Orthographical Dictionaries: How Much Can You Expect?
The Danish Spelling Dictionary Revis(it)ed

Henrik Lorentzen

Dept. for Digital Dictionaries and Text Corpora, DSL Society for Danish Language and Literature

Orthographical dictionaries constitute a particular and rather specialised subclass of dictionaries. This contribution offers a presentation of the ongoing revision of a spelling dictionary (for Danish) and a discussion of some of the general and specific issues that have arisen during the project. Firstly, the background is sketched, a brief overview of the many editorial changes is provided, and lemma selection, variant forms and definitions are discussed in some detail. Secondly, the field of orthographical dictionaries in Denmark is compared to some other countries of northern Europe. Finally, the conclusion engages in a discussion of the necessity of this particular type of dictionary.

1. Introduction

The tradition of official orthographical dictionaries for Danish goes back about 140 years. The first official spelling dictionary was published in 1872 and authorised by the Ministry for Ecclesiastical Affairs and Public Instruction (Kultusministeriet). For the following 80 years, a whole series of spelling dictionaries succeeded each other, edited by different scholars and teachers, but all of them officially recommended by the government. In 1955, the Danish Language Council was established as a governmental institution under the Ministry of Culture, and among its primary tasks it was to codify and publish the official orthographical standard of Danish in a spelling dictionary. The first dictionary edited by the Language Council was published in 1986 under the title of *Retskrivningsordbogen* (*The Orthographical Dictionary*, henceforth RO), two revised editions came out in 1996 and 2001, and now a completely updated version is planned for 2011.

Since this is the first real revision in 25 years, the project implies quite a number of changes in the dictionary text whereas the original concept of a printed dictionary in three columns is retained. The changes take place at all levels; I will, however, restrain myself to a discussion of the following general modifications: *lemma selection*, *variant forms* and *definitions*, but other important enhancements are information about word division, more examples of compound forming and more explicit information in the actual entries instead of references to general rules in the outside matter (cf. Schack 2007).

2. Lemma selection

The actual number of words or entries in a dictionary has always been a rather sensitive issue, and a high number of entries is in itself often regarded as an indisputable sign of quality. Statistically the chance of success when looking up a word is expected to be higher if the total number of words is higher, but there is more to it than that, two central factors being (a) the user's needs and (b) the purpose of the dictionary.

In principle, the user of a spelling dictionary may be virtually anyone seeking information about the spelling of a Danish word. Combined with the fact that emerges from several user surveys, namely that spelling, together with meaning, are the most looked-up information types in dictionaries (cf. Béjoint 2000: 141pp. and Jackson 2002: 76p.), it is obvious that a Danish spelling dictionary has a potentially very large user group, including all the pupils and students striving to master Danish orthography (which is probably as far removed from modern pronunciation as is the case for English). A primary conclusion from this may be that in order to satisfy most users in most situations the number of entry words should be relatively high, and that the lemma list should include words presenting difficulties on the

orthographical level. Furthermore, it seems evident that new words should be included since there may be uncertainty as to their spelling (or inflection), especially in the case of loanwords. It is far more doubtful whether old, rare and technical words have a right to be in an orthographical dictionary. As it happens, of the 60,000 words in the 1986 version of the Danish spelling dictionary almost 10,000 words were not included in the first draft version of *Den Danske Ordbog* (*The Danish Dictionary*), a corpus-based dictionary of modern Danish (Lorentzen 2008). This gives food for thought and has undoubtedly urged the editors of the future revised version to exclude a considerable amount of this part of the vocabulary.

The question remains, however, how many words should be included: should the number of headwords be reduced because of this elimination of marginal vocabulary or should it, on the contrary, be augmented with words that actually occur in the language? My personal view is in accordance with the latter option, but discussions in the editorial group reveal different opinions. If the principal purpose of the orthographical dictionary is to provide help in producing language, i.e. correct spellings and inflections, it is argued that a good stock of simplex words is important, combined with a certain amount of compounds and a good deal of examples illustrating how to form more compounds. Filling the dictionary with occasional compounds is not considered relevant. If, on the other hand, the purpose is documenting the vocabulary of the language, this is a good reason to include a lot of compounds, provided they occur in a corpus as evidence of them being used in actual language. Seen from the user's perspective it is also an advantage if the lemma list contains many headwords as it is rather comforting for a user to find the particular word he or she is looking for, as an indication of its usability. Reception (decoding) seems a less relevant purpose when spelling or inflection is concerned, but it comes into play when talking about meaning (see section 4).

