Remarks on Dictionaries – with Special Reference to Grammar, Pronunciation and Orthographic Variation

Niels Davidsen-Nielsen

Copenhagen Business School Faculty of Modern Languages 15 Dalgas Have 2000 Frederiksberg

Abstract

The languages dealt with in this paper are English and Danish. In a section on pronunciation it is shown that English words such as disperse, distract, discover with medial /sp, st, sk/ were formerly transcribed incorrectly in British English dictionaries. As a result of an experimental investigation carried out by the author, however, errors of this type have subsequently been corrected. A section on grammar in English language dictionaries discusses so-called multi-word verbs, i.e. verbs like turn down (= reject), wait on (= serve), put up with (= tolerate) and take place (= happen). The author proposes a classification of these verbs, arguing that two widely used English dictionaries tend to operate with too many of them. Finally, a section on spelling variation discusses the approach to such variation adopted in the Danish Orthographic Dictionary. Here the author argues that many instances of spelling variation can justifiably be eliminated.

Introduction

When I was asked to give a plenary lecture at the 10th International EURALEX Congress, I could honestly say that I am not a lexicographer. I could equally honestly add, however, that throughout my career I have been an ardent user of dictionaries. This was evidently sufficient for the organisers, so here we are. I am honoured to be able to address such an august assembly of professional lexicographers, publishers, researchers, scholars and others interested in dictionaries of all types.

As announced, I am going to talk about grammar, pronunciation and orthographic variation in dictionaries, and the languages involved will be English and Danish. As the international auxiliary language of today, English is undoubtedly mastered by everybody at this congress. Danish is obviously not, but then Danish can be regarded as an exotic language which also needs its dictionaries. Among the 5 or 6000 languages in the world, incidentally, Danish ranks among the top 100 in terms of number of speakers.

Pronunciation

The only contribution I have made to lexicography involves English pronunciation and dates back to the seventies. In 1974 I published an article called 'Syllabification in English words with medial sp, st, sk'. In the words investigated, word-medial sp, st, sk are followed by a stressed vowel, as in *despise, aristocracy, discuss*. Words of this type were recorded by native speakers of British and American English together with words like *pool, tool, cool* and *spool, stool, school* serving as a frame of reference.

In words like *pool*, *tool*, *school* the initial consonant is aspirated, i.e. pronounced with a puff of air, and in words like *spool*, *stool*, *school* the consonant after *s* is unaspirated. Now if the stop consonant in the words investigated with medial sp, st, sk turned out to be aspirated, my interpretation was that there is a syllable boundary immediately before this stop consonant: *des-pise*, *aris-tocracy*, *dis-cuss*. If it turned out to be unaspirated, on the other hand, I assumed that the syllable boundary was placed before the *s*: *de-spise*, *ari-stocracy*, *di-scuss*.

In the recorded words – which had been pseudo-randomised so that words of the same type did not occur conspicuously together – I measured the release stage duration (i.e. the degree of aspiration) by means of a pitch meter and an intensity meter and by using a so-called mingograph as a registering apparatus.

The results were very clear and in accordance with my hypothesis. The stop consonants in words with medial sp, st, sk were normally unaspirated, so in e.g. despise, aristocracy, discuss the syllable boundary turned out to be placed before the s. The only exception to this pattern involved words containing a prefix ending in -s followed by an intuitively transparent morpheme boundary, i.e. words like mis-time, dis-courteous, mis-calculate were pronounced with a syllable boundary between the s and the stop consonant.

An important factor motivating my investigation was that British English dictionaries had until that time indicated syllable division between the s and p, t, k in all cases, i.e. also in words without a transparent morpheme boundary like despise, aristocracy, discuss. When looking up words with medial sp, st, sk users would therefore erroneously assume that the stop consonant should be pronounced with aspiration not only in e.g. mistime, discourteous, miscalculate – where there is a transparent morpheme boundary – but also in e.g. despise, aristocracy, discuss where there is no such boundary. In many instances the dictionaries simply let their users down in indicating pronunciations which were wrong. Only one of the British English dictionaries correctly prescribed syllable division before s in a large majority of cases, namely Jack Windsor Lewis's A Concise Pronouncing Dictionary of British and American English. In that dictionary, however, syllable division before s was also prescribed in a number of cases where there is a transparent morpheme boundary, for example in displace, distrust, discomfort. Altogether, then, the British English dictionaries were not reliable guides to the pronunciation of words with medial sp, st, sk.

