanation. The explanation appears, of course, only once, but pointers to that occurrence are given from every word of the collocation.

- Collocations were tagged by part-of-speech tags: nominal, verbal, adjectival, adverbial, etc. When necessary, morphological variants were added.

- Possible additions, omissions, replacements, etc, in the collocation text were marked clearly, in a uniform way.

With these steps and more, we indeed believe that the collocational component of RM has made an important contribution to the clarification and systematic study of collocations in Hebrew.

To sum up: RM was a bold step taken to bring modern methodologies, trends and techniques to Hebrew dictionary making, applying overwhelmingly a computerized approach to its compilation and checking procedures. We believe that it has thus set a new standard of precision, coverage, methodology and systematization that will be hard to ignore.

### Acknowledgements

*Rav-Milim* was the product of a large team of dedicated and competent professionals whom I have had the good fortune of leading and directing. I thank them all for the wonderful job they did, and I wish I had the space to mention them all by name. I would like still, at least, to thank personally Uzi Freidkin, Dr Haym Cohen, Yoni Ne'eman and Sara Choueka. May they all be blessed in their future endeavors.

### Dictionaries and articles

**Alcalay, Reuven.** 1969-1971. *Milon 'Ivri Shalem (The Complete Hebrew Dictionary)*. 3 vol. Ramat-Gan: Massada.

- Ben-Yehuda, Eliezer.** 1908-1959. *Milon ha-Lashon ha-Ivrit ha-Yeshana ve-ha-Hadasha (A Complete Dictionary of Ancient and Modern Hebrew)*. 16 vol.
- Cnaani, Yaacov.** 1960-1982. *Otsar ha-Lashon ha-Ivrit (Treasure of the Hebrew Language)*. 18 vol. Ramat-Gan: Massada.
- Even-Shoshan, Avraham.** 1947-1952. *Ha-Milon he-Hadash (The New Dictionary)*. 5 vol; 1966-1970, 7 vol; 1997, 5 vol. Jerusalem: Qiryat Sefer.
- Grazovsky (Gur), Yehuda.** 1934-1936. *Milon ha-Safa ha-Ivrit (Dictionary of the Hebrew Language)*. 3 vol; 1947, 1 vol. Tel Aviv: Dvir.
- Medan, Meir.** 1954. *Me-'Aleph 'ad Tav – Milon Ivri Shimushi (From A to Z – a Practical Hebrew Dictionary)*. Jerusalem: Achiasaf.
- Choueka, Y., S.T. Klein and E. Neuwitz.** 1983. 'Automatic retrieval of frequent idiomatic and collocational expressions in a large corpus'. *ALLC Journal*, 4.34-38.
- Choueka, Y.** 1988. 'Looking for needles in a haystack or: locating interesting expressions in large textual databases'. *Proceedings of the RIAO Conference (Cambridge, MA)*. 609-623.
- Firth, J.R.** 1957. 'A Synopsis of Linguistic Theory, 1930-1955'. *Studies in Linguistic Analysis*, 1-32. Oxford: Blackwell.
- Gove, P.B.** 1961. 'Linguistic Advances and Lexicography'. *Word Study*, October 1961, 3-8.
- Krishnamurthy, R.** 2002. 'The Corpus Revolution in EFL Dictionaries.' *Kernerman Dictionary News*, 10.23-27.

### Melingo

Melingo is a Tel Aviv-based subsidiary of Encyclopaedia Britannica, focused mainly on Natural Language Processing for Arabic and Hebrew. Its tools are incorporated in a variety of technologies, including leading enterprise and web search engines, data mining and extraction, automatic speech applications and computerized dictionaries.

Yoni Ne'eman

Rachel Finkel

www.melingo.com

## Milon Kis-Ariel

Maya Fruchtman

*Milon-Kis Ariel* [*Ariel Pocket Dictionary*] was compiled in response to concerns of teachers and families from the south of Israel, who wanted an up-to-date, learner-friendly Hebrew dictionary, which will be aimed explicitly at present-day young pupils, including new immigrants. This need arose because the existing school dictionaries were generally viewed as out-of-date, offering sloppy definitions, using archaic language, and lacking current words and meanings.

To achieve the goal of creating this Modern Hebrew learning tool, we established a competent editorial team consisting of experienced educators and linguists as well as professional specialists. In cooperation with a group of teachers, parents and students, we studied the specific requirements from such a dictionary, in relation to what was really necessary for the pupils at school and at home. The school dictionary was conceived as a first step in a larger project, having its own lexicographic database sources developed in relation to the target audience.

We decided to focus, in particular, on the following subject areas: flora and fauna, youth life, sports, civics and state institutes, communications, computers and technology, geography and history

(particularly of Israel), literature, Judaism, loan words, current events and economics. We wanted to include old words besides modern ones, but to define them briefly and clearly, in a modern way.

The dictionary is arranged in a straightforward alphabetical order, according to the first letter of the word, no matter what the Hebrew root is. The verbs are presented in the traditional way – based on the model of the past tense masculine singular. The entries include only definitions, not examples of usages, but they are often exemplified by the sub-entries. Vocalized spelling is used, in accordance with the rules of the Academy of the Hebrew Language. There are illustrations, for further emphasis or clarification of certain entries, and in order to make the book more lively and attractive.

