

Johan De Caluwe

Dictionary entries as windows on the onomasiological aspects of word formation

1. Introduction

The theme of this volume of papers is *Word formation in electronic dictionaries* and in fact this reflects the perspective of linguists/lexicographers on the subject. The average user of the dictionary is likely to be more interested in words and meanings, not in word formation as such. If we try to combine the perspective of both linguist/lexicographer on the one hand, and user of the dictionary on the other hand, I would reformulate the theme of this volume as: What, if any information on word formation should we build into the make-up of the dictionary (lexicographic perspective) in order for the user of that dictionary to find what he may be looking for (user perspective)?

As a consequence we cannot but start from the needs of the user. We all know that users may have either semasiological or onomasiological needs, or – again, in the language of the user:

- a) What does word *X* mean?
- b) Which word could best express meaning/concept ‘*Y*’?

I will concentrate on word formation in monolingual dictionaries, since this context is most challenging to lexicographers, especially from the onomasiological perspective.¹ When for example a language learner is looking in a monolingual dictionary of that language for a particular lexical item that would best serve his/her naming needs, there is no mediating term from his/her mother tongue to start from, as would be the case in a bilingual dictionary.

Throughout this paper I will use examples from Dutch from one particular paradigm. I will focus on

- 1) adjectives that can be used to characterise people,
- 2) nouns naming people with these characteristics,
- 3) the relation between these adjectives 1) and nouns 2),

¹ Cf. Koefoed (1991, 1993) for his pioneering work on the relation between word formation and onomasiology.

- 4) the question whether information of type 3) should be encoded in the e-dictionary, and if so, how?

In Dutch we can distinguish between the following word formation patterns for naming people on the basis of a typical (adjectival) feature:

- | | | | |
|-----|--|---|-----------------------------------|
| (1) | [A + <i>eling</i>] _N | | |
| | <i>woest</i> ‘savage, wild’ | > | <i>woest-eling</i> |
| | <i>creatief</i> ‘creative’ | > | <i>creatiev²-eling</i> |
| (2) | [A] _N (conversion) | | |
| | <i>revolutionair</i> ‘revolutionary’ | > | <i>revolutionair</i> |
| | <i>conservatief</i> ‘conservative’ | > | <i>conservatief</i> |
| (3) | [A + <i>e</i>] _N | | |
| | <i>bejaard</i> ‘elderly’ | > | <i>bejaard-e</i> |
| | <i>blank</i> ‘white’ | > | <i>blank-e</i> |
| (4) | [A + <i>erd/aard</i>] _N | | |
| | <i>eng</i> ‘creepy, scary’ | > | <i>eng-erd</i> |
| | <i>laf</i> ‘cowardly’ | > | <i>laf-aard</i> |
| (5) | [A + <i>erik</i>] _N | | |
| | <i>lui</i> ‘lazy’ | > | <i>lui-erik</i> |
| | <i>onbeleefd</i> ‘impolite, rude’ | > | <i>onbeleefd-erik</i> |
| (6) | [A + <i>je</i>] _N (diminutive) | | |
| | <i>blond</i> ‘blonde’ | > | <i>blond-je</i> |
| | <i>oud</i> ‘old’ | > | <i>oud-je</i> |
| (7) | [A + <i>o</i>] _N | | |
| | <i>positief</i> ‘positive’ | > | <i>positiv-o</i> |
| | <i>lesbisch</i> ‘lesbian’ | > | <i>lesb³-o</i> |

So Dutch has at least seven different patterns – be it with varying productivity – operating in one particular onomasiological domain, each with its own morphological, phonological, lexical, semantic restrictions on the base words. Consequently there are complex relations of competition and/or complementarity between these word formation rules, and even between these morphological naming strategies and „phrasal names“ (De Caluwe 1990, 2010; Booij 2010, p. 169ff.) such as *grote baas* (‘big boss’).

² The alternation between [f] and [v] is a phonological matter, but in this paper I will not go into the phonological changes involved in the word formation patterns under discussion.

³ The *o*-suffix is not added to the base adjective, but rather replaces the original adjectival suffix *-isch*.