Fortunately, the dictionaries are now reliable with respect to the pronunciation of the words I investigated, and have been for quite some time. In the 14th edition of *Everyman's English Pronouncing Dictionary* from 1977 by A. C. Gimson a complete revision of syllable division in words with medial sp, st, sk was made on the basis of my experimental investigation. Permit me to quote from the introduction to that dictionary:

"/p, t, k/ are typically accompanied by aspiration ... especially when initial in a stressed syllable ... However, in the stressed syllable-initial sequences /sp-, st-, sk-/, /p, t, k/ lack such aspiration. Within a word, therefore, the situation of the stress accent in relation to such sequences will denote presence or absence of aspiration ... I have in many cases ... found it necessary to shift the stress accent given in previous editions in order to denote or preclude aspiration of the plosive... The criterion determining the separation of /s/ from /p, t, k/

appears to be the presence of an 'intuitively transparent morpheme boundary' (i.e. a separation based on clearly felt sense units)."

In a footnote Gimson kindly adds the following about 'intuitively transparent morpheme boundary':

"The term is used by Niels Davidsen-Nielsen, to whom I am much indebted for advice; see his article 'Syllabification in English words with medial sp, st, sk', *Journal of Phonetics*, vol. 2, no. 1 (1974)."

Gimson's revision of the pronouncing dictionary was naturally gratifying. Occasionally research does give tangible practical results. An error in the British English dictionaries with respect to the pronunciation of quite a large class of words had been corrected.

As already stated, users of dictionares today can rely on the syllable divisions shown in the words in question. The same also goes for John Wells's influential *Longman Pronunciation Dictionary*. Experimental phonetic work is thus important in that it can bring about revisions of dictionaries so that these become more reliable tools for their users.

Grammar

It is a characteristic feature of the English language that it has developed a large class of so-called multi-word verbs, i.e. verbs like TEAR UP, WAIT ON (= serve), TAKE PLACE and LOOK DOWN ON which begin with a verb form and contain one or more words from other classes. What were originally single-word verbs followed by another sentence constituent have developed into complex verbs which the following constituent has been absorbed into. Multi-word verbs have existed in English for several centuries, but since the middle of the 19th century their number has increased sharply. Multi-word verbs are also familiar from other languages. In German we find examples of the type Hör auf mit deinem Quatsch 'Stop drivelling', but German differs from English in that the particle can also be prefixed to the verb: Nun muss dein Quatsch aufhören. The same pattern is found in Danish, as illustrated by examples like Vi stiller en ny kandidat op and Vi opstiller en ny kandidat 'We'll nominate a new candidate'.

In a multi-word verb like TURN DOWN the two elements have fused syntactically, so in an example like We turned down the offer the sentence constituents we find are subject (We), verbal (turned down) and object (the offer). In other words, DOWN is here part of the verbal and does not belong to a separate adverbial as it does in We turned down the street. But the elements of TURN DOWN in the sense of 'reject' have also fused semantically: TURN DOWN has idiomatic meaning, for both the verb and the adverb have lost their original literal meaning – that of 'move round a central point' in the case of TURN and concrete directional meaning in the case of DOWN. In this way the meaning of TURN DOWN is clearly different from that of TURN + DOWN as in We turned down the street.

In those cases where both syntactic and semantic fusion have taken place it would seem natural for a dictionary to operate with individual lexical items. This is not always what we find, however. In a case like TAKE PLACE the dictionary user has to look under PLACE, so

a verb that means 'happen' has to be found under a noun that means 'a particular area or position' (see Oxford Advanced Learner's Dictionary (ALD) and Longman Dictionary of English Language and Culture (Longman)).

As I pointed out, the class of multi-word verbs is a very large one in contemporary English, though precisely how large depends on how much fusion lexicographers insist on before they are prepared to include a multi-word construction in the class. If we check under the verb GO, we see that ALD and Longman list about 30 multi-word verbs (in both dictionaries the term phrasal verb is used). As illustrated by GO OFF in some of its senses, for example 'fall asleep' (Has the baby gone off?) and 'become bad and unfit to eat' (Has the food gone off?), there is often semantic fusion, but in many other cases there is only syntactic fusion, as in the case of GO IN. Like GO, verbs like TAKE and COME form part of a large number of multi-word verbs. Also here it appears that semantic fusion is not a requirement for the recognition of multi-word verbs. Examples illustrating this are TAKE AWAY and COME IN.