The total number of entries includes 14,000 words and 4,000 phrases, of which as many as 3,000 appear for the first time in a Hebrew dictionary, for example: *Alzheimer's (disease)*, *anorexia*, *Intifada*, *Eurovision*, *Druzi* [*Druse*], *hashman* [*cardinal*], *divkit* [*sticker*], *hor ba'-ozon* [*hole in the ozone*], *duty-free*, *telemarketing*, *tampon*, *home'opatia* [*homeopathie*], *alpaka*, *hetsion* [*median*], *štrudel* [the symbol '@', officially *grukhit*].



Maya Fruchtman studied Hebrew at Tel Aviv University and teaches at Bar Ilan University. Professor Fruchtman is the editor of the *Ariel* dictionary and lexicographic works, and has written extensively on the Hebrew language. [frucht@zahav.net.il](mailto:frucht@zahav.net.il)



### Milon-Kis Ariel

Maya Fruchtman  
Kor'im, Qiryat Gat, 2001  
1030 pages  
ISBN 965-515-000-3

## Milon Even-Shoshan, Revisited

Moshe Azar



Moshe Azar has a PhD in General Linguistics from the University of Nancy (1970). He teaches Hebrew Linguistics at University of Haifa, and is a member of the Academy of the Hebrew Language. Professor Azar has written books and many articles concerning the Hebrew language, and the argumentation and interpretation of legal texts. [azar@research.haifa.ac.il](mailto:azar@research.haifa.ac.il)

The lexicographer Avraham Even-Shoshan entitled his dictionary, published in 1948, *Milon Hedash* [*New Dictionary*]. This title was changed, in a later edition, to *ha-Milon ha-Hadash* [*The New Dictionary*]. Now that Even-Shoshan is no longer alive, the title was changed to *Milon Even-Shoshan* [*Even-Shoshan Dictionary*, (ESD)]. This name was deemed proper, on one hand, to express the esteem and admiration felt toward the founder and first editor, and, on the other hand, to formally endorse the name by which the first *New Dictionary* has become known. Now, with this publication, Avraham Even-Shoshan joins Noah Webster, Paul Robert, Pierre Larousse, and other compilers of monolingual dictionaries that bear the name of their author. Even-Shoshan not only gained the privilege of having his name become synonymous with Israeli Hebrew lexicography, but he is also seen by many Israelis as the symbol of standard Hebrew.

The sub-heading of ESD is “A General and Comprehensive Lexicon of Contemporary Hebrew Combined as Derived from all Hebrew Periods”. This was indeed the description of the dictionary from its inception. The dictionary attempts to be, as far as possible, a complete compendium of everyday spoken and written Hebrew, and, being a general dictionary, it intentionally does not include a lot of scientific and technological terminology, sufficing with words and expressions that educated people may encounter in non-professional reading and conversation.

Another characteristic of the dictionary is that it is constructed in a manner that describes the vocabulary of contemporary Hebrew in combination with all its historical periods. This method is based on the assumption that present-day Hebrew cannot be properly understood without considering it as yet another stage in the long history of the language. No language exists on its own without reference to its history, certainly not Hebrew, especially regarding the most common words inherited from the Biblical and Talmudic periods. Thus, the vocabulary of the Hebrew language has increased without deleting old words and senses. Moreover, disconnecting Modern Hebrew words from their past would mean disconnecting them from the heart of the religious, cultural and literary legacy

stored in the Hebrew language throughout the generations. Even-Shoshan correctly decided to preserve the uniformity of the Hebrew language by indicating the period (Biblical, Talmudic, Medieval, Modern and Modern Foreign) in which the word and meaning became part of the lexicon for Hebrew speakers.

ESD, then, is characterized by documenting and interpreting Israeli contemporary Hebrew (Modern Hebrew), including many foreign and non-standard words and idioms, as well as all the words that appear in the Old Testament with their accepted meanings, and a large selection of Talmudic and medieval words and meanings. The reason for this is, as Even-Shoshan wrote in the Preface to the first edition: “There is almost no word, as uncommon as it might be, which the Hebrew reader might not encounter by chance, through reading or conversing, and it is his right that his dictionary explain it to him. It is, then, better for the dictionary to ‘err’ here and there by including more words, than by leaving out words which in contemporary Hebrew may be considered as ‘dead’. Yet, it is not impossible that, one of these days, our language will revive these words in their original form or with some minor change.” Even-Shoshan well expressed the dominant passive role of any general monolingual dictionary. A dictionary of this kind is intended first of all for understanding unfamiliar and unclear words, which a reader may encounter, and also for enriching the linguistic knowledge about these words, including linguistic history.

Even-Shoshan’s mixture of old and new is not very different from that of the great European monolingual dictionaries. These dictionaries, especially those dedicated to national languages, are distinguished by associating a contemporary language with its historical sources and roots. A national language is not a mere vehicle for communication; it also preserves culture. Therefore, listing words and describing their present usage without associating them to their past, strips them of nuances that can be grasped only by knowing their historical background.

The language that is inscribed, described and interpreted in ESD is standard, normative and correct Hebrew, provided that the words defined are not marked as colloquial or slang. Evidence of how Even-Shoshan’s various editions have

been considered as the authoritative dictionary of normative language can be found in written verdicts of Israeli courts, which based their decisions concerning the interpretation of disputed words on Even-Shoshan's definitions. Sometimes the court accords the dictionary the explicit title *normative*. For example, a verdict given by the Jerusalem District Court sitting as an appellate court in 1996 (criminal appeal 96/19; see also *Leshonenu La'am* 47.65, 1997) says: "...we have brought above the definition of Even-Shoshan dictionary, the normative dictionary of the last decades".