The information on the hundreds of word formation rules in a language like Dutch or German is too complex for an exhaustive treatment in dictionaries. Therefore people with a primary interest in (Dutch) word formation *as such* should better consult a Dutch grammar like the *ANS* (1997) (*Algemeene Nederlandse Spraakkunst* – ‘General Dutch Grammar’), Booij’s *Morphology of Dutch* (2002), or the *Morfologisch Handboek van het Nederlands* (‘Morphological Handbook of Dutch’) (De Haas/Trommelen 1993).

In the next sections I will try to answer the question: if it is impossible to give all relevant data on word formation in a dictionary, then how do we select the data that *are* lexicographically relevant, and how do we present them in an electronic dictionary?

2. Information on the morphological make-up of words

2.1 Specifying the latest morphological operation

If we focus on the receptive use of the dictionary, i.e. on the semasiological needs of the user, we have in mind for example learners of Dutch as a second language, looking up *woesteling*, *revolutionair*, *bejaarde*, *lafaard*, *luierik*, *oudje*, etc.

Should we give the users more than the meaning of these words? Should we offer them the morphological make-up of these words as specified in (1)-(7)?

The answer in this case is: yes. It can be done at a relatively small cost, at least if word formation is a simple process of concatenation, as with *bejaard-e* or *onbeleefd-erik*.

For each derived word that is listed as a separate entry in a (paper-based or electronic) dictionary, it would be useful to have its morphology explained in detail, to the extent that it is grammatically and practically possible. This does not mean that each derivation should be analysed to its deepest level, but that the most recent step taken in the morphological history of that word be made clear to the dictionary user.

In the case of *bejaarde* ‘elderly’ the most recent morphological operation is suffixation of the adjective with *-e*: *bejaard-e*, and the same holds for many more words in Dutch: *blank-e* ‘white’, *blind-e* ‘blind’, *volwassen-e* ‘adult’, etc.

In the case of the adjective *bejaard* the most recent operation is circumfixation with *be-* and *-d*: *be-jaar_N-d*, meaning: ‘with (many) N’. The same holds for

- *be-wolk-t* ‘with (many) clouds’,
- *be-middel-d* ‘with (many) means’, i.e. ‘rich’,
- *be-vooroordeel-d* ‘with (many) prejudices’.

The specification of the morphological make-up of complex words will help the users of the dictionary if they want more information on the systematic meaning of that word. With *bejaard-e* for example, they can click on the suffix *-e* to read more about that particular word formation pattern, and to find more words that have been formed in a similar way. This will help the reader to situate a word and its set of meanings – be they idiosyncratic or not – within a paradigm of morphologically related words, which can allow the reader to gain insight into how the consulted word developed its meaning. In the case of *bejaard* for example, readers will discover that its present meaning ‘elderly’ relates to the basic meaning of ‘having (many/much) N’, which it shares with *bewolkt*, *bemiddeld*, *bevooroordeeld*.

The added advantage of morphological marking in an electronic dictionary is that it allows us to easily identify all words of the same morphological type in the dictionary (De Caluwe/Taeldeman 2003, p. 121f.). As we all know, it is of little use to search for **e* in an electronic dictionary if we want to identify all nouns of the type *bejaard-e*, since the search engine would also generate *ziekte* ‘disease’, *aardolie* ‘petrol’, and thousands of other words ending in *e*, but not the suffix *-e*.

2.2 Technical problems

Specifying the morphological make-up is sometimes easier said than done. We are aware of the technical problems that will arise if you want to enrich the dictionary with information on the derivational history of complex items. Lexicographers should devise technical solutions for problems that are typical for either concatenative or non-concatenative word formation. Under (8) we just mention a few different types of problems to overcome:

- (8a) *domm-erik_N* ‘stupid person’

Orthographic: the user may think the base word is *domm*, but it is *dom*. In *dommerik* the consonant is doubled to indicate that the preceding vowel is short.

- (8b) *creatiev-eling_N* ‘creative person’

Phonological: the user may think the base word is *creatiev*, but it is *creatief*, due to the process of devoicing of the final consonant in the adjective.