Let me now turn to the grammar of multi-word verbs. As I analyse them, multi-word verbs are divided into four subclasses (see Bache & Davidsen-Nielsen 1997: 85ff):

1. Phrasal verbs

Here it is an adverb that has fused with the verb. Phrasal verbs can be intransitive, as in *I give in*, or transitive, as in *She tore up the letter*. In transitive phrasal verbs, the verb and the adverb are separable, as appears from *She tore the letter up* where the adverb comes after the object. If the object is an unstressed pronoun, separability is even obligatory: *She tore it up*. The adverbs of phrasal verbs are drawn from a fairly small class of one-syllable or two-syllable adverbs with locative meaning. Furthermore, they are pronounced with primary stress. Another characteristic is that phrasal verbs can often be turned into nouns, cf. examples like *break-in*, *comeback*, *go-ahead*, *take-off*.

2. Prepositional verbs

Here it is a preposition that has fused with the preceding verb. Prepositional verbs are always transitive, as illustrated by *Miranda was waiting on the customers* where the object is *the customers*. In spite of the fact that the preposition belongs to the verb, many grammarians speak of 'prepositional objects' here. Unlike phrasal verbs, prepositional verbs are inseparable. The preposition cannot be moved to the position after the object. Unlike adverbs in phrasal verbs, furthermore, prepositions in prepositional verbs are typically unstressed, as illustrated by the example with WAIT ON just given. Occasionally, however, the second element of a prepositional verb is pronounced with primary stress. This can be exemplified by a sentence like *Stephen took after his father* (where *took after* means 'resembled'). That the multi-word verb here is clearly a prepositional verb and not a phrasal verb is evident owing to the fact that its elements are inseparable, i.e. AFTER cannot be moved to the end of the sentence. Prepositional verbs cannot normally be turned into nouns.

3. Phrasal-prepositional verbs

In this subclass the verb is followed first by an adverb and then by a preposition, as illustrated by the following examples:

Cassandra looked down on the nurses. (= despised)
She came up with a solution in no time. (= found)
I won't put up with your behaviour. (= won't tolerate)
I grew out of that habit a long time ago. (= abandoned)

Like prepositional verbs, phrasal-prepositional verbs are transitive and inseparable. Primary stress always falls on the adverb. As is apparent from the examples, not only syntactic fusion but also semantic fusion is normally involved. For example, the meaning of PUT UP WITH cannot be inferred from the meanings of the three words involved. Unlike phrasal verbs and prepositional verbs, phrasal-prepositional verbs consitute a small class.

4. Other multi-word verbs

Finally, there is the category 'others'. Here the verb is followed not by an adverb or a preposition but e.g. by an adjective, a noun or a verb:

Sam *cut short* his sermon. We'll *break even* later this year.

The peace talks *take place* in Oslo. I *caught sight of* my old friend in town today.

And we had to make do with gin. I can't get rid of these 20 copies.

This subclass, which also includes multi-word verbs like BE ON THE POINT OF and BE ABOUT TO, is characterised by inseparability like prepositional verbs and phrasal-prepositional verbs. On the other hand it resembles the subclass of phrasal verbs in that its members can be transitive or intransitive. Primary stress falls on the word after the verb.

Now a major problem facing the lexicographer is obviously where to draw the boundary between multi-word verbs and single-word verbs followed by an adverb, preposition, adjective, noun or another verb. For example, while *Miranda was waiting on the customers* clearly contains a multi-word verb and *Miranda was waiting on the corner* clearly does not, it is more difficult to decide whether *Miranda was looking at the customers* contains a multi-word verb or not. Here LOOK AT means exactly what it says, and the preposition AT has the same relational meaning as it has when not serving as part of a prepositional verb. Is LOOK AT sufficiently fused to be analysed as a multi-word verb?

