

Are space-saving strategies relevant in electronic dictionaries?

Marie-Hélène Corréard

5 place Hubert Dubedout - 38000 Grenoble - France

correard@worldnet.fr

Abstract

A major challenge for lexicographers compiling paper dictionaries has always been one of available space. Different strategies have been devised to pack as much information as possible into traditional dictionaries. This paper addresses some of these strategies and their usefulness in electronic dictionaries.

In the first part of this article, the more straightforward methods used by dictionary makers are examined. These range from using thinner paper, choosing smaller fonts, to replacing headwords by swung dashes or listing collocates with slashes, etc.

Part two examines in detail the special treatment of lexical sets. The technique that consists in giving all the information about one item and cross-referring to a 'model entry' has been used for a long time for items like months and days of the week. In the Oxford-Hachette Dictionary (1994) this approach was taken further and developed into a feature.

Finally, in the third part of this paper, the need for such an approach in electronic dictionaries is discussed. Even when space is not an issue, some of the devices remain useful. They need to be enhanced and adapted to the electronic medium in order to give users the possibility to acquire knowledge (e.g. extracting lexical sets) in an organised and effortless fashion.

Introduction

A major challenge for lexicographers compiling paper dictionaries has always been one of available space and controlling length [Landau 1984]. The switch to the electronic medium has made this problem less acute. However, it may be worth considering what space-saving strategies have done for dictionary users, apart from enabling lexicographers to put in more information. So the first question to ask is what exactly are these space-saving strategies?

Traditional space-saving strategies

The strategies used to 'save space' in dictionaries, range from purely superficial methods to deeper ones where action is taken on the text itself.

The most superficial and crudest methods to increase the amount of space available in a dictionary, include making the book larger by having a bigger page size, or thicker by having more pages, or both by having more pages of a bigger size. However, there are traditions, presumably based on production limitations and customers' surveys, whereby 'desk dictionaries' rarely exceed 30 cm in height and 7-8 cm in thickness. As one can imagine, the limits are reached fairly quickly. If it is not possible to make a dictionary larger and thicker, it is possible sometimes to use thinner paper. Here again, there are some limitations: the paper must be thick enough so that it is possible to read one page at a time without the interference of text on the verso. If all else fails, the dictionary can be published as two (or more) volumes, but then it becomes a 'product' which is totally different from the one

planned initially. Having several volumes instead of one may cause objections within the publishing company.

When all has been tried with the physical object, it will be necessary to work on typography and layout. A clever designer will find ways to help fit more characters on the page. The most obvious one is to select smaller and/or narrower fonts. When the text becomes unreadable, one has to stop or supply a magnifying glass, as was done with a reduced version of the Oxford English Dictionary. Reducing the margins could be another option; this option can be combined with the previous one. These are tricks and have limitations. In the age of "user-friendliness" and "accessibility", weight and readability are criteria a dictionary buyer will certainly take into account while selecting an item in a bookshop.

The next levels of intervention will be with lexicographers. They will devise methods to cram more characters in by shortening the text. In the first instance they will try not to alter the contents of the dictionary significantly. Therefore the obvious first victim is metalanguage. It is used across the dictionary and it is already coded information. So modifying and tightening this coded language is not going to change the actual contents of the dictionary. All those familiar with dictionaries are used to abbreviations such as *n.* or *n* without the full stop in more radical cases for *noun*, *tr.* or *tr* for *transitive* and so on. A quick calculation based on the French section of the Oxford-Hachette French Dictionary [henceforth OHFD 1994] shows that using abbreviated parts of speech, for nouns only, saved around 20 pages; the simple dropping of full stops (as in *nf.*) saved over two pages. Most other elements of metalanguage can be abbreviated: semantic field labels, grammatical information about translations in the case of bilingual dictionaries, cross-references, GB/US variations, etc. In some cases, symbols are used to save space, notably for register labels, trademarks, archaic or old-fashioned words and cross-references [Svensén 1993].