- (8c) *il-legaal*_A ‘illegal’
Phonological: in all cases the prefix *in-* is involved, but that is obscured by the different types of regressive assimilation.
- (8d) *revolutionair*_{AN} ‘revolutionary’
Morphological: how to specify that *revolutionair*_N is derived from *revolution-air*_A by conversion?
- (8e) *social-ist*_N
Morphological: how to specify that *socialist*_N is not derived from the adjective *social*, but is the product of affix substitution on the base word *socialisme*_N, as is the case with *realist*, *naturalist*, *impressionist*, etc.
Mutatis mutandis the same question holds for cases of back-formation (*Rückbildung*): *hongerstaker*_N ‘hunger striker’ is no simple deverbal noun; its base is another noun *hongerstaking* (‘hunger strike’).

2.3 Giving the users more than they were looking for

Specifying the morphological make-up of a word in a strict sense does not correspond to the principles we started from at the beginning of this paper. The user of the dictionary, looking up *dommerik*, *illegaal*, etc. is most likely to be only/primarily interested in the meaning of these words and perhaps also the way they can be used in an idiomatically correct way.

And yet there is a good reason for lexicographers to specify the morphological make-up of lexical entries since it adds greatly to the practical value of the dictionary. As ten Hacken/Abel/Knapp (2006, p. 243) put it:

In second language acquisition, word formation is important for the decoding of words the learner does not know, for the production of regular new words when the learner has not acquired the standard word, and for the creation of a tighter network structure in the mental lexicon, which facilitates vocabulary acquisition.

The quotation reflects ten Hacken/Abel/Knapp's focus on (electronic) learners' dictionaries in the context of language acquisition, but mutatis mutandis the same argument holds for native speakers consulting a monolingual dictionary. Providing the user looking for the meaning of a word with information on the paradigmatic, in casu morphological relations, of that word with other items in the lexicon really constitutes an added value to the user, on the condition of course that it will not lead to *information stress*.⁴ The way in

⁴ On information stress, see also Henning Bergenholtz in his article „Wortbildungsangaben als Hilfe für den Zugriff auf andere Datentypen und als Hilfe bei kommunikativen und kognitiven Informationsbedürfnissen“ (this volume).

which the morphological information is presented should reflect its secondary status, while semantic information should by all means keep its primary status on the screen.⁵

Specifying the morphological make-up of a word also raises another question: where does information on word formation end and where does etymology start? Words may have been derived from a particular base many hundreds of years ago, and the relation to that base word may have long been lost in the mind of native speakers. For example:

- (9) Du. *bepalen*_v ‘to determine’
 verb stem: *be-paal*
 base word: *paal* ‘post, stake, pole’
 word formation type: [*be + N*]_v: ‘to provide with N’
 semantic evolution:
 ‘to put stakes in the ground to stake out a piece of land’
 > ‘to demarcate’
 > ‘to determine’

This word formation type is still productive in Dutch, which may be a good argument to specify the morphological make-up of this type of words too. It will help the user of the dictionary to create a tighter network structure in the mental lexicon, facilitating vocabulary acquisition in the case of language learning. The unproductivity of the word formation type involved is, however, no sufficient ground for not specifying morphological make-up. In Dutch for example, the following pattern has become unproductive now:

- (10) [*N + in*]_N: ‘female N’
vriend-in ‘girl friend’
koning-in ‘queen’
leeuw-in ‘lioness’

Yet, the morphological make-up of these complex words is still highly transparent to all native speakers of Dutch, which is a good reason to specify this make-up in the dictionary.

On the other hand, there will be cases where unproductive word formation types are involved, resulting in complex items that are no longer transparent

⁵ This raises the general question of course which information can count as added value, and where it should be specified. For example, some dictionaries add data on the etymology of a word. In the Dutch dictionary series of Van Dale – the most comprehensive and the most popular dictionary in both Belgium and the Netherlands – etymological information used to be specified immediately after the details on the word form, so before the definition of the word sense. The editors have recently decided to move this secondary type of information to the end of the dictionary entry, i.e. after the description of the semantics (cf. Van Dale 2005).

even to native speakers. Specifying the morphological make-up of these words would perhaps be a step too far in a synchronic (learners') dictionary, but of course would be welcome in a historic dictionary.