ESD is in fact the only new Hebrew dictionary constructed in an integrated manner, which is characteristic of dictionaries known by the designation 'academic'. As with other academic dictionaries, it is general and comprehensive, and, at the same time, standard. Like them, it describes the contemporary linguistic facts with a strong inclination to historical and etymological facts. Like them, technical and scientific terms are only partially represented, and encyclopedic items and information is restricted. And, as in any regular academic dictionary, the most important element is the standard language. Standard language entries are treated with the utmost attention and are described in full detail. They are usually accompanied by quotations with exact references. The grammar, morphology (roots and stems), and syntax appendices provide indispensable tools for the comprehensive description of standard language.

The decision to retain the policy of the former editions regarding vocalized spelling, and fully punctuate all words according to the rules of the Academy of the Hebrew Language, is also congruent with viewing the dictionary as intended for all Hebrew speakers and students. Native speakers of Hebrew are also included in its student population, not because Hebrew is unique in being still in a state of transformation from a language preserved in books into a modern living one, but because every language that serves as a cultural and literary medium is always in the process of being learned. There can be no doubt about the importance of vocalized spelling for all Hebrew speakers, and especially for students of Hebrew, whether native or non-native speakers. If a dictionary attempts to be user-friendly, it cannot do without sufficient spelling,

not only with respect to words which users look up, but also for definitions and quotations. As Even-Shoshan wrote in the preface to the first edition: "The full and exact spelling illuminates the eyes of the reader, it eliminates doubts and mistakes when reading a quotation, and thereby facilitates comprehension of the subject matter. Also, vocalized spelling opens the dictionary before larger and more popular circles – students, new immigrants and others".

The new and updated ESD undertook the task of including all the lexical, grammatical, vowel indications, and syntactical innovations that were created after the last edition of the dictionary was published, as well as those created before the last edition but which, for whatever reason, were not included in it.

The lexical innovations may be in the form of new Hebrew words (neologisms), standard or non-standard, or foreign. They may also appear as new meanings added to old words or as idiomatic phrases. Thousands of new lexical items found a place in the new edition, from every corner of life and all types of writing and speech, whether created spontaneously by speakers or writers, or coined by the Academy. Among the sources from which they were drawn, it is worth mentioning the vast corpus of fiction written since the early 1970s, which was not cited in previous editions.

Finally, although ESD assembles colloquial words and meanings, and to some extent even words that may be considered gross and vulgar, spelling always follows the standard grammatical rules. The dictionary also takes no consideration of the fact that many users of the language, including interviewees on radio and television, pronounce certain words in a non-standard way. An academic general language dictionary is not meant to reflect everything that happens in the language. Ungrammatical forms and pronunciations may become part of language (or of a subset of it) some day, but they may also disappear. ESD, as any general monolingual dictionary, is pre-eminently designed to serve as a practical, academic, standard dictionary, and as such, does not include information about inconsistency or instability concerning pronunciation.



### **Milon Even-Shoshan**

Renewed and Updated for the 2000s

Moshe Azar and a team

Am Oved, Kineret, Zmora-

Bitan, Dvir and Yediot

Aharonot, Tel Aviv, 2003

6 volumes, 2200 pages

ISBN 965-517-059-4

# The Historical Dictionary of the Hebrew Language

Doron Rubinstein



Doron Rubinstein was born in 1964. He studied Hebrew Language and Bible at Tel Aviv University, and taught Hebrew at high school, at Tel Aviv University and at the David Yellin College of Education in Jerusalem. He has been working at the Academy of the Hebrew Language since 1996, and in 2000 became the co-ordinator of the department of modern literature of the Historical Dictionary Project. [drubinst@fastmail.fm](mailto:drubinst@fastmail.fm)

## Founding the Academy of the Hebrew Language and setting up the Historical Dictionary Project

This article introduces readers to the major research project of the Academy of the Hebrew Language – the Historical Dictionary of the Hebrew Language.

The Academy of the Hebrew Language (AHL) was established following a decree of the *Knesset* [Israeli parliament], “The Supreme Institute of the Hebrew Language Act, 5713-1953”, passed on 27 August 1953. The Academy replaced an earlier institution, ‘The Hebrew Language Council’. The first plenary session was held on 16 November 1953, about three months following the legislation. Its first President was Naphtali Herz Tur-Sinai (Torczyner, 1886-1973, President of the AHL from its foundation until his death). The next Presidents were Professor Ze'ev Ben-Hayyim (born 1907, President 1973-1981), and Professor Joshua Blau (born 1919, President 1981-1993). The current AHL President, since 1993, is Professor Moshe Bar-Asher (born 1939).

About a year after its foundation, on 20 December 1954, at its eighth session, the Academy's plenum approved an agreement between the AHL and Bialik Institute publishing house of the Jewish Agency, “to publish a historical dictionary of the Hebrew language, containing the lexicon of Hebrew words and their meanings throughout history, from ancient times to our age. This dictionary shall be called ‘The Historical Dictionary of the Hebrew Language of the Academy of the Hebrew Language’” (*Proceedings of the AHL* 1-2.45, 1954-1955). At its next plenary session, on 2 March 1955, the plenum appointed the editorial board of the Dictionary. However, four more years passed before its shape was determined.

The founder and first editor of the Dictionary was Ze'ev Ben-Hayyim. He served as editor until his retirement in 1992, when the current editor, Professor Abraham Tal (born 1931), was appointed.

