

Space restrictions in paper and electronic dictionaries and their implications for the design of production dictionaries

Robert Lew
Adam Mickiewicz University
rlew@amu.edu.pl

Abstract:

One important consideration in dictionary-making has been that of space. To conserve space in paper dictionaries, a number of principles, strategies and conventions have been employed. With the gradual transition of the dictionary to the electronic medium, some of these strategies and conventions have lost, or changed, their significance.

For one thing, it is no longer sufficient to speak of *dictionary space* alone, as was customary with paper dictionaries. One should now distinguish between at least two types of space: *storage space* and *presentation space*. In fact, this distinction is also valid for paper dictionaries.

By storage space I mean the capacity to hold the total content of the dictionary. Storage space restrictions are no longer a major concern in electronic dictionaries except for the most space-consuming content, such as high-resolution video. Presentation space refers to the display of lexicographic information, and here the restrictions are very real in the case of electronic dictionaries.

The (potentially) dynamic character of lexicographic presentation in electronic dictionaries redefines classical metalexigraphic notions, such as *microstructure* and *access structure*, or *entry element* and *cross-reference*.

The move of dictionaries to the electronic medium has also opened up new possibilities for dictionaries designed specifically to aid production in the second/foreign language. I offer some suggestions on how these new proposals could be improved to take advantage of the electronic medium.

1. Space in dictionaries: background

Space has traditionally been a concept central to dictionaries, and, consequently, an essential consideration in dictionary making (Landau 2001). There is a widespread underlying assumption that space in dictionaries is a highly valuable commodity. To describe even the core lexicon of a natural language requires a substantial volume of data. Fitting it into a printed book is usually a challenge. Splitting a dictionary into multiple volumes is an option only for some dictionary types. Dictionaries need to be affordable to the users, so they cannot be excessively large. Many are intended to be portable, so they cannot be too bulky. Paper dictionaries are often reproduced in many copies and thus even small savings get multiplied manifold. It is an overarching principle of lexicography that dictionary space needs to be conserved. As a consequence, a host of conventional principles, strategies, and devices have been employed in lexicographic practice of the past and present, contributing to the peculiar properties of the canonical dictionary text (cf. the concept of textual condensation, Hausmann & Wiegand 1989; Wiegand 1996).

1.1. From principle to device

We have referred above to *principles*, *strategies*, and *devices*. To illustrate the difference between the three in this context: one rather obvious *principle* would be to *avoid the duplication of information*, and from this principle follows the *strategy* of cross-referencing, yielding a number of specific cross-referencing *devices*.

1.2. Some space-saving conventions

Some space-saving devices may become conventionalized in lexicography: this means they they become popular across a range of lexicographic projects. Thus, they become lexicographic conventions. Some examples of such conventions would be the following:

- use of abbreviations is conventionalized in many lexicographic traditions; in English-language lexicography, for example, part of speech information has often been presented through abbreviations (n, v, adj); likewise, parts of definitions representing placeholders for subjects and complements have often been abbreviated (sb, sth)
- concise defining styles have traditionally been used for reasons of space in monolingual dictionaries; pocket monolingual dictionaries favour defining by synonym if possible; synonyms, being intralingual equivalents, are naturally shorter than phrasal or clausal definitions
- niching and nesting can be employed, resulting in run-on entries (see Gouws 2003 on the distinction between niching and nesting)
- restricted treatment is often given to derivatives, such as when no definition is provided at all
- cross-referencing to avoid duplicating information that is already available in another place; this could be in another entry, but a cross-reference may also be entry-internal, as in Figure 1 below

congruence

SYLLABICATION: con·gru·ence

PRONUNCIATION:  kɒŋ'gru-əns, kən-gru-'

NOUN: **1a.** Agreement, harmony, conformity, or correspondence. **b.** An instance **of this** “*What an extraordinary congruence of genius and era*” (Rita Rack).
2. Mathematics **a.** The state of being congruent. **b.** A statement that two quantities are congruent.

Figure 1: An example of entry-internal cross-reference from the online American Heritage Dictionary of the English Language¹. Subsense 1b cross-refers to subsense 1a.

2. Electronic dictionaries and dictionary space

There is no doubt that lexicography is now undergoing a steady transition to the electronic medium. One reason that dictionaries and other reference works are taking the lead in the electronic revolution is that they tend to be rather more voluminous than e.g. fiction, and so a transition to the electronic medium can save more paper (and buyers' money). Another, functional, reason is related to access: reference works, unlike works of fiction, are not typically meant to be read in linear order, and there is a potential for greater flexibility of access in the electronic format (that this potential has not always been utilized, especially in the early products, is still another matter).