In both the dictionaries I have referred to (ALD, Longman) LOOK AT is recognised as a multi-word verb. Syntactically, an argument in favour of this analysis is that sentences with LOOK AT can be passivised (e.g. *The customers were often looked at with curiosity*). Since the noun phrase that comes after LOOK AT can become the subject of a corresponding passive sentence, it behaves like an object in an ordinary S V O sentence. Furthermore, unless we adopt this analysis, the preposition is left 'stranded' by passivisation.

Whenever passivisation is possible, then, it would seem that we can safely assume that a potential multi-word verb is a genuine multi-word verb, also in the absence of semantic fusion. This is not quite the case, though, as appears from an example like Someone has slept in the bed. Here passivisation is possible (The bed has been slept in), but as there is no fusion between SLEEP and IN, it does not seem possible to analyse the bed as an object. What we find in this case, then, is passivisation of an intransitive S V A(dverbial) sentence. Conversely, not all transitive multi-word verbs can be made passive. For example, I've gone off beer and Jane seems to be going off Peter have no passive equivalents. Finally, it should be recalled that the passivisation test cannot be applied to instransitive multi-word verbs. If in doubt about the status of FALL DOWN in a sentence like She fell down and hurt her knee, for example, we obviously cannot turn to passivisation for an answer. While passivisation is an important test for lexicographers in their attempts to tackle borderline cases, it is therefore not a test which can be used in all cases.

The basic problem posed by multi-word verbs is clearly that there is quite a large grey area where a verb and the following adverb, preposition, adjective, noun or verb have fused somewhat but not a lot. Here are some examples – all taken from ALD – which in my view illustrate such intermediate cases:

The case comes before the court next week.

The daffodils are just beginning to come up.

The main road was flooded so we had to go round by a narrow country lane.

Let's go in, it's getting cold.

We only take goods back if the customer can produce a receipt.

Workmen arrived to take down the scaffolding.

If we compare these examples with an example like'This chapter takes up where the last one left off', it is evident that the degree of fusion is relatively low. And many grammarians would probably analyse the italicized word after the verb as a separate adverbial (up, round, in, back, down) or as part of a separate adverbial (before the court). The general approach adopted in the two dictionaries I refer to is to operate with multi-word verbs also in cases like these with rather little fusion. As fusion is likely to increase over time, such a catholic approach may perhaps demonstrate foresight. In some cases, incidentally, ALD practices an even more open policy of admission than Longman, for example by recognising GO AWAY, GO BEFORE and GO BEYOND as multi-word verbs.

No matter how the lexicographer attempts to tackle the problem of grey zones and intermediate cases, the task requires that a cut-off point be proposed, and as far as multiword verbs are concerned that is by no means easy.

Finally, I would like to take a brief look at how the two dictionaries handle multi-word verbs grammatically. In both dictionaries these verbs are called phrasal verbs; and in such phrasal verbs the initial verb is said to be followed by an adverb, a preposition or both. As a result, not only TURN DOWN but also EAT INTO and PUT UP WITH are considered phrasal verbs. From a grammatical point of view I see three problems in conflating the three subtypes in this way:

- 1. One misses the point that prepositional verbs and phrasal-prepositional verbs can never be intransitive. In the dictionaries we only read that phrasal verbs can be transitive or intransitive.
- 2. One misses the point that it is only multi-word verbs consisting of verb + adverb which may have a noun related to them. In the dictionaries we read that particular multi-word verbs have this property but not which ones.
- 3. One misses the point that it is only multi-word verbs consisting of verb + adverb which are separable. In the dictionaries it is stated that some multi-word verbs are separable but not which ones.

Finally, it is obviously a problem not to include into the class of multi-word verbs those types where the initial verb is followed by an adjective, a noun or another verb, as illustrated by CUT SHORT, TAKE PLACE and GET RID OF. In ALD the reader has to look up under SHORT, PLACE and RID, and the fusion involved in these multi-word verbs then has to be inferred from the label 'idiom'. In Longman the reader has to look up under CUT, PLACE and RID.

It is obviously unreasonable to demand that a dictionary should provide the same amount of grammar as a grammar book. So the account of the grammar of multi-word verbs has to be kept simple. In their degree of simplification I think that both ALD and Longman have struck a reasonable balace. On the other hand, I find their open door policy a bit excessive.