The reluctance to add so-called transparent compounds to the lemma list may also stem from the idea that the end product is a printed book with a limited number of pages in order to cater for its use in classroom and examination situations. As emerged from several papers presented at the eLexicography conference in Louvain-la-Neuve in October 2009 (e.g. Nesi and Piotrowski, both forthcoming) the situation is rapidly changing from printed books to small hand-held devices, so there seem to be very few good arguments in favour of restricting the amount of vocabulary, although a relevant and realistic exception is that of funding and manpower.

3. Variant forms

As stated by Davidsen-Nielsen (2002: 35) there are always cases where words can be spelt in different ways. He further claims that in the case of Danish the official norm allows too many spelling variants, taking the example of *virus* that used to have 14 different official forms including variants in both singular and plural:

virus (singular indefinite)
virusen, *virussen*, *viruset*, *virusset* (singular definite)
viruser, *virusser*, *virus*, *vira (*plural indefinite)
viruserne, *virusserne*, *virusene*, *virussene*, *viraene* (plural definite)

In 2001 this rather confusing abundance of forms was reduced to 9 because the optional gemination of the s was abandoned. Davidsen-Nielsen argues that the current revision of the spelling dictionary could further reduce the number of forms to 4 thus bringing the lemma

virus in line with the vast majority of Danish nouns that have only 4 inflectional forms[1]:

virus (singular indefinite)
virussen (singular definite)
virusser (plural indefinite)
virusserne (plural definite)

The very fact of having an official orthographical norm somehow conflicts with having a lot of optional forms within the norm, and language users often express a need for clear answers, not a variety of possibilities to choose from. On the other hand, many forms have existed alongside each other for a long time, and discarding one in favour of another might be felt as too authoritarian. The policy of the editorial group is to accept a number of optional forms but only those that are found necessary. This of course implies invidual decisions generally based on corpus studies combined with systematical principles such as analogy and consistency. Thus a number of variant forms can be eliminated at word level, e.g. *federalisme/føderalisme* ('federalism') where the Danicized form *føderalisme* is predominant in actual use, and the opposite case: *schweizisk/svejtsisk* ('Swiss', adjective) where the more Danish-looking variant *svejtsisk* tends to be rare in corpus texts. The consistency principle applies to a number of inflectional forms, for example nouns ending in *-el*, which at present have an optional *-e-* in the singular definite form: *kabel – kablet/kabelet* ('cable'), *cykel – cyklen/cykelen* ('bicycle'). The tendency in actual usage is to prefer the syncopated form (*kablet*, *cyklen*) and by excluding the non-syncopated forms for this group of words, a considerable reduction of inflectional variants could be obtained as well as a transparent and consistent principle.

An alternative to the possibilities now applied in the RO, i.e. either one or more variant forms vs. one single authorised form, might be the one used in SAOL (Wordlist of the Swedish Academy); here the concept of prescription is introduced. When two variant forms are considered equally usable they are given like this: **sprej** or **spray** (no prescription), but where one of the variant forms is regarded less usable, it is introduced by the comment *also*: **sjal** also **schal** ('shawl') meaning that the former word form is prescribed. This compromise has been discussed in the Danish Language Council but this more fine-grained distinction between straightforward variants and prescribed variants has not been considered suitable, the rationale being that if you can recommend one form and not the other you might as well make a clear-cut decision and eliminate the non-recommended form.

4. Definitions

Up till now there has only been little information about meaning in the Danish spelling dictionary. In fact a frequent reaction when first becoming acquainted with it, is 'what do I actually get in this dictionary?' In orthographical dictionaries for German and Swedish it is much more common to give semantic information but in the Danish spelling dictionary the guiding principle was only to give enough information to distinguish homonyms or other cases where confusion is likely to occur, such as the following entries from the 2001 edition:

1. **bark** *sb. ...* (*lag på træ* 'layer on wood')
2. **bark** *sb. ...* (*fartøj* 'vessel')

amnesi *sb. ...* (*svækkelse af hukommelse* 'weakening of memory')
amnesti *sb. ...* (*eftergivelse af straf* 'remission of sentence')

---

1 And also with English dictionaries such as MACMILLAN, which only gives the forms *virus* (singular) and *viruses* (plural).

In the revised edition, the plan is to provide explanations not only when confusion is possible but also whenever the word in question is held to be marginal or outside the vocabulary of pupils of about 16 years of age (who use this particular dictionary when preparing for their final examination in Danish). According to the editorial plan the explanations are meant to 'ensure quick identification of the headwords', which implies a certain laconism, the meaning being conveyed by synonyms, short definitions, or labels, e.g.:

> **kapitulere** *vb. ...* (*overgive sig* 'surrender')
> **scherzo** *sb. ...* (*let, muntert musikstykke* 'light, lively piece of music')
> **dolce** *adv. ...* (*(mus.) blidt* 'gently')

There is little doubt that a larger amount of semantic information will improve the user value of the dictionary considerably as the decoding or reception function can be catered for to a much larger extent. On the other hand, enhancing the dictionary with more explanations turns out to be less straightforward than might at first be expected. An obvious difficulty is to decide which words need an explanation, and as in many other cases the decision depends on the judgment of the individual lexicographer and generalisations are hard to make.