The enthusiasm with which the electronic revolution is embraced invites comments on the superiority of the new medium; some of them fully justified, others only partially true. A frequently voiced sentiment is that space in electronic dictionaries is unrestricted, and, consequently, that space-saving becomes a non-issue.

Corréard (2002) argues that space-saving is still relevant in the context of electronic dictionaries. In her paper, though, she only tackles a rather narrow selection of details. It should be clear that in view of the new medium, the notion of space deserves a little more attention.

2.1. Storage space

On careful inspection, it appears that the notion of *dictionary space* is not specific enough as a technical term, because it is ambiguous. The suggestion that dictionary space is unrestricted is actually largely correct, but only when space is understood as the capacity to hold the total content of the dictionary – this sense of *dictionary space* could provisionally be called *storage space*. There is at least one more important sense of *dictionary space* which I will here call *presentation space*; but let me first deal with storage space.²

¹ At the time of original writing: <http://www.bartleby.com/61/> ; at publication time available at <http://dictionary.reference.com/browse/congruence>

² Perhaps a third category of space could be introduced (I am grateful to Michael Rundell for this suggestion), that of *perceptual space* (of the user), reflecting the dictionary user's capacity to perceive and process lexicographic data. Unlike storage space and presentation space, perceptual space would not be a property of the dictionary, but rather the user; it might,

2.1.1. Storage space in paper dictionaries

In traditional paper dictionaries, storage space is relatively restricted, and it is determined by the number of volumes, format, weight, layout, font density and font size of the finished book. Non-textual content, such as pictorial illustrations, takes up more space on paper than text, and some content, such as sound, video or animation cannot be stored at all.

2.1.2. Storage space in electronic dictionaries

Storage space in electronic dictionaries is relatively unrestricted, thanks to the modern high-capacity storage media, as well as content-sharing over high-speed networks. At present, only the most space-demanding content is restricted, such as high-resolution video (although there may still be real storage concerns in handheld devices). Clearly, it is the storage aspect of space that is meant in claims of space restrictions being irrelevant in electronic dictionaries. But here we come to another, just as important aspect of dictionary space.

2.2. Presentation space

In contrast to *storage space*, *presentation space* refers to how much can be presented (displayed, visualized) at a given time to the dictionary user.

2.2.1. Presentation space in paper dictionaries

In paper dictionaries, presentation space typically comprises two facing pages of an open book. Paper presentation space is static: that is, the content and appearance of a paper page does not change in time. Also, in paper dictionaries, storage space is a simple multiple of presentation space; the multiplication factor is the number of paper sheets, or half the number of pages of a given dictionary, assuming that two facing pages can normally be viewed at the same time. Thus, in paper dictionaries there is a simple relation between presentation and storage space, which is perhaps why not much has been made of the distinction between the two types of space, even though it is also perfectly valid for paper dictionaries.

2.2.2. Presentation space in electronic dictionaries

In electronic dictionaries, in contrast to the paper medium, some kind of visual display device is used for displaying the content of dictionaries, the most common at present being an LCD screen of various size (the more traditional CRT display is currently on the way out and will likely be a museum piece before soon). The absolute viewing area of today's typical PC screen is roughly comparable to two pages of a large-format paper dictionary, but the resolution of a standard screen is still a few times lower than that of typical print, and so such a screen is capable of carrying accordingly less information in the visual channel. For handheld devices, presentation space is of course still much more limited due to practical restrictions on physical dimensions and energy consumption.

However, the visual channel may be complemented by the audio channel: it is quite common today for dictionaries to offer audio recordings of headwords. Some products can read the definitions aloud.

Advances in electronic paper design (Graham-Rowe 2007) raise hopes for the resolution of commonly available electronic displays to equal that of traditional paper. Unlike traditional paper though, electronic display may be *dynamic*, and this important feature can be exploited to compensate for the lower momentary information content. The (potentially)

however, be relevant for dictionary design because the interaction between the dictionary and the user is likely to be affected by the user's perceptual capacity.

dynamic presentation in electronic dictionaries redefines classical lexicographic notions. As an example, in the following section let us consider *microstructure* and *access structure*.