Still more space can be saved by working on the dictionary text itself. Firstly, without altering it, just using abbreviations to replace the headword (swung dash or sometimes the initial letter of the headword) in examples. Equally, using abbreviations for words repeated in examples : e.g. *quelqu'un* is replaced by *qn* and *quelque chose* by *qch*. An explanation about the conventions chosen will also result in space saving: e.g. regular inflections are not shown, but formation rules explained in the *How to use the dictionary* section : e.g. "*Most nouns form their plurals by adding -s [...]. Nouns which end in -s, -ss, -ch, -x, and -z make their plural by adding -es.*" in Cambridge Learner's Dictionary [CLD 2001]. Alternatively, regular inflections are shown in tabular form in appendices and cross-references indicate where to find the information, e.g. *chanter* /... / [1] [OHD 1994].

Secondly, dictionary text can be condensed by means of various techniques:

- Nesting derivatives within an entry (e.g. including *dramatically* as a run-on item under the headword *dramatic*), will considerably shorten the text. It is not customary in bilingual dictionaries.
- Using slashes: in examples *changer qch/qn en and in translations to change sth/sb into* , in translations to supply information about the target language *audacieux/-ieuse* [OHD].
- Giving examples in a 'canonical' form e.g. *to be able to do something* in Collins Easy Learning French Dictionary [CELFD 1998], *to be able to do sth* [CLD], where *to do*

something or *to do sth* could be replaced by any verb; they can also be given in 'economical' yet productive form: *to be able to [walk|travel|sleep...]* in Oxford Starter French Dictionary [OSFD 1997].

- Giving collocates in a formal manner, as lists, rather than fully fleshed examples, e.g. under *émanciper*: to emancipate [**people**]; to liberate [**country**] [OHFD].

When selecting examples, every character counts and lexicographers are known to have replaced first names or surnames in real corpus examples to gain a few characters. Why are *Peter*, *Paul* and *Mary* so common in dictionaries? Because they take less space than *Richard*, *Geoffrey* or *Margaret*!

Another method commonly used in learners' and bilingual dictionaries was mentioned by Berkov [1988]: *All languages have series of homogeneous lexemes that can be used in homogeneous constructions, e.g. names of the days of the week, cf. on Friday/Tuesday..., on Fridays/Tuesdays..., last Friday/Tuesday..., etc. [...] Much space is saved when constructions are given only once, e.g. under the headword which comes alphabetically first, e.g. under the headword Friday [...].*

Lexical notes in the Oxford-Hachette French Dictionary

The technique that consists in giving all the information about one item and cross-referring to a 'model entry' has long been used in monolingual and bilingual dictionaries for the most obvious closed sets: days of the week, months, etc. In the OHFD this approach was taken further: more lexical sets were identified and given extensive treatment. The resulting notes were presented in boxes spread over the dictionary; all the headwords related to a note were given cross-references to the page where the note is located. There are 45 notes and 3,325 cross-references pointing to lexical notes in the English-French section, and 40 in the French-English with 3,325 cross-references. These notes were written by a team of external contributors (Richard Wakely and Henri Béjoint) in close collaboration with the Chief Editors under B.T.S. Atkins' guidance. In the introduction to the first edition of the dictionary, the boxes containing the information traditionally given under model entries, are called *Lexical Notes*; they are described as follows: *These are notes which can be accessed from certain entries. They give the user facts about certain types of words that behave alike [...] and provide ways of discussing topics [...].*

The sets were selected according to several criteria: there are enough members in a set to make it worth treating; enough members of a set behave in the same way; examples involving items of the set in real language and presented in the dictionary must be useful. There are several sorts of notes. Some notes include words which are syntactically interchangeable in examples, e.g. *The signs of the Zodiac, Rivers, British regions and counties, US states*. The note on *The signs of the Zodiac* (see appendix 1) is a good illustration of a closed lexical set. It lists all the items belonging to the set, then it gives what is seen as the most common and useful examples, showing implicitly that all the members of the set work in the same way in English (the note is for English speakers). On the other hand, the note does mention explicitly that all the signs work in the same way in French. Cross-references to the note can be found under 18 headwords: all the entries for the signs of the

Zodiac (12) as well as *Aquarian, Libran, Piscean, Sagittarian, Taurean* and *sign*. As the note is located near the entry for *zodiac*, there is no cross-reference.

