
Grammatical information in dictionaries

Eric Hoekstra¹

Fryske Akademy, Leeuwarden, The Netherlands

A dictionary is an encyclopaedia of linguistic information about words. It presents to a target group of laymen and professionals general information about words belonging to various disciplines of linguistics such as

- *semantics (the meaning of words and phrases)*
- *phonology (the pronunciation of words)*
- *syntax (the syntactic category of words and the collocations in which they partake)*

My contribution discusses the question whether (new) insights from these disciplines may change the content of dictionaries, seeing that an evaluation of these insights does not take place very often.

It is a shortcoming of dictionaries that a paraphrase of the meaning of function words is often not very insightful with respect to their use (Coffey 2006).

- *What is the meaning of Dutch *er* 'there'?*
- *What is the meaning of articles like *the*?*
- *What is the meaning of the complementiser *that*?*

*Some Dutch dictionaries muddle the description of the various uses of *er*, ignoring the distinctions drawn by *de Algemene Nederlandse Spraakkunst, the Standard Dutch Grammar (Haeseryn et al. 1997)*. Those distinctions are practical and well-motivated (Hoekstra 2000). It is proposed to use syntactic knowledge to structure articles about function words. In addition, dictionaries can covertly use example sentences to illustrate syntactic phenomena. Such measures strengthen the encyclopaedic character of a dictionary.*

1. Dictionary information is encyclopaedic and multi-disciplinary in nature

This contribution focuses on the type of grammatical information about words which is represented in dictionaries, and, more specifically, how function words are dealt with. In a way, a dictionary can be likened to an encyclopaedia. Like an encyclopaedia, a dictionary presents information that must be relevant for its target group of laymen, among which higher-level learners of language may be distinguished as an important subgroup (Coffey 2006: 159). Of course, the specific information contained in a dictionary may be tailored to the needs of the specific target group which it is written for (Bergenholz; Mugdan 1985: 15).

A dictionary is a repository of various kinds of linguistic information about words. Thus the following pieces of information about a word may be found in a dictionary:

- its meaning
- its pronunciation
- its etymology
- its substructure with respect to derivational morphology (distinguishing affixes)
- its membership of a paradigm of inflectional morphology
- the syntactic category to which it belongs
- the collocational or idiomatic structures in which it may be found

The information which the dictionary presents belongs to various branches of linguistics. When it gives the meaning of words, it is concerned with semantics. The semantics of content words has always been the focus of investigation, but in addition, a lot of work has been done in the past forty years on the semantics of quantifiers, that is logical words, like *all*, *every*, *none*, *many* and the like. Furthermore, semantics has gained a lot of insight into the phenomenon of negative polarity (Fauconnier 1975, Ladusaw 1979, Zwarts 1981). Negative polarity items characteristically occur in negative contexts, while being banned from

¹ The author would like to thank Willem Visser and Anne Dykstra for useful comments.

affirmative sentences. Their distribution may be very complex. For example, the Dutch negative polarity item *ook maar iets* 'anything' is allowed in clauses introduced by the complementiser *voordat* 'before', but not in clauses introduced by the complementiser *nadat* 'after'. This is especially important for learners' dictionaries.

Despite the advances being made in semantics, the question has not been addressed in lexicography whether new insights from that field of linguistic inquiry have consequences for the way in which semantic information in dictionaries is structured, or, more specifically, for the way in which information about a negatively polar word or expressions is presented. One could imagine that examples are given of a negative polarity item for every semantic context in which semanticists have established that it is allowed to occur. Thus, the entry for *iets* should contain the collocation *ook maar iets*, and the examples should specify the contrast between *voordat* 'before' and *nadat* 'after'.

When a dictionary concerns itself with pronunciation, it is actually concerned with the field of phonetics and phonology. Developments in this field will not generally have consequences for the phonemic representations of words which a dictionary may provide. However, a better understanding of syllabification led to more accurate phonetic representations in the case of the Frisian dictionary of Duijff; Van der Kuip; De Haan; Sijens (2008). Furthermore, technological developments may make it possible to include sound files to digital dictionaries containing a recording of the word's actual pronunciation.