3. Traditional metalexigraphic concepts in electronic format dictionaries

3.1. Microstructure versus access structure

Microstructure and access structure are two of the more important structural-theoretical

The screenshot shows the Longman Dictionary of Contemporary English ONLINE interface. At the top, there is a search bar with the word 'disaster' entered and an 'OK' button. Below the search bar, the word 'disaster' is displayed as a noun with its phonetic transcription and grammatical information. A definition is provided, and a 'COLLOCATIONS' section lists various phrases like 'air disaster', 'nuclear disaster', and 'natural disaster'. A 'damage' entry is highlighted in blue, indicating a cross-reference. A separate window titled 'LDOCE Online' is overlaid on the right, showing the 'damage' entry in detail, including its phonetic transcription, definition, and a 'COLLOCATIONS' section with phrases like 'do/cause damage', 'serious/severe/extensive damage', and 'irreparable/irreversible/permanent damage'.

offer, and so modern electronic dictionaries often help save both time and effort with regard to cross-referencing behaviour. In order to follow a cross-reference in a technologically enhanced lexicographic environment, all the user may need to do is click the mouse, and is instantly taken to the target location. In contrast to traditional paper *non-immediate* cross-references, electronic dictionaries may offer *immediate* cross-reference. Further, following a cross-reference no longer requires losing sight of the original context of the article: devices such as fanouts, callouts or popups can utilise some of the display area to give extended treatment targeted by the cross-reference, while still retaining most of the original article on screen; an example of this is shown in Figure 2.

Figure 2: An example of immediate cross-reference. Longman Dictionary of Contemporary English 4 online entry¹ for *disaster*. The target entry *damage* pops up on the screen but the source entry is still retained and partially visible in the background window.

It is also possible in electronic dictionaries for (some) cross-references to be activated on mere mouse hover. In those cases, users do not even have to resort to the proverbial point-and-click duo: they simply point at the relevant area of online text.

4. Dictionaries for production

At present, it seems that most bilingual dictionaries are best fit for receptive activities, mostly reading comprehension and L2→L1 translation. Only a small minority are designed as production dictionaries, a type also referred to as *active dictionaries* (though there are authors who insist on making a finer distinction between active and production dictionaries).

One may wonder why production dictionaries are so few in number, even though calls to make such dictionaries have been heard from time to time. The immediately obvious reason is that production dictionaries require more work of lexicographers, because such dictionaries need to be richer in content than dictionaries streamlined for reception. However, I believe that there is another important reason for the scarcity of production dictionaries: the very concept conflicts rather acutely with one of the fundamental space-saving principles of *paper* lexicography which follows from the limitation of storage space: that of avoiding redundancy.

Now, with the electronic medium rapidly becoming mainstream, the factors in the production dictionaries equation are changing, opening up new possibilities for dictionaries designed specifically to aid production in the second/foreign language. While lexicographers still have to produce (and get paid for) the richer content, storage space is no longer an issue, and so redundancy is no longer a lexicographic villain.

4.1. Redundancy: an example

To illustrate the new status of redundancy, consider phonetic information in bilingual dictionaries. A bidirectional paper bilingual dictionary will not give phonetic information for target language equivalents, instead requiring the user to look them up in the other direction by consulting the potential equivalents as headwords (if they *are* headwords in the other part—this need not necessarily be the case!). This traditional solution is non-redundant, as each pronunciation is represented just once in the whole dictionary; the alternative would require the information to be repeated for each occurrence of a word as equivalent (as well as next to the lemma sign in the other part), thus consuming substantial storage and presentation space.

In contrast, in an electronic dictionary we do not have to worry that much about storage space, and can thus present the pronunciation at every occurrence. In fact, by using a relational database storage structure, we can achieve this at no significant cost to storage space: it is enough to store the pronunciation information in a single record, and reuse it as necessary by invoking it via some record key.

This being said, *presentation* space is still restricted (even more so on screen than on paper, as I argue above), and there is also a real danger of information overload (call it screen clutter). Still, to remedy the restrictions on visual presentation, electronic dictionaries can draw on what I have referred to above as immediate cross-references (fanouts, popups, etc.) and take advantage of the dynamic potential of electronic displays in other ways (Sobkowiak 2007). Also, for information on pronunciation, electronic dictionaries can also utilize the audio channel, presenting the user with spoken pronunciation, which a paper dictionary cannot do. In embracing spoken pronunciation, however, we should not too hastily discard the old-fashioned graphic transcription, however, because it offers at least two advantages for

non-native speakers lacking in an audio representation. First, graphic transcription is explicit with regard to phonemic representation: a learner may not be able to notice phonemic distinctions when only presented as audio, since her perception is filtered through the phonological system of the native language (including the phonemic system, but also the stress system). Second, graphic transcription has an important indexical function: it allows access based on phonetic criteria.