Formulating the definitions and adjusting them to a common standard is a question that has been much debated in the editorial group. It turns out that the principle of short, identifying explanations is not easy to handle in practice when it comes to describing complicated matters. An example of the explanation serving as a pure identifier is

> **backgammon** *sb. ...* (*et brætspil* 'board game')

where the user gets no information about the game (its purpose or rules for instance). The adjective *organisk* ('organic') is an example of the reverse situation where the amount of semantic information given is very close to that of a genuine definition dictionary (for ease of presentation the definitions are only given in English):

> **organisk** *adj. ...* ('concerning living organisms; (chemistry) containing carbon compounds; having a natural, harmonic coherence')

To the user, this type of entry is no doubt a real improvement compared to earlier versions of the dictionary, but to the lexicographer, the more elaborate definitions may be a problem in different respects. One issue is to provide correct and sufficient definitions and the amount of labour it takes to do so, and another difficulty is the normative status that the average user will ascribe to the definitions, generally because users tend to believe what dictionaries tell them, and more specifically because the Danish spelling dictionary communicates the official norm in terms of orthography and inflection: the users can hardly be blamed for thinking that the definitions are just as official. This may be particularly problematic if a word or an expression is used in a controversial way not accepted by all speakers of Danish. An example of this are the words *blindskrift* ('touch-typing') and *blindeskrift* ('Braille writing'), which tend to be mixed up by language users, chiefly *blindskrift* being used in the sense of Braille. The problem for the lexicographer is now whether (and how) the users should be informed about this and (at least) three possibilities emerge:

1. Ignore the confusion and keep the senses apart in two different entries
2. Give one sense for *blindeskrift* and two senses for *blindskrift* describing the use 'Braille' as incorrect
3. As in 2. but leave out the usage information

The possibilities reflect the dichotomy norm vs. description with no. 1 being strictly normative, no. 3 strictly descriptive and no. 2 in between. In view of the above-mentioned tendency to interpret dictionary data normatively, no. 2 may represent a reasonable compromise by informing of actual usage and at the same time displaying a warning sign saying 'this use may not be accepted by everyone'. This is in fact the solution adopted in the corpus-based DDO (The Danish Dictionary), which intends to combine description and prescription. See Trap-Jensen (2002) for further discussion of descriptive and normative aspects in dictionary making.

5. Other orthographical dictionaries

As mentioned above, the Danish spelling dictionary is official and authorised by law as the way the Danish Language Council communicates the orthography standard to the general public, a standard the national and local authorities as well as the educational system are obliged to use. This means that the spelling dictionary of the Language Council enjoys monopoly status, a status that was nevertheless infringed in 1996 and 2001 so that Denmark now has two unofficial spelling dictionaries (GRO and PRO). Among their main features are a higher number of entry words and more definitions, features that are partly taken into account in the revision of the official RO.

Germany has a long history of orthographical dictionaries but not one authorised by government. However, DUDEN functions in practice as conveyor of the norm and not only that, it also gives a lot of additional information on, for instance, grammar, pronunciation, etymology and meaning, thus approaching the general-purpose dictionary (Nerius 1990: 1302).

In Sweden, there is no official spelling dictionary either but unofficially SAOL has the role of defining the standard of Swedish orthography. The wordlist, which in its latest edition is a book with 1130 pages and 125,000 headwords (RO 2001: 749 pages and 64,000 headwords), has been published since 1874, two years after the publication of the first Danish spelling dictionary. Apart from orthography and inflections it provides a certain amount of information on pronunciation, meaning and word division.

The example entries in Table 1 clearly illustrate the differences in the amount and type of information provided by four spelling dictionaries. For ease of understanding and comparison the entry word is the English loanword *pacemaker* (smaller differences from the original notation may occur).

DUDEN:	SAOL:
Pace|ma|ker [...*me:*kE], der; -s, - (Pacemacher; *Med.:* Herzschrittmacher)	**pace|maker** [pej'smejk-] s. *-n pacemakrar •* hjärtstimulator; <sport.> farthållare i löpning
PRO:	RO 2001:
pace·maker sb. (apparat) *-en, -e, -ne*	**pacemaker** *sb.*, -en, -e, *bf. pl.* pacemakerne.

Table 1: The entry **pacemaker** in four different spelling dictionaries.