A word's substructure with respect to derivational morphology is given in some dictionaries. Yet it would be interesting to any user to add to, for example, a word like *construction*, the information that it contains the suffix *-ion*, and to add to the dictionary an entry *-ion*, which mentions that the suffix produces nouns, and that it is usually added to verbs such as *destruct*. After all, we have identified the dictionary as an encyclopaedia of linguistic knowledge about words, and information such as just mentioned should not be missing from dictionaries. As far as developments in the field of derivational morphology are concerned, it does not seem likely that they have consequences for the way in which derivational morphological information could be represented in dictionaries. However, it is fair to conclude that distinguishing affixes in words should be part and parcel of any dictionary.

Information about words is condensed by bringing words belonging to one inflectional paradigm together under one dictionary article, as members of the same inflectional paradigm tend to behave the same with respect to the various types of information that were listed above. Irregular forms are generally listed, as is the case for exceptional behaviour of words compared with other members of the inflectional paradigm. Advances in inflectional morphology seem unlikely to affect the structuring of information in dictionaries. Nevertheless, there is a minor point to be addressed here. It is well-known that full pronouns are regularly reduced in spoken language, a process that is sometimes dependent on syntactic position. This process of reduction may be phonologically regular (Dutch *jij* > *je* 'you'), *wij* > *we* 'we'), but it is often irregular (Frisian *hy* > *er* 'he', Dutch *hij* > *ie* 'he'). However, dictionaries do not regularly list this type of variation, as should be the case, since the words mean the same and they tend to be highly frequent. Furthermore, dictionaries should also specify whether the reduced forms may or may not occur in sentence-initial position; for example, Dutch *ie* and Frisian *er*, mentioned above, are banned from the sentence-initial position.

The study of syntactic categories has led to the discovery of many differences between members of categories such as adverbs, determiners or quantifiers. It should be investigated whether these discoveries are of interest for laymen, and whether dictionary practice needs to be updated in this respect. In some cases, articles in translation dictionaries can almost certainly profit from the new insights that have been gained in this field. My point is that such an evaluation of new insights does not seem to take place in lexicography.

A lot of research has been done on idiomatic structures and collocations, and part of this research has been done by lexicographers on the basis of corpus investigation (Sinclair 2004). Collocations belong to the discipline of syntax. It is known that idioms may partially or wholly freeze syntactic structure (Cutler 1982 and subsequent work). According to Siepmann (2005: 430), collocational phenomena span the entire range of morpho-syntactic constructions. However, recent advances in syntax do not seem to have consequences for the structuring of dictionary information: they are mostly of interest to the syntactic theoretician only. On the other hand, collocations are also an important part of departure for the study of semantics, see for example Moon (2008) who (2008: 246) quotes Sinclair (2004: 148) saying that the phrase is a better starting point for the description of meaning than the word (also Otani 2005: 10ff). This may have consequences for structuring the information about collocations in dictionaries.

To sum, it is important that lexicography stays in touch with the advances that are made in the disciplines of phonology, morphology, syntax and semantics as these disciplines may provide tools for structuring the encyclopaedic information about words and collocations that is presented to the laymen who are the primary target group of dictionaries. Furthermore, the discoveries themselves may be of interest to laymen and deserve a place in the dictionary conceived of as an encyclopaedia of linguistic knowledge. For example, the Frisian form *datst* 'that-2SG' is nowadays analysed as a form of complementiser agreement. Listings of agreeing complementisers are both interesting and useful for the user who, besides, often doesn't know how to spell them, and looks in vain in the dictionary to find the correct spelling. For example, Zantema (1992:180) does not specify in the article for *dy't* 'which' how the complementiser agreement for the second person singular should be spelled (*dyst*).

In what follows, I will focus on a specific problem involving the way in which dictionaries treat function words.

2. The problem of dealing with function words in dictionaries

Most space in a dictionary is devoted to semantics, more specifically, to paraphrases of the meaning of words. These paraphrases are couched in natural language, not in the sort of logic which semanticists often use. The reason for this is partly, as mentioned earlier, that dictionaries by and large intend to be accessible to laymen. Apart from that, it would not be insightful to try and define content words with, say, the help of lambda-abstraction.