4.2. Electronic dictionaries for production: new proposals

Tradition in lexicography is an important consideration. It is often said within the user perspective that dictionaries should serve the needs of the users. The paradoxical situation, however, is that the lexicographic needs of the users are largely shaped by their consultation habits, and those are of necessity based on older dictionaries. What we are dealing with, then, is a mild version of the vicious circle, where it may be difficult to part with conservative features (de Schryver 2003; Nesi 2000). However, interesting new proposals for electronic production dictionaries are (slowly) being put forward, such as the one by Batia Laufer and Tamar Levitzky-Aviad (Laufer & Levitzky-Aviad 2005; Laufer & Levitzky-Aviad 2006). In some ways, though, they are still hostage to the old paper principles.

4.3. Bilingual Dictionary Plus

Laufer and Levitzky-Aviad (2005; 2006), following up on an original idea by Laufer (1995), suggest a novel structure for a bilingual production dictionary consisting of the following four principal microstructural components (Laufer & Levitzky-Aviad 2006: 136):

1. L1→L2 translations
2. L2 information (definitions, examples, etc.) about each translation option
3. thesaurus-like information, i.e. words semantically related to each translation option
4. additional L1 meanings of the L2 translations

ברר (BERER)

find out (vt, past tense, past participle **found out**)

to learn something by study or inquiry: *Find out the cost and let me know. Please find out when the next train leaves.*

Related words:

get at (v prep, infml) to manage to find out

גילה (GILA)

I'm afraid we just can't get at the information; no one will help us.

determine (vt, fml) to find out exactly: **חישב וקבע** (CHISHEV VE-KAVA)

The police wanted to determine all the facts/what happened.

detect (vt, fml & tech) to find out:

הבחין (HIVCHIN)

We have been able to detect some improvement as a result of the medicine.

Other meanings of **find out**:

You've broken the vase and if your mother finds out she'll be angry. **ידע על כר**

(YADA AL KACH)

Don't steal pens; if you're found out there'll be trouble. **תפס** (TAFAS)

Figure 3: Example entry for Hebrew word BERER from Laufer and Levitzky-Aviad's Bilingual Dictionary Plus (after Laufer & Levitzky-Aviad 2006: 140).

The proposal is an interesting one. The basic idea of the Bilingual Dictionary Plus seems to be to “import” the L2 and L2→L1 information into each L1→L2 entry, thus aggregating in a single view all the lexicographic information that a user would get if they

looked up all of the L2 translations found in the L1→L2 dictionary section. Of course, such entry structure in a paper dictionary would produce a dictionary with a very high degree of redundancy and thus extremely bulky and uneconomical. Both these drawbacks carry much weight in paper lexicography and would likely far outweigh the benefits. But, here Laufer and Levitzky-Aviad capitalize on the fact that storage space is not an issue in an electronic medium, and so much richer content can be offered to the user. However, as I have tried to show, storage space is not the end of the story: careful consideration must also be given to the management of presentation space. In what follows, several preliminary suggestions are given that would be relevant to new-generation lexicographic tools such as the Bilingual Dictionary Plus.

5. Suggestions for dictionary making

There is evidence to indicate (e.g. Lew 2004) that presenting too rich a microstructure can lead to information overload. As a result, users find it difficult to extract the relevant information and may be less willing to proceed beyond the initial sense(s) of an entry. While it is important that the additional information (the *Plus* part in Bilingual Dictionary Plus) be easily accessible, not all of it need (should?) be presented at the same time. I believe that such dictionaries should utilize immediate cross-referencing to avoid the dangers of information overload. Likewise, reverse direction (in a production dictionary, L2→L1) content should also be dynamically cross-linked (= immediately cross-referenced, see section 3.2).

6. Suggestions for dictionary research

User research is needed to establish, firstly, what content should be displayed immediately on the screen, and what content should be deferred (for the concept of lexicographic information deferral, cf. Pujol et al. 2006). Further empirical study is called for to establish which specific cross-referencing devices work best for which uses, and with what user profiles.