Ben-Hayyim spent a few months in Europe on his scientific work (end 1957-beginning 1958), and took advantage of this stay to travel and study historical dictionary projects across the continent. In a series of sessions of the editorial board (May-June 1958), he reported his findings and proposed a plan to organize the preparatory work of the Historical

Dictionary (*Proceedings of the AHL* 5.62, 1958). Ben-Hayyim suggested basing work on the Dictionary from its outset on the use of computers, a revolutionary idea in those days.

On 2 January 1959, the editorial board endorsed Ben-Hayyim's proposal to “decide on one comprehensive historical dictionary to embrace all periods” (*Leshonenu* 23.118, 1959, and *Proceedings of the AHL* 6.87, 1959). This dictionary “may also serve as an excellent basis for the preparation of special Period Dictionaries at some later date.”

This decision determined the shape of the Project from that day onwards, and during the following months, in April-May 1959, the preparatory work for the Dictionary, headed by Ben-Hayyim, started according to that program (*Proceedings of the AHL* 6.87, 1959).

The first attempt to compile such a comprehensive dictionary was made by Eliezer Ben-Yehuda (1858-1922), known as the “reviver of the Hebrew language”: *Millon ha-Lashon ha-Ivrit ha-Yeshana ve-ha-Hadasha* (*A Complete Dictionary of Ancient and Modern Hebrew*, 1908-1959).

## The Historical Dictionary Project through the prism of its publications: from the *Source Book* of ancient literature to the *Ma'agarim* CD-ROM

The publications of the Historical Dictionary Project (HDP) offer a fine prism to describe its history: each one of the publications constituted a mid-stage in the process of the Project, and each of them reflected its accomplishments until its publication date. In this section we will review the history of the Project through its publications, practically all of which are publications of the department of ancient literature. One more important publication – the sample pamphlet of the root 'RB – will be discussed at the end of the article.

The beginning of the Project was devoted to seeking an appropriate way within the framework set by Ben-Hayyim. What did this imply? The HDP had to undertake two tasks before the Dictionary writing could begin: the first – preparing the foundations of the Dictionary and determining its scope, and the second – assembling the sources, that is assembling the texts from which the references would be derived and on which the Dictionary would be based.

This article was translated from Hebrew by IJK, the translation was edited by Raphael Gefen, then additions to the English version have been introduced by DR and IJK. The Hebrew version is available online: <http://kictionaries.com/kdn/kdn12-3-7-heb.html>

Preparing the foundations of the Dictionary – setting the precise working procedure and determining the database structure – lasted several years. The results in their different stages appeared in the first two publications of the HDP: *Barayta di-Melekhet ha-Mishkan* and *Megillat 'Ahima'az*, as described below.

Compilation of sources for the Dictionary was completed in 1963, with the publication of *Sefer ha-Meqorot* [Source Book] for sources “from the canonization of the Bible until the conclusion of the Geonic period”. A second edition was published in 1970. The essence of the book was the HDP’s intention to process all Hebrew writings of the time, which are known to us. The book provided a chronological list of sources that were candidates for processing in the Dictionary, indicating the literary genre of each: inscriptions, documents of the Judean Desert, Rabbinical literature (Talmudic and Midrashic), etc. Further details were given in a separate list: a list of the *mesivot* [transmissions], that is the “good” manuscripts of each work, a list of its important printings, a list of important studies dealing with it, etc. In due course these data were concentrated in the bibliographical records of the Academy’s database. Each source in the HDP database, without exception, has a bibliographical record that includes the above-mentioned data.

The introduction to the *Source Book* served Ben-Hayyim as a forum to raise some of the Project’s problems, two in particular: the problem of chronology and the problem of manuscript selection.

The problem of chronology concerns the time gap between the source itself and the surviving manuscripts in which it was transmitted, i.e. appeared in writing. The transmission of the *Mishnah* is a good example of this problem: the *Mishnah* was compiled in the beginning of the third century CE, but its best manuscript, the Kaufmann Manuscript, dates from the twelfth century CE, nearly a thousand years later! In any assertion whatsoever about Mishnaic Hebrew (and about the language of any other such source) the researcher must take into account the huge gap in time between the *Mishnah* and the Kaufmann Manuscript.

The second problem concerns determining the choice of manuscripts that will be utilized for the purposes of the Dictionary. The Project did not wish to exploit in full all the manuscripts of every given source. Instead, it wanted to utilize manuscripts that met certain chosen criteria, the principal being that only the

earliest manuscripts of each source were to be utilized for the purposes of the Dictionary. A single manuscript out of these was chosen to serve as the “main transmission”, the manuscript according to which the text of the said source was installed in the Academy’s database.

At the same time that the *Source Book* was being completed, first attempts were made to process full sources by computer. In 1961, before the third World Congress of Jewish Studies, there appeared the booklet *Samples of a Concordance and Word Collections of Barayta di-Melekhet ha-Mishkan – an automation attempt in Hebrew research with IBM computers*. This booklet presents the first attempt to process a Hebrew text with the aid of computers, although the processing procedure that is presented in it is not the one that the Project has finally adopted.

Four years later, in 1965, before the fourth World Congress of Jewish Studies, there appeared the booklet *Megillat 'Ahima'az – text, concordance and lexical analysis*. The editorial board wanted thereby to show its method of preparing the material for the Historical Dictionary by means of the computer and of the technological means at its disposal then – from setting up the source to setting up its concordance. This booklet already presents the method of the Dictionary, as was destined to be applied with very few changes almost until the present time.