Nevertheless, it is a shortcoming of dictionaries that a paraphrase of the meaning of function words is often not very insightful (Bergenholtz 1985: 236ff, Coffey 2006: 159ff). The meaning of function words like *the* hardly lends itself to an insightful paraphrase which teaches the user anything. Dictionary makers nevertheless present such paraphrases in the interest of consistency (Coffey 2006: 162). Furthermore, dictionary makers present a fine-grained polysemous analysis of such function words, which is totally uninformative (Coffey 2006: 164ff). Coffey proposes in his article to do away with superfluous polysemous detail in

the case of function words, as that sort of information is not of interest to the user. After all, dictionaries also don't contain fine-grained phonetic or syntactic information. In order to present relevant information, Coffey proposes to present more information about the use of functional words in collocations, information which can usually be listed under the content word that is part of the collocation (if any), and otherwise under the most salient function word.

While agreeing with Coffey's proposal, I would like to add two other methods by which useful information of a general nature can be added to a dictionary, conceived of as an encyclopaedia of words. Although certain function words do not lend themselves to a semantic, polysemous analysis that is useful for the general audience, they may be amenable to a syntactic analysis that is clear and useful for the general audience. A case in point is the Dutch word *er*, which is related to English *there*, although it behaves differently in many respects.

Some Dutch dictionaries muddle the description of the various uses of *er* (see Hoekstra 2000), ignoring the distinctions drawn by de *Algemene Nederlandse Spraakkunst*, the Standard Dutch Grammar (Haeseryn; Romijn; Geerts; De Rooij; Van den Toorn 1997). These distinctions are practical and well-motivated. The ANS distinguishes four uses of *er*, which are summarised below:

1. Locative adverb

Hij woont er sinds 2001
he lives there since 2001
 'He has lived there since 2001.'

2. Prepositional complement (possibly discontinuous)

Ik zal er met hem over spreken
I shall it with him about talk
 'I will talk with him about it.'

3. Presentational function with indefinite subjects (possibly discontinuous)

Er danst iemand
there dances somebody
 'There's somebody dancing.'

4. Presentational function with indefinite objects without nominal head (possibly discontinuous)

Ze hebben er toen drie verkocht
they have there then three sold
 'They sold three at that time.'

(Compare:

Ze hebben toen drie boeken verkocht
they have then three books sold
 'They sold three books at that time.')

The syntactic distinctions will be understandable to those users who have had a higher education, but the example sentences can be generally understood. It is not the case that adding grammatical information in dictionaries will automatically lead to dictionary articles that are difficult to understand (contra Schumacher 1985: 174): just selecting relevant examples will not make the dictionary more complex for the average user.

The syntactic distinctions drawn above will also be useful to foreign learners of Dutch. Native speakers of English learning Dutch will immediately see that the first and the third use of *er* is

highly similar to the English use of *there*, whereas the other two usages are distinct from English *there*. Thus the dictionary will be informative in the way in which an encyclopaedic work intended for the general audience should be.

Syntactic distinctions are often hard to explain to the layman, because they involve too much syntactic jargon. In that case, it is nevertheless possible to *illustrate* the relevant distinctions by means of appropriately chosen example sentences. The importance of using appropriate examples should not be underestimated (Heath 1985: 340, Ickler 1985: 376). Such example sentences will be intuitively grasped by laymen, while professional users will grasp their relevance. Two examples will be given. Dictionary articles for modals can include examples illustrating the Infinitivus-pro-Participio Effect, which is very frequent in Dutch:

Hij	heeft	het	niet	gewild	
<i>he</i>	<i>has</i>	<i>it</i>	<i>not</i>	<i>wanted</i>	<i>(PAST PTC)</i>

‘He didn’t intend it (to happen).’

Hij	heeft	het	niet	willen	doen
<i>he</i>	<i>has</i>	<i>it</i>	<i>not</i>	<i>want</i>	<i>(INF)</i>

‘He didn’t want to do it.’

This phenomenon entails that an infinitive shows up where a past participle would otherwise be expected. This phenomenon takes place in case the relevant verb is not the main verb. The phenomenon need not be explained. It can just be *shown* in the example sentences in the dictionary article for the modal *willen* ‘want’. This has been done for example in the very informative dictionary of Visser; Dyk (2002), where example sentences are used in the articles on modal verbs showing that the Infinitivus-pro-Participio Effect is absent.

Similarly, consider question formation in English. The dictionary article for the English complementiser *that* can include examples with question words (on these, see Chomsky 1977):

Who do you think came?
 Who do you think (that) I saw?