Other notes are focussed on topics and regroup all the relevant words. The note on *Age* (see appendix 2), for instance, explains how to express age by asking the question *How old are you?* It ensures that the reader knows about transfer problems between French and English (*Note that where English says to be X years old French says avoir X ans (to have X years), Note the use of de after âgé and à l'âge: a woman aged thirty = une femme âgée de trente ans*).

Cross-references to this lexical note will be found under entries for numbers up to a hundred and to the entries for *age, old* and *year*.

Both kinds of notes will be useful for natives speakers of the source language who want to 'encode' in the foreign language. There is also a third category of notes which are a mixture of both. The note for *Shops, trades and professions* (given in appendix 3) is a good example. It is clear that most of the professions which have a cross-reference to the note would not have had examples in the paper version of the dictionary, for lack of space. With the help of the note, users can generate correct utterings such as *Paul is a dentist/an aviculturist/baggage handler Paul est dentiste/aviculteur/bagagiste*, for any profession.

A quick calculation for these three notes enables us to see how efficient they are as a space-saving strategy: in order to give the information contained in the note at all the entries cross-referring to it, this information would have had to be repeated at least 12 times for *The signs of the Zodiac*, 42 times for *Age* and 1,114 times for *Shops, trades and professions*. Is saving space the only function of these boxes? Clearly not.

Lexical notes are features in electronic dictionaries

When space ceases to be an issue, is there a need for such an approach in electronic dictionaries? Obviously, with more space, all the general material given in a note could be given within the entry. It would make looking up a fairly simple entry a more complicated task simply because of the amount of information and it would obliterate the notion of lexical set. In the same way as using a swung dash to replace the headword ensures that a dictionary user does not forget what the headword is while reading examples, having some sort of mention of lexical sets is useful for language learners. It emphasizes the systematicity of language and the issue of interrelationship between the lexicon and general syntactic rules [Aitchison 1987]. It helps the readers realise that if they know how to say *Paul is a dentist* in French, they will be able to extrapolate to all other professions, as long as the hint is provided. By comparison with its original paper counterpart the electronic version of the OHFD already supplies more facilities; it is possible to extract a list of all the headwords cross-referring to a given note. However, some functionalities could profitably be added. When a word is part of a lexical set, it is marked as such and access to the note itself is gained by clicking on the cross-reference symbol. It would also be useful to have access to the full list of all the members of a set and this could be achieved without too much cost for the publisher, as the information exists. Of course, one could dream of having access to Mel'cuk's lexical functions but this is not built into the data and would involve more editorial work. Also, on demand, it should be possible to generate all the examples using the

headword which is part of a set. To go even further in the direction of the dictionary as a resource for language teaching/learning, it would be useful to add exercises to check that all the information given in a note has been assimilated.

Furthermore, some sets which were eliminated for the paper version of the dictionary as not productive enough but which were included in a book called *French Usage* could easily be added to the electronic version of the dictionary and possibly new ones developed. Such simple additions would transform a traditional general bilingual dictionary into a more powerful one and also give users an integrated tool for learning a language or, as the authors of the *Lexical Notes* themselves write [Wakely 2002]:

Thus we may be witnessing an evolution of the bilingual dictionary towards an all-purpose self-teaching book including all sorts of information that might be needed by users, and going far beyond the areas of meaning and basic syntax that were the domains of the traditional bilingual dictionary.

Conclusion

Some of the strategies used in the OHFD were more than simple devices to save space. They ensure comprehensiveness and systematicity of description for similar entries [Swanepoel 1994]. They have a didactic purpose and still have a place in electronic dictionaries, making them a complete resource for learning, understanding and producing the foreign language.