The German DUDEN and the Swedish SAOL both give more information than just orthography and inflection, indicating pronunciation, word division and two different senses. PRO, one of the unofficial Danish spelling dictionaries, is more sparing but provides

nevertheless the word division and a short gloss for one of the possible senses[2]. RO, however, restricts itself to conveying a minimum: orthography and inflection. As stated above, especially in Section 4, this is going to change in the edition under preparation: word division will be provided for all headwords and semantic information in the form of glosses or definitions will be given in many more cases than was common in earlier editions. As a result of this policy, the future entry for **pacemaker** is likely to look like this:

RO 2011:
pace|maker *sb*., -en, -e, -ne *(elektronisk apparat der stimulerer hjerterytmen)*.

6. Conclusions

When considering the German DUDEN and the Swedish SAOL as well as the unofficial Danish spelling dictionaries, there is no doubt that some of their main features have served as inspiration for the next edition of the Danish RO, in particular information on meaning by way of definitions and information on word division. Now the question is: Why have a special spelling dictionary? It could be argued that spelling is just one information type among others. And once other elements than orthography and morphology such as definitions are introduced, it might no longer be reasonable to maintain the spelling dictionary as a subtype of its own (cf. discussions of polyfunctionality by Bergenholtz 1997 and Nerius 1990 among others).

Seen from the narrow Danish perspective, there seem to be good arguments in favour of preserving the spelling dictionary since the law stipulates that the Language Council is responsible for defining the official orthography and for making it accessible to the Danish population. But it is not absolutely certain that it need be in the form of one particular, physical dictionary: the important thing is to communicate this highly relevant information type to the relevant users (probably most of the potential users), and why not as part of a more general reference work as long as the spellings are clearly indicated to be in accordance with official recommendations?

---

2 The sports sense of *pacemaker* ('someone who sets the pace in a race') is uncommon in Danish, this sense generally being conveyed by the word *pacer*.

References

Béjoint, H. (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.

Bergenholtz, H. (1997). 'Polyfunktionale ordbøger'. In *LexicoNordica 4*. 15-29.

Davidsen-Nielsen, N. (2002). 'Remarks on Dictionaries – with Special Reference to Grammar, Pronunciation and Orthographic Variation'. In Braasch, A.; Povlsen, C. (eds.). *Proceedings of the Tenth EURALEX International Congress*. Copenhagen. 29-38.

DDO = *Den Danske Ordbog 1-6*. København: Det Danske Sprog- og Litteraturselskab og Gyldendal. 2003-2005.

DUDEN = *Duden – Die deutsche Rechtschreibung*. 22. Auflage. Mannheim: Dudenverlag. 2001.

GRO = *Retskrivningsordbog*. København: Gyldendal 1996, 2001.

Jackson, H. (2002). *Lexicography: An Introduction*. London/New York: Routledge.

Lorentzen, H. (2008). 'Korpusbaseret lemma-eksklusion eller 'Hvor gik ordene hen da de gik ud?''. In Jóhannesson, K. et al. (eds.). *Nog ordat? Festskrift till Sven-Göran Malmgren*. Göteborg. 259-267.

MACMILLAN = *Macmillan English Dictionary* online: http://www.macmillandictionary.com [access date: 26 Feb. 2010].

Nerius, D. (1990). 'Das Orthographiewörterbuch'. In Hausmann, F. J. et al. (eds.). *Wörterbücher / Dictionaries / Dictionnaires*. Bd. 1–3. 1989–1991. Berlin/New York: Walter de Gruyter. 1297-1304.

Nesi, H. (forthcoming). 'E-dictionaries and Language Learning: Uncovering Dark Practices'. In Granger, S.; Paquot, M. (eds.). *Proceedings of eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve.

Piotrowski, T. (forthcoming). 'Mobile Dictionaries: Situations of Use'. In Granger, S.; Paquot, M. (eds.). *Proceedings of eLexicography in the 21st century: New challenges, new applications*. Louvain-la-Neuve.

PRO = *Politikens Retskrivningsordbog*. København: Politikens Forlag. 2001.

RO = *Retskrivningsordbogen*. København: Dansk Sprognævn. 1986, 1996, 2001.

SAOL = *Svenska Akademiens ordlista*. 13. upplagan. Stockholm: Norstedts Akademiska Förlag. 2006.

Schack, J. (2007). 'Eventuelle ændringer i en kommende udgave af Retskrivningsordbogen". In Sandersen, V.; Schack, J. (eds.). *Dansk Sprognævn – godt og vel 50 år efter*. København. 93-103.

Trap-Jensen, L. (2002). 'Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases'. In Braasch, A.; Povlsen, C. (eds.). *Proceedings of the Tenth EURALEX International Congress*. Copenhagen. 503-509.