With a subject question, no *that* is allowed. *That* is allowed with an object question. Such facts can be presented in a dictionary by just presenting appropriate example sentences.

It is my proposal to systematically use syntactic knowledge to structure articles about function words. Acceptance of this proposal implies that the encyclopaedic character of dictionaries will be reinforced.

References

- Bergenholtz, H. (1985). 'Vom wissenschaftlichen Wörterbuch zum Lernerwörterbuch'. In Bergenholz, H.; Mugdan, J. (eds). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28-30.6.1984*. Tübingen: Max Niemeyer Verlag. 225-256.
- Bergenholtz, H.; Mugdan, J. (1985). 'Vorwort'. In Bergenholz, H.; Mugdan, J. (eds.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28-30.6.1984*. Tübingen: Max Niemeyer Verlag. 7-19.
- Chomsky, N. (1977). 'On Wh-Movement'. In P. Culicover; T. Wasow; A. Akmajian (eds.). *Formal Syntax*. New York: Academic Press. 71-132.
- Coffey, S. (2006). 'High-frequency grammatical lexis in advanced-level English learners' dictionaries: from language description to pedagogical usefulness'. In *International Journal of Lexicography* 19 (2). 157-173.
- Cutler A. (1982). 'Idioms: the older the colder'. In *Linguistic Inquiry* 13. 317-320.
- Duijff, P.; Van der Kuip, F.; De Haan, R.; Sijens, H. (2008). *Frysk Hânwurdboek*. Ljouwert: Fryske Akademy, AFUK.
- Fauconnier, G. (1975). 'Pragmatic scales and logical structure'. In *Linguistic Inquiry* 6. 353-375
- Haeseryn, W.; K. Romijn; G. Geerts; J. de Rooij; M.C. van den Toorn (1997). *Algemene Nederlandse Spraakkunst (ANS)*. Groningen: Martinus Nijhoff.
- Heath, D. (1985). 'Grammatische Angaben in Lernwörterbüchern des Englischen'. In Bergenholz, H.; Mugdan, J. (eds.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28-30.6.1984*. Tübingen: Max Niemeyer Verlag. 332-345.
- Hoekstra, E. (2000). 'ER in de dikke Van Dale: over functiewoorden in het woordenboek' [on line in Internet magazine Trefwoord] <http://www.fa.knaw.nl/fa/uitgaven/trefwoord/jaargang-2000/jg2000/er-in-de-dikke-van-dale>. [Access date: 9-6-2000].
- Ickler, T. (1985). 'Valenz und Bedeutung. Beobachtungen zur Lexikographie des Deutschen als Fremdsprache'. In Bergenholz, H.; Mugdan, J. (eds.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28-30.6.1984*. Tübingen: Max Niemeyer Verlag. 358-377.
- Ladusaw, William (1979). *Polarity sensitivity as inherent scope relations*. Ph.D. Dissertation, Austin, University of Texas (Bloomington: Indiana University Linguistic Club, 1980).
- Moon, R. (2008). 'Sinclair, phraseology and lexicography'. In *International Journal of Lexicography* 21 (3). 243-254.
- Otani, H. (2005). 'Investigating intercollocations - towards an archeology of text'. In *International Journal of Lexicography* 18 (1). 1-24.
- Schumacher, H. (1985). 'Grammatik in semantisch orientierten Valenzwörterbuch'. In Bergenholz, H.; Mugdan, J. (eds.). *Lexikographie und Grammatik. Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28-30.6.1984*. Tübingen: Max Niemeyer Verlag. 159-174.
- Siepmann, (2005). 'Collocation, colligation and encoding dictionaries, part I: lexicological aspects'. In *International Journal of Lexicography* 18 (4). 409-443.
- Sinclair, J. (2004). *Trust the text*. Routledge, London.
- Visser, W.; S. Dyk (2002). *Eilander Wezenbúek. Woordenboek van het Schiermonnikoogs*. Ljouwert/Leeuwarden: Fryske Akademy.
- Zantema, J.W. (1992). *Frysk Wurdboek Frysk-Nederlânsk*. Drachten / Ljouwert: A.J. Osinga.
- Zwarts, Frans 1981. 'Negatief-polaire uitdrukkingen I'. In *Glott Tijdschrift voor Taalwetenschap* 6. 35-132.