Appendix 1

The signs of the Zodiac

Aries	=	le Bélier	[belje]	21 mars–20 avril
Taurus	=	le Taureau	[tOlo]	21 avril–20 mai
Gemini	=	les Gémeaux	[Zemo]	21 mai–21 juin
Cancer	=	le Cancer	[k~AsE1]	22 juin–22 juillet
Leo	=	le Lion	[lj~O]	23 juillet–22 août
Virgo	=	la Vierge	[vjE1Z]	23 août–22 septembre
Libra	=	la Balance	[ba~As]	23 septembre–23 octobre
Scorpio	=	le Scorpion	[skO1pj~O]	24 octobre–21 novembre
Sagittarius	=	le Sagittaire	[səZitE1]	22 novembre–21 décembre
Capricorn	=	le Capricorne	[kap1ikO1n]	22 décembre–19 janvier
Aquarius	=	le Verseau	[vE1so]	20 janvier–18 février
Pisces	=	les Poissons	[pwas~O]	19 février–20 mars

I'm Leo	=	je suis Lion
I'm Gemini	=	je suis Gémeaux
born in Leo <i>or</i> under the sign of Leo	=	né sous le signe du Lion
born in Gemini	=	né sous le signe des Gémeaux
Leos/Arians are very generous	=	les Lions/les Béliers sont très généreux
what's the horoscope for Leos?	=	que dit l'horoscope pour les Lions?
the sun is in Leo	=	le soleil est au Lion

All the signs work in the same way in French.

Appendix 2

Age

Note that where English says to be X years old French says avoir X ans (to have X years)

How old?

how old are you? = quel âge as-tu?
 what age is she? = quel âge a-t-elle?

The word *ans* (years) is never dropped:

he is forty years old *or* = il a quarante ans
 he is forty *or*
 he is forty years of age
 she's eighty = elle a quatre-vingts ans
 the house is a hundred = la maison a cent ans
 years old
 a man of fifty = un homme de cinquante ans
 a child of eight and a half = un enfant de huit ans et demi
 I feel sixteen = j'ai l'impression d'avoir seize ans
 he looks sixteen = on lui donnerait seize ans

Note the use of *de* after *âgé* and *à l'âge*:

a woman aged thirty = une femme âgée de trente ans
 at the age of forty = à l'âge de quarante ans
 Mrs Smith, aged forty *or* = Mme Smith, âgée de
 Mrs Smith (40) = quarante ans

Do not confuse *que* and *de* used with *plus* and *moins*:

I'm older than you = je suis plus âgé que toi
 she's younger than him = elle est plus jeune que lui
 Anne's two years younger = Anne a deux ans de moins
 Margot's older than Suzanne = Margot a cinq ans
 by five years de plus que Suzanne
 Robert's younger than Thomas = Robert a six ans de
 by six years moins que Thomas

X-year-old

a forty-year-old = quelqu'un de quarante ans
 a sixty-year-old woman = une femme de soixante ans
 an eighty-year-old pensioner = un retraité de quatre-vingts ans
 they've got an eight-year-old = ils ont un enfant de
 and a five-year-old huit ans et un autre de cinq ans

Approximate ages

Note the various ways of saying these in French:

he is about fifty = il a environ cinquante ans *or* il a une cinquantaine d'années *or* (*less formally*) il a dans les cinquante ans

(Other round numbers in *-aine* used to express age are *dizaine* (10), *vingtaine* (20), *trentaine* (30), *quarantaine* (40), *soixantaine* (60) and *centaine* (100).)

she's just over sixty = elle vient d'avoir soixante ans
 she's just under seventy = elle aura bientôt soixante-dix ans
 she's in her sixties = elle a entre soixante et soixante-dix ans
 she's in her early sixties = elle a entre soixante et soixante-cinq ans
 she's in her late sixties = elle va avoir soixante-dix ans *or* (*less formally*) elle va sur ses
 soixante-dix ans
 she must be seventy = elle doit avoir soixante-dix ans

he's in his mid-forties	=	il a environ quarante-cinq ans <i>or (less formally)</i> il a dans les quarante-cinq ans
he's just ten	=	il a tout juste dix ans
he's barely twelve	=	il a à peine douze ans
games for the under twelves	=	jeux pour les moins de douze ans