Orthographic Variation

In dictionaries there will always be cases where a word is spelled in more than one way. A word like *judg(e)ment* can be spelled in two ways: with or without an *e* at the end of the first syllable. Even disregarding the differences between British and American English, for example the use of one or two consonants in words like *fulfil(l)* and *bus(s)es*, many words have more than one spelling. In British English, words ending in /alz/ can be spelled with – *ise* or –*ize*, for example *realise/realize*, *idealise/idealize*, *generalise/generalize*, *hypnotise/hypnotize*, *memorise/memorize*, *theorise/theorize* and *vandalise/vandalize*. Words spelled with *oe* like *foetus*, *foetal*, *oesophagus* can also be spelled without an *o*. Words spelled with *ae* like *haemophilia*, *haemorrhage*, *haemorrhoids* can also be spelled without an *a* in the first syllable. A word like /'kæftæn/ can be spelled with *c* or *k*, and so on.

Let me now turn to Danish. In Denmark we have an official orthographic dictionary. It is published by the Danish Language Council – which I am vice-chairman of – and it defines what correct Danish spelling is. In this dictionary there is also spelling variation, quite a lot in fact, and for some time that has been a controversial political issue.

When Danes consult their official orthographic dictionary, they typically want one answer only. Their view is that if they go to the trouble of looking a word up, they do not wish to be given a choice between different spelling possibilities. Fair enough, perhaps, but on the other hand it is a fact of life that in many cases what we find in texts written by meticulous, educated Danes are two different spellings of the same word, sometimes evenly represented. For example, the Danish word for 'eleven' is written by some Danes as *elve* and by others as

elleve. Similarly, the Danish word for 'café' is written by some Danes with an acute accent over the e (cafe) and by others without this diacritic mark (cafe). Quite a number of words, furthermore, have two different plurals. A word like høvding 'chief' is by some Danes pluralised with -e (høvdinge) and by others with -e (høvdinge).

My own view is that there is more spelling variation in the Danish orthographic dictionary than necessary, so to some extent I understand the reaction from many Danes who object to what they see as indecision on the part of the Language Council, and who want us to make up our minds and prescribe – particularly since the orthographic dictionary is normative.

Let me illustrate by means of the word VIRUS ('virus'). According to the 1996-version of the orthographic dictionary this word has three plural forms: virus(s)er, virus or vira. Secondly, the gender of VIRUS may be common or neuter, with the result that the user has to choose between virus(s)en and virus(s)et. Thirdly, the final s may be doubled or not, both in the singular and the plural: virussen or virusen, virusset or viruset, virusserne or viruserne. Fourthly, the definite form of plural vira is viraene. And finally, the definite form of the plurals virus(s)er and virus are respectively virus(s)erne and virus(s)ene (with or without an r). What all this adds up to is that the lexeme VIRUS has fourteen different forms. The mind boggles, and I sympathise with the suffering Danish public.

In the 2001-edition of the orthographic dictionary the number of forms was nearly halved, for now it was only correct to write two s's: virussen, virusset, virusser, virusserne, virussene. This could be justified partly because the vowel preceding ss is short and qualitatively unreduced, partly because the spelling with two s's is used more often than that with a single s.

Are there grounds for further simplification of VIRUS in the next edition (2006) of the orthographic dictionary, you may well wonder? In my view it is certainly worth considering whether we could eliminate the neuter form of this word, for it appears to be much rarer than the common gender form. Secondly, it is worth considering whether we could eliminate the plural zero-form *virus*. That, too, appears to be relatively rare, and altogether the zero plural that we find in words like *mus* 'mice' and *ting* 'things' is not really productive.

If we were to eliminate these forms, the dictionary entry for VIRUS would look like this:

virus n., -sen, -ser or vira, def. pl. virusserne or viraene

Gender, incidentally, is the cause of a good deal of spelling variation, and in some cases this variation has to be represented. For example, *gummi-et* and *gummi-en* 'the rubber' are equally common among educated, careful Danes.

But I am not quite through with VIRUS. Yet another possible simplification would be to eliminate the plural forms *vira* and *viraene*. To be sure, these forms are not infrequently used, particularly by members of the medical profession. But in classical Latin *virus* has no plural form at all! Like *pelagus* 'sea' and *vulgus* 'populace' it is an uncountable neuter noun following the second declension. Cicero – the most eloquent Roman of all – would never

have dreamed of using a plural form like vira. To be sure, this form is regular enough if virus could be pluralised – but it couldn't. If the Romans could do without it, so perhaps can we. By eliminating vira and viraene the Danish Language Council would surely be criticised, but we could face such criticism with some equanimity, for these forms are less frequent than virusser and viruserne, and vira is 'incorrect' in the sense that it has no direct support in classical Latin.