This method was described in full detail in the booklet *The HDP and the Ways of its Making*, published in 1969, before the fifth World Congress of Jewish Studies. An earlier version of this method was described already in the *Source Book*, pp.232-235 (reprinted in *Leshonenu* 27-28.171-175, 1964).

It is important to emphasize one feature of the concordance. Its structure and its code system enable a uniform treatment of texts from all layers of the Hebrew language: Biblical Hebrew, Mishnaic and Talmudic Hebrew, Medieval Hebrew and Modern Hebrew. The mere existence of this possibility – the possibility to treat texts from all the language layers in one system – is a remarkable illustration of the historical uniformity of Hebrew morphology. Despite all the – very many! – differences between the historical layers of the language, all of them together and each and every one of them on its own represent aspects of **one** language – Hebrew.

Eight years after the publication of *Megillat 'Ahima'az*, in 1973, there appeared *The Book of Ben Sira: text, concordance and analysis of the vocabulary*. The book contained the Hebrew source of the *Book*



### Ma'agarim CD-ROM

The Academy of the Hebrew Language, Jerusalem, 1998  
Second Edition, 2001  
Databases of the *Historical Dictionary of the Hebrew Language*, from the second century BCE to the eleventh century CE

## HEBREW SPELLING

Hebrew is a Semitic language, related to Aramaic and more distantly to Arabic. The Hebrew alphabet (*aleph-bet*) consists of 22 letters, all of which represent, primarily, consonants. Four of these – *Aleph, He, Waw* and *Yod* – serve also as vowel letters [*matres lectionis*].

Throughout history, two systems of Hebrew spelling have evolved: vocalized and unvocalized, with many variations within each system. In vocalized spelling, called also ‘defective’ or ‘grammatical’ spelling, all of the vowels are indicated by diacritical vowel points (*niqqud*), and some of them also by vowel letters. In unvocalized spelling, called also ‘plene’ spelling, the vowel points are omitted, but some of them are substituted by additional vowel letters (*Waw* and *Yod*). The latter is the spelling system commonly used in Israel today.

Rules regulating unvocalized spelling were first issued by the Hebrew Language Council and later by the Academy of the Hebrew Language, most recently in 1993. However, some of these rules are often disputed, and many writers and publishers do not fully apply them. Moreover, since the very existence of two spelling systems within the same language is quite confusing, it is gradually being realized that the whole subject should be reconsidered. A concrete proposal for a modest reform was raised recently by Mordechay Mishor in a special session of the Academy of the Hebrew Language (May 2004).

of *Ben Sira* according to the manuscripts that served the HDP, the concordance of the book, as it was set up in the Project, as well as numerous lexicological lists; an index of the *tevot* and their lexicographic entries, a list of the entries in descending frequency order, etc. As pointed out by Ben-Hayyim in the preface, “the editorial board of the Historical Dictionary sought to present this work in the usual form of a printed book *without its being typeset*. The reason for this was that typesetting and everything connected with it were liable to impair the accuracy of the material and its fidelity to the MSS, which had been achieved with such great effort by the mechanographical method. [...] This is perhaps the first time that an attempt has been made to apply this form of printing to a Hebrew book of so extremely complicated [...] nature, both because of the way the contents are arranged and the range of letters and symbols used [...]” (p.viii).

The next two publications of sources of the Historical Dictionary and their concordances were no longer printed on paper. In 1988 a part of the database appeared on microfiche: *Materials for the Dictionary – Series I – 200 BCE-300 CE*. These microfiches were edited in a similar manner as the *Book of Ben Sira* from 1973.

In 1998 there appeared the *Ma'agarim [Databases]* CD-ROM, including the Historical Dictionary’s database from the period of the second century BCE to the first quarter of the fifth century CE. A second edition of this CD was published in 2001, where the database was expanded and included texts up to the first half of the eleventh century.

### The department of modern literature and the department of medieval literature

Until 1969 the activity of the HDP focused on ancient literature. That year the department of modern literature was established. This department was intended to process sources from the years 1750-1947. Two events demarcate this 200-year period: beginning in the sixth decade of the eighteenth century, with the publication of the journal *Qohelet Musar* (edited by Moses Mendelssohn; only two issues of which appeared), among the heralds of the *Haskalah* period in Europe, and ending in the year 1947, before the establishment of the State of Israel. This date, which is some decades ago now, offers a better perspective for examining the state of the language in that period.

Two historical facts are at the essence of the work of the department of modern

literature: one – the multitude of Hebrew sources in the said period, and the other – the printed book (or journal) being the major means for disseminating the writings of that period.

Thus, the editorial board made two decisions that differentiated between the work of the department of modern literature and that of the department of ancient literature: first – to process only a selection of sources from the relevant period, not all of them, and second – to process each source according to its first (and sometimes only) publication in print. Excluded from the first decision were three great authors, all the work of whom it was decided to process: Mendeli Mokher Sefarim, Hayyim Nahman Bialik and Shmuel Yosef Agnon. The practical meaning of the second decision made it necessary to find the place of first publication of, for example, each of Agnon’s hundreds of stories – and to locate the journal or the book where it was printed.

In 1977 the first pamphlet of the *Source Book* appeared “for the period from 1750 onwards [...] a selection of writings from the Hebrew belles-lettres (1860-1920)”.

The department of medieval literature was established at the end of 1999, with the aim of treating literature of the 700-year period between 1050-1750 CE. At present this department deals with the Geonic literature.