Appendix 3

Shops, trades and professions

Shops

In English you can say *at the baker's or at the baker's shop*; in French the construction with *chez* (at the house or premises of ...) is common but you can also use the name of the particular shop:

at the baker's	=	chez le boulanger <i>or</i> à la boulangerie
I'm going to the grocer's	=	je vais chez l'épicier <i>or</i> à l'épicerie
I bought it at the fishmonger's	=	je l'ai acheté chez le poissonnier <i>or</i> à la poissonnerie
go to the chemist's	=	va à la pharmacie <i>or</i> chez le pharmacien
at <i>or</i> to the hairdresser's	=	chez le coiffeur/la coiffeuse
to work in a butcher's	=	travailler dans une boucherie

Chez is also used with the names of professions:

at <i>or</i> to the doctor's	=	chez le médecin
at <i>or</i> to the lawyer's	=	chez le notaire
at <i>or</i> to the dentist's	=	chez le dentiste

Note that there are specific names for the place of work of some professions:

the lawyer's office	=	l'étude f du notaire
the doctor's surgery (GB)	=	le cabinet du médecin
<i>or</i> office (US)		

Cabinet is also used for architects and dentists. If in doubt, check in the dictionary.

People

Talking of someone's profession, we could say *he is a dentist*. In French this would be either *il est dentiste* or *c'est un dentiste*. Only when the sentence begins with *c'est* can the indefinite article (*un* or *une*) be used.

Paul is a dentist	=	Paul est dentiste
she is a dentist	=	elle est dentiste <i>or</i> c'est une dentiste
she's a geography teacher	=	elle est professeur de géographie <i>or</i> c'est un professeur de géographie

With adjectives, only the *c'est* construction is possible:

she is a good dentist	=	c'est une bonne dentiste
-----------------------	---	--------------------------

In the plural, if the construction begins with *ce sont* then you need to use *des* (*or de* before an adjective):

they are mechanics	=	ils sont mécaniciens <i>or</i> ce sont des mécaniciens
they are good mechanics	=	ce sont de bons mécaniciens

Trades and professions	
what does he do?	= qu'est-ce qu'il fait?
what's your job?	= qu'est-ce que vous faites dans la vie?
I'm a teacher	= je suis professeur
to work as a dentist	= travailler comme dentiste
to work for an electrician	= travailler pour un électricien
to be paid as a mechanic	= être payé comme mécanicien
he wants to be a baker	= il veut devenir boulanger

References

A. Dictionaries

- [CLD 2001] *Cambridge Learner's Dictionary*, Walter, E. (ed), Cambridge University Press, Cambridge, UK
- [CELF 1998] *Collins Easy Learning French Dictionary*, Kopleck, H. (ed), HarperCollins, Glasgow, UK
- [OHD 1994] *Oxford-Hachette French Dictionary*, Corréard, M.-H. & Grundy, V. (eds), Oxford University Press, Oxford, UK and Hachette, Paris, France.
- [OSFD 1997] *Oxford Starter French Dictionary*, Corréard, M.-H. & O'Neill, M. (eds), Oxford University Press, Oxford, UK.

B. Other references

- [Aitchison 1987] Aitchison, J., 1987. *Words in the Mind - An Introduction to the Mental Lexicon* -, Basil Blackwell.
- [Berkov 1988] Berkov, P.V., 1988. *A Modern Bilingual Dictionary - Results and Prospects*, *BudaLEX '88 Proceedings*, Budapest, Hungary.
- [Landau 1984] Landau, S.I., 1984. *Dictionaries - The Art and Craft of Lexicography*, Cambridge University Press, Cambridge, UK.
- [Swanepoel 1994] Swanepoel, P., 1994 Problems, Theories and Methodologies in Current Lexicographic Semantic Research, in W. Martin et al. (eds), (11-26) *Euralex '94 Proceedings*, Amsterdam, The Netherlands.
- [Svensén 1993] Svensén, B., 1993. *Practical Lexicography*, Oxford University Press, Oxford, UK.
- [Wakely 1996] Wakely, R. & Béjoint, H. 1996. *French Usage*, Oxford University Press, Oxford, UK.
- [Wakely 2002] Wakely, R. & Béjoint, H. 2002. *Word Groups in Bilingual Dictionaries: OHFD and After* (Forthcoming).