7. A case for customizing lexicographic presentation

In an electronic dictionary, presentation of lexicographic content need not be static and may be made customizable. Here, at least two possible approaches to customization can be envisaged:

1. user-controlled customization
2. application-controlled customization, based on on-line monitoring and analysis of user behaviour

In the first case, the user would explicitly specify (select), either directly the lexicographic data types to display, or the type of task that she is currently engaged in, and the choice of the microstructural elements to be displayed would be predetermined by the lexicographers (ideally based on the type of research described above).

In the second scenario, no explicit user querying would be used, but there would be a tracking module in the dictionary software that would monitor the activities of the user, and determine the likely task type (e.g. on-line reading, composition) that the user is involved in, adjusting the presentation mode accordingly.

Some of the above suggestions may sound a little far-fetched, but they are pretty much achievable with today's technology, in a technical sense. What we do not yet have is a clear picture of the optimal set and shape of data to display to different users in different situations.

References:

Corréard, Marie-Hélène. 2002. Are space-saving strategies relevant in electronic dictionaries? *Proceedings of the Tenth EURALEX International Congress, EURALEX 2002, Copenhagen, Denmark, August 12-17, 2002, Vol.2*, Braasch, Anna & Claus Povlsen (eds.), 463-70. Copenhagen: Center for Sprogteknologi, Copenhagen University.

- de Schryver, Gilles-Maurice. 2003. Lexicographers' dreams in the electronic-dictionary age. *International Journal of Lexicography* 16(2): 143-99.
- Gouws, Rufus H. 2003. Types of articles, their structure and different types of lemmata. *A practical guide to lexicography, (Terminology and Lexicography Research and Practice 6)*, van Sterkenburg, Piet (ed.), 34-43. Amsterdam: John Benjamins.
- Graham-Rowe, Duncan. 2007. Electronic paper rewrites the rulebook for displays. *Nature Photonics* 1(5): 248-51.
- Hausmann, Franz Josef & Herbert Ernst Wiegand. 1989. Component parts and structures of general monolingual dictionaries: A survey. *Wörterbücher/Dictionaries/Dictionnaires. An international encyclopedia of lexicography, Vol.1*, Hausmann, Franz Josef, Oskar Reichmann, Herbert Ernst Wiegand & Ladislav Zgusta (eds.), 328-60. Berlin: Walter de Gruyter.
- Landau, Sidney I. 2001. *Dictionaries: The art and craft of lexicography*, 2nd edition. Cambridge: Cambridge University Press.
- Laufer, Batia. 1995. A case for a semi-bilingual dictionary for productive purposes. *Kernerman Dictionary News* 3: 2-4.
- Laufer, Batia & Tamar Levitzky-Aviad. 2005. Towards a bilingualized dictionary for second language production. *AsiaLEX Singapore Proceedings*.
- Laufer, Batia & Tamar Levitzky-Aviad. 2006. Examining the effectiveness of 'bilingual dictionary plus' – A dictionary for production in a foreign language. *International Journal of Lexicography* 19(2): 135-55.
- Lew, Robert. 2004. *Which dictionary for whom? Receptive use of bilingual, monolingual and semi-bilingual dictionaries by Polish learners of English*. Poznań: Motivex.
- Nesi, Hilary. 2000. Electronic dictionaries in second language vocabulary comprehension and acquisition: The state of the art. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000, Stuttgart, Germany*, Heid, Ulrich, Stefan Evert, Egbert Lehmann & Christian Rohrer (eds.), 839-41. Stuttgart: Institut für maschinelle Sprachverarbeitung, Universität Stuttgart.
- Pujol, Didac, Montse Corrius & Joan Masnou. 2006. Print deferred bilingualised dictionaries and their implications for effective language learning: A new approach to pedagogical lexicography. *International Journal of Lexicography* 19(2): 197-215.
- Sobkowiak, Włodzimierz. 2007. Innovative phonetic interfaces for electronic dictionaries. *Dictionary visions, research and practice: Selected papers from the 12th International Symposium on Lexicography, Copenhagen 2004*, Gottlieb, Henrik & Jens Erik Mogensen (eds.), 39–52. Amsterdam/Philadelphia: John Benjamins.
- Wiegand, Herbert Ernst. 1996. Textual condensation in printed dictionaries. A theoretical draft. *Lexikos* 6: 133-58.