### Intermediate summary

As of April 2004, the HDP databases have been supplied with about 3,500 sources ranging from one or a few words (inscriptions and old coins) to tens of thousands or hundreds of thousands of words (the *Mishnah* and the *Tosefta*, books from the eighteenth and nineteenth centuries, etc). The total number of word tokens in these sources is about 18,500,000, about 8,500,000 of which have received lexicographical entries.

At present, the computer department of the Academy is developing a new software for the HDP, and the three Project departments are gradually proceeding to use it. The transformation to the use of the new software is accompanied by a real revolution in the structure of the Dictionary’s databases, and it necessitates their comprehensive updating. The task of updating the ancient literature database has now been completed, and serves as the base for updating the modern literature database, which is currently being done.

There are several aspects to this updating, the principal ones of which are: completing the vowel points that

for economy reasons was omitted before (indicating the first two vowel points in each word sufficed to distinguish between most of the homographic words in the language); setting standard pointing to those entries, which for various reasons had different spelling from the Hebrew standard today; attributing the Dictionary entries to their roots. If the word's root was not determined, it is attributed to another selected form – the “*neta*”.

**The form of the Historical Dictionary:  
The Root 'RB – Specimen Pamphlet  
and a thought about the future**

We have so far dealt with describing the work of the HDP, with its source books and its databases. But what will be the form of the Dictionary itself?

The editorial board tried to answer this question in 1982 and published – on the pages of the journal *Leshonenu* – a specimen pamphlet, containing the complete lexicographic treatment of one root of the Hebrew language – the root 'RB (*Leshonenu* 46.3-4.165-267). The root 'RB was chosen as a sample because of the many difficulties it poses to the lexicographer in determining its branches of meanings, whether because of homonymy or because of polysemy. As Ben-Hayyim, who edited the pamphlet, wanted to present in it the continuity of use of the words derived from the root 'RB, he included references not only from the Dictionary's databases – the ancient literature database and the modern literature database – but also from sources from historical layers of the language not yet processed in the HDP.

However, this pamphlet was written and printed when the personal computer and information networks such as the Internet were still in their infancy. The enormous development that has occurred in information technology in recent years requires the Project to adapt continuously, and it will naturally have an influence on the design of the Dictionary. Nevertheless, the infrastructure work that has been done, is being done and will be done on the Project, offers a solid base for the compilation of the Historical Dictionary of the Hebrew Language, whatever end form it may have.

## GLOSSARY

### Historical terms

**Amora** *pl.* **Amora'im** אמוראים a sage of the Talmudic period (3rd-5th centuries CE); hence **Amoraic**.

**barayta** (more commonly **baraita**) *pl.* **baraytot** ברייתא a Talmudic quotation from a Tannaitic source outside the Mishnah; in a wider sense: any piece of Tannaitic material not incorporated in the Mishnah.

**Barayta di-Melekhet ha-Mishkan** ברייתא דמלאכת המשכן (**barayta**) “on the Building of the Tabernacle”.

**Book of Ben Sira** (The) ספר בן סירא a book of the Apocrypha; composed by Shim'on ben Jeshua ben 'El'azar ben Sira (2nd century BCE); a.k.a. The Wisdom of Ben Sirach, Ecclesiasticus, etc.

**Gaon** *pl.* **Geonim** גאונים a post-Talmudic sage, mainly in Babylonia (6th-11th centuries CE); hence **Geonic**.

**Haskalah** השכלה the Jewish ‘Enlightenment’ movement in Europe (c. 1770s-1880s) that promoted the adoption of secular European culture.

**Megillat 'Ahima'az** מגילת אהימעץ The Scroll of 'Ahima'az; originally: ספר יוהסין **Sefer Yuhasin** (The Book of Genealogy); composed in Italy by 'Ahima'az ben Palti'el (11th century CE).

**Midrash** מדרש a legal and legendary commentary on the Bible often characterized by non-literal interpretation; also: an anthology of pieces of this; hence **Midrashic**.

**Mishnah** משנה the collection of oral law compiled by Rabbi Judah ha-Nasi (beginning of the 3rd century CE); also: a single paragraph of this; hence **Mishnaic**.

**Talmud** תלמוד a compilation of the Mishnah and its Amoraic commentary (**Gemara** גמרא); there are the Jerusalem (or Palestinian) Talmud (beginning of the 5th century CE) and the more authoritative Babylonian Talmud (end of the 5th century CE); hence **Talmudic**.

**Tanna** *pl.* **Tanna'im** תנאים a sage of the Mishnaic period (1st-2nd centuries CE); hence **Tannaitic**.

**Tosefta** תוספתא a compilation of baraytot, arranged according to the order of the Mishnah (end of the 3rd century CE).

### Linguistic terms

**millon** מילון a dictionary; derived from **milla** מילה a word; said to be the first new word introduced into Modern Hebrew by Eliezer Ben-Yehuda (1880).

**neta** נטע a stem.

**niqqud** ניקוד (diacritical) vowel points, vowel marks.

**shoresh** שורש a root; a morpheme that consists of a sequence of (usually) three consonants and carries the basic meaning of a word.

**teva** *pl.* **tevoth** תיבות a word token.

## A Glance at Porto Editora

Graciete Teixeira



Graciete Teixeira was born in Porto. Her studies of Roman Philology arose her interest in the Portuguese language and lexicography in particular, and in languages in general. She has broad experience in teaching at Strasbourg and Poitiers Universities, and in translation and terminology at the European Union, and a strong passion for dictionaries. Dr Teixeira is managing director of Porto Editora.  
gracietet@portoeditora.pt

Porto Editora is a publishing house based in Porto, Portugal, specializing in educational books, dictionaries, and off and online multimedia products.