What we could end up with in the orthopgraphical dictionary is thus the following:

virus n., -sen, -ser, def. pl. virusserne

That would bring us into line with English dictionaries, incidentally, for what we find by looking up VIRUS in ALD is virus, pl. viruses.

In the preface of the 2001 orthographic dictionary the following is stated about spelling variation [my translation]:

"... in about 1,000 of the second edition's word-articles, changes and adjustments have been made. A great many of the changes result from the Language Council's efforts to reduce the number of double forms in the Orthographic Dictionary."

Personally, I wholeheartedly support these changes, and in future editions of the orthographic dictionary I think there is room for further reduction in the number of double forms.

Closing Words

In Denmark it was the general rule for a long time that the spelling of loanwords was adapted to the rules governing the relationship between spelling and pronunciation. For example, the French loanword milieu was changed to miljø, and the Low German word for 'strain' or 'sieve' was changed from dorchslach to dørslag. In 1986, however, a so-called mayonnaise war broke out in Denmark. The reason for that was that in the edition of the Orthographic Dictionary that came out that year a handful of loanwords, including mayonnaise, could now also be spelled the Danish way: mayonnaise or majonæse, remoulade or remulade, creme or krem, ressource or resurse. Even the then Minister of Education played an active part in the popular uprising and deplored the low standards of the Language Council.

Considering the orthographic practice followed earlier on in Denmark, and considering the fact that orthographic adaptation of loanwords has since 1862 been consistently practiced in Norwegian – a very closely related language indeed – this may seem surprising. It is not quite as surprising as it sounds, though, for with respect to spelling there is now a basic difference between Danish and Norwegian. In Norwegian, spelling is basically governed by the phonetic principle: words are spelled the way they are pronounced, and spelling a word like *mayonnaise* in the French way would be unthinkable. In Danish, on the other hand, spelling is no longer governed by the phonetic principle. We write *mord* 'murder' with a

final d, for example, in spite of the fact that no final d is pronounced and in spite of the fact that mord rhymes with hor 'adultery', where there is no written d.

The advantage of spelling *mord* with a final d is that this way of writing connects *mord* with the derived word *morder* 'murderer' where the d is pronounced. In Danish spelling it is not only derived words we take into consideration but also inflected forms of the same word. For example, we spell bage 'bake' with a g which is not pronounced, and in so doing we can connect this word with its inflected forms bagte and bagt 'baked' where the g is pronounced. In our Danish spelling practice, finally, we also consider the origin of words. For example, we spell chaiselong the French way (apart from the fact that two final vowels have dropped), and that makes it easier for Danes who are familiar with French to understand the word than if we had written sjeselong the way they do in Norway.

With respect to spelling, then, there are a great many factors that need to be taken into consideration before the Danish Language Council can for each individual word arrive at a solution in its Orthographic Dictionary. That cannot be helped, though, and all things considered the job of the lexicographer is more rewarding if instead of mere plodding it involves the cracking of some really hard nuts.

References

[Bache & Davidsen-Nielsen 1997] Bache, C. & N. Davidsen-Nielsen, 1997. Mastering English. An Advanced Grammar for Non-native and Native Speakers. Mouton de Gruyter, Berlin/New York. [Davidsen-Nielsen 1974] Davidsen-Nielsen, N. Syllabification in English words with medial sp, st, sk. Journal of Phonetics 2, 15-45.

Dictionaries:

A Concise Pronouncing Dictionary of British and American English. Oxford University Press 1972. Everyman's English pronouncing Dictionary. London 1977 (14th edition).

Longman Pronunciation Dictionary. Longman 1990.

Longman Dictionary of English Language and Culture. Longman 1998 (2nd edition).

Oxford Advanced Learner's Dictionary. Oxford University Press 2000 (6th edition).

Retskrivningsordbogen. Dansk Sprognævn. Aschehoug 1996 (2nd edition).

Retskrivningsordbogen. Dansk Sprognævn. Alinea A/S – Achehoug Dansk Forlag A/S 2001 (3rd edition).