It was founded in 1944 by Vasco Teixeira, my father, who managed to interest a group of teachers in his idea of creating high quality schoolbooks. Motivated by a commitment to promote high standards in education, the company was highly successful, and an independent organization developed from the visionary project. Over the years, Porto Editora established itself as a leading educational publisher, and in the process underwent periods of great change, which, in fact, mirror the major transitions within the Portuguese society.

The need for high quality printing at specific periods of the year soon resulted in buying a printing house, Bloco Gráfico, which, to this day, takes care of all production. From the beginning, a large amount of money was invested in advanced typesetters and printing machines, which allowed for high efficiency and regular production. In the late 1990s, a modern building was designed to fit the increasing printing turnover, with innovative technical capacity, and an automatic warehouse.

Our products are distributed by ourselves or by subsidiary companies, Arnado, in the central area of Portugal, and Fluminense, in the south. This policy allows us to be present all over the country.

Meanwhile, the implementation of new technologies has enabled us to step into multimedia products, providing the Portuguese market with educational and reference CD-ROMs and DVDs. Our presence on the Internet is both institutional and commercial, with a bookstore online and four sites addressed to students, parents and teachers. Since 2003, we offer a paid service – Infopedia.pt – with encyclopaedic articles, dictionaries and multimedia resources.

### Dictionaries at Porto Editora

The publication of the first edition of the *Dicionário Editora da Língua Portuguesa* (DELP) took place in 1952, as Vasco Teixeira soon realized dictionaries were an essential part of culture. That year really marked the beginning of a trend in dictionary making that has prevailed up to the present. New editions of Portuguese dictionaries, as well as bilingual dictionaries for English, French, Spanish,

Italian, German, Japanese, Modern Greek, Polish, Serbo-Croatian, Romanian, Dutch, Swedish and Latin were launched, providing the market with learning tools that have grown increasingly popular among students, teachers and the general public.

At first all the work on dictionaries was done on paper, and relied on the author's capacities. Authors were chosen by their reputation and language competence, and many titles appeared. It soon became clear that the process and quality should be controlled in-house, and, as a result, some of the staff became lexicographers. This was suitable for that time, as entries were hand-written, notes were kept in a system of card-based inventory, and texts had to be manually typeset. New editions were revised using the same hard time-consuming methods.



Vasco Teixeira, 1916-1987

When I joined the company, coming from a translation service where I had done a lot of work in terminology, I had a strong feeling that we need to control the content much more than the form – as was the practice until then – if we wanted to maintain our leading position in dictionaries and be able to face the rising competition. So, in the mid-1990s, a new dictionary division was set up, and a whole new team of lexicographers was recruited. Most of them went through training programs, while they developed their own skills as language researchers, as at that time there were no university courses on lexicography in Portugal. By then the dictionary program of the company had, as its prime concern, the

formation of in-house, full-time, skilled staff, and the development of editing software for dictionary making, using available computer technologies. It took a few years before the work done in the dictionary division became almost completely automated: dictionaries were converted into databases, information was gathered and tagged so as to permit data retrieval, and there was a huge investment in electronic devices for storing information and optimizing the productive process. We were very happy to produce our first dictionaries coming directly out of our database in 1998, and we took the opportunity to redesign the cover and change the format, while maintaining



Main building in Porto

the colour orange which immediately identifies Porto Editora's dictionaries in the Portuguese market.

The use of databases has allowed us to produce three complete CD-ROMs: Portuguese, English and French, including dictionaries of verbs, and to introduce other features. At the same time, we put DELP online for free, then English/Portuguese, Portuguese/English, French/Portuguese and Portuguese/French dictionaries. A Portuguese illustrated dictionary, dictionaries specially created for primary school students and substantial revisions of different dictionary ranges were undertaken, helped by the advance of new technologies.

Today, our policy is to produce the most updated dictionaries suitable for the target users. Thus, DELP has had the year of publication on the cover for the past two years, in order to stress its actuality.

#### **A brand new masterpiece**

For the 60th anniversary of our publishing house this year, we decided to publish a dictionary of one volume, including the main work done by Porto Editora on the Portuguese language database, and thus to create a new lexicographical reference book in Portugal.

Our *Grande Dicionário da Língua Portuguesa* appeared in May 2004. It presents the actual state-of-the-art of the Portuguese language not only in

Portugal, but also in Portuguese-speaking countries, including Angola, Brazil, Cape Verde, Guinea-Bissau, Mozambique, and São Tomé and Príncipe. This important geographical coverage includes both a large degree of common language as well as special domains.

It was a great challenge to coordinate the collaboration of many specialists who updated our databases and introduced new concepts in different areas, from mathematics to medicine, music or biology. This specialized revision was done on paper, and our lexicographers were in charge of dealing with the experts, adapting their explanations to a dictionary definition and finally introducing the new version in the database. Besides, our team dealt with common language entries and some areas of their knowledge, such as linguistics or literature.

With its prestigious physical features, the *Grande Dicionário* stands as a symbol of the company's long experience in dictionary making and of its high standard of quality.

#### **PE Group details**

- 7 companies
- 600 persons
- 44% market share in schoolbooks
- 70% market share in dictionaries

#### **Dictionaries in numbers**

- 15 languages
- 102 titles
- 9 ranges



#### **Grande Dicionário da Língua Portuguesa**

- 220 cm x 302 cm
- 1624 pages
- Cloth hardcover with engraved design and gold lettering
- Protection box
- ISBN 972-0-05000-4

## **Lexicography organization founded in Taiwan**

Taiwan's first lexicography organization, the Dictionary and Corpus Research Center (DCRC), was founded on 7 March 2004 in Taipei. This formation was initiated at the plenary meeting of the Taiwanese Association for Translation and Interpretation, held on 20 December 2003, and the DCRC will function as an interest group within the Association. Founding members include lexicographers, corpus linguists and publishers. The DCRC aims to become an active organization of lexicography theorists and practitioners, and to serve as a confluence of dictionary and corpus resources. Its short-term goals include making the general readers and specialized researchers in Taiwan aware of the different kinds of dictionaries and corpora that are available, advising the public on dictionary purchases and use, and evaluating dictionaries and related reference books that are on the market. Its long-term goals include publishing newsletters, books and academic journals, and hosting conferences. The DCRC intends to hold quarterly meetings with keynote lectures. The current coordinator is Dr. Hugo T. Y. Tseng, of the English Department of Soochow University, Taipei. The contact person is Ms. Meihua Sun, [meihua@bookman.com.tw](mailto:meihua@bookman.com.tw).

# Kernerman Dictionary Research Grants

## The Creation of an Innovative Kiswahili-English Online Dictionary

Gilles-Maurice de Schryver and Sarah Hillewaert

AFRILEX has allocated a two-year grant for 2003 and 2004 to Gilles-Maurice de Schryver (principal investigator) and Sarah Hillewaert (co-investigator) for the project 'The Creation of an Innovative Kiswahili-English Online Dictionary', the outline of which appears herewith.

The Kernerman Dictionary Research Grants are allocated by independent Assessment Committees administered by AFRILEX, ASIALEX and EURALEX.

Funds are still available for 2004 from ASIALEX and EURALEX for research projects in the following fields:

- The study of the dictionary-using behavior of language learners at the elementary-to-high-school levels, and the design of these dictionaries.
- Specialized corpora for foreign-language learners.
- The function of lexicography in vocabulary acquisition.
- Trilingual and multilingual lexicography.
- Lexicography in the service of language preservation.

More details are available in *Kernerman Dictionary News*, Number 10, July 2002 (<http://kdictionaries/newsletter/kdn10-11.html>). Applications should be made to the relevant committees directly.

Kiswahili (Swahili) is one of Africa's major languages, spoken throughout East Africa as a lingua franca by tens of millions of people. Despite its official status, substantial number of speakers and relatively long dictionary tradition, the state of lexicographical research as well as the availability of modern and up-to-date dictionaries for Kiswahili is far from satisfactory. The numerous Kiswahili monolingual and bilingual dictionaries compiled for over a century are largely based on Western compilation principles, and until now the most commonly used dictionaries remain rooted in lexica originally derived by missionaries.

Kiswahili is an *agglutinating* language, meaning that morphemes are juxtaposed to form linguistic words. In all current dictionaries, 'orthographic words' are decomposed into their formatives, with only the latter being lemmatised. As a result, not all native speakers of Kiswahili can look up 'words' in their own language – as this implies being able to cut off prefixes and suffixes – and even trained scholars often need more than one look-up round before they hit on what they are looking for (since vocal changes between formatives are not always predictable).

This research project attempts to deal with all these problems simultaneously. The aim is to create the first corpus-based Kiswahili dictionary that is also intuitive in nature, and to research the feasibility of this approach in real time. Instead of lemmatising stems as in traditional dictionaries, the idea is to lemmatise full orthographic words (in addition to stems), and to provide full translations for these strings. In order to sensibly limit the number of items one can physically treat, the items will be selected from a frequency list derived from a large corpus. Concordance lines will be called up for each frequent orthographic word, and the various translations will be recorded in order of frequency. A user will thus be able to look up words directly, as they are spoken or written, and the translations will be arranged from most likely to least likely. An English search index will additionally enable searches in the reverse direction. Since, obviously, such an

approach will require much more 'space' than in a traditional stem-based dictionary, the dictionary will be developed and made available in an electronic environment right from the start, primarily on the Internet, where it is also possible to keep a log of all searches. Analyzing these log files will enable further research on whether or not this hybrid approach is feasible and to amend the approach if need be.

Given the intuitive lemmatisation approach, native speakers and learners at the elementary and intermediate levels will for the first time be able to effectively look up words, and find meanings of 'real' words, which should help to develop a dictionary culture. Furthermore, the log files will be utilised to full potential by tracking each individual dictionary-use behaviour, including vocabulary retention. For the first time, truly unobtrusive data will be collected and true look-up behaviour in an electronic environment will be recorded. Finally, this project will also ensure that Kiswahili, an increasingly popular language on the Internet, is also kept alive in a modern online reference work based on sound lexicographical principles.

Mr de Schryver has participated in building a 13-million-word Kiswahili corpus, which will be queried in this project. He is the main compiler of a recent online dictionary for Sesotho sa Leboa (<http://africanlanguages.com/sdp>), from which valuable information will be drawn. Ms Hillewaert will be responsible for the compilation and editing of the Kiswahili-English translations. The two have worked together on various Kiswahili projects over the past few years, and produced a lexicon for Sheng (a language that is largely based on Kiswahili and spoken by the youth in Nairobi).

It is expected that a first lexicon, consisting of translations for the top 1500 orthographic words, could be uploaded following three months of work, and batches of 1500 items could be added to the online dictionary every three months. The project is planned to run for at least three years (and thus reach 18,000 items) before the great majority of the searches will be successful.