3. Information on derivations and compounds with a particular base word

3.1 Motivation

Until now we discussed information on the morphological make-up of a word, but how about the complementary question: should the e-dictionary specify for each lexical item the derivations and compounds that have been formed on the basis of that lexical item?

For example, returning to the examples we discussed earlier, there are good reasons to specify the morphological make-up of Du. *bejaarde* as *bejaard-e*, or of *bejaard* as *be-jaar-d*. But should we specify in the dictionary entry for *bejaard*_A that Dutch has a deadjectival noun *bejaard-e* 'elderly person'? And is it interesting for the reader looking up *jaar* 'year' to learn that there is a derivation in Dutch *be-jaar-d*, meaning 'elderly'? And shall we tell the reader in the entry for *revolutionair*_A that the word form can also be used as a name for people who are revolutionary (by means of conversion)?

One option is to list all derivations and compounds with any base word in the dictionary. This option is only worth considering if the format of the e-dictionary will enable the user to avoid information stress. So it would certainly be wise to present all compounds and derivations on secondary screens only if the user decides to use a particular link on the main screen.

In chapter 2.3 we raised the question whether we should specify the morphological make-up of words, if that make-up is of no particular synchronic relevance any more. The same question holds in the opposite direction: should we specify in the entry for *paal*_N, as in (9) above, that the verb *bepalen* has been derived from it? And how about the relation between the words in (11)?

- | | | | |
|------|---|---|---|
| (11) | Du. <i>vliegen</i> _V 'to fly' | > | <i>vlieg</i> _N 'fly' |
| | Du. <i>boek</i> _N 'book' | > | <i>boeken</i> _V 'to book' |
| | Du. <i>hond</i> _N 'dog' | > | <i>zeehond</i> _N lit. sea-dog/'seal' |
| | Du. <i>haven</i> _N 'port, harbour' | > | <i>luchthaven</i> _N 'airport' |

Shall we specify in the lexical entry for the Du. verb *vliegen* that the name for a fly (*vlieg*) has been derived from it by conversion of the verb stem? And is it relevant for the reader consulting the dictionary for Du. *hond* 'dog' to learn that the name for a seal (*zeehond*) is based on that word?

We would suggest to list all these derivations and compounds in a secondary screen, but particular words should be brought under the attention of the user more prominently than others because they may be more relevant to the user.

3.2 Selection on onomasiological grounds

Which word formations deserve to be highlighted? One option is to only list or to highlight those word formations that reach a particular level of frequency in the relevant corpora of a language. This option should be considered with due caution since the most frequent word formations may not always be the most relevant for a particular base word. For example: *luchthaven* ‘airport’ is very frequent in Dutch, but is that relevant when lexicographers have to decide which complex words to highlight in the lexical entry for *haven* ‘harbour’?

We would propose to list and/or highlight not the most current word formations with a particular base word, but rather those words that stand in a current onomasiological relation to that base word, even if these words are morphologically unrelated to that base word. To illustrate with an example under (12):

- (12) Du. *vliegen*_V ‘to fly’ > *vlieg*_N ‘fly’
*vlieg-angst*_N ‘fear of flying’
*vlieg-enier*_N ‘flier, airman’
*vlieg-er*_N ‘flier / kite’
*vlieg-ramp*_N ‘plane crash’
*vlieg-tuig*_N ‘aeroplane’
 ...
 PILOOT_N ?

Many words have been derived and compounded on the basis of the Dutch verb *vliegen* – see the right hand column under (12) – and there is nothing wrong with listing these formations in the dictionary in the lexical entry for that verb. However, a word like *piloot* ‘pilot’ is of much more onomasiological relevance to the user and should consequently be much more prominent in the window of related words than the name for the well-known flying insect.

3.3 Implementation

With reference to the *Algemeen Nederlands Woordenboek (ANW)*⁶ – ‘General Dutch Dictionary’, a combined semasiological and onomasiological dictionary being compiled at the Institute for Dutch Lexicology in Leiden, Van Sterkenburg (2003) states:

⁶ <http://anw.inl.nl>

By improving existing dictionaries we will probably never succeed in making the electronic monolingual dictionary impeccably onomasiological. That can only be achieved if a new electronic dictionary is compiled [...] (Van Sterkenburg 2003, p. 138)

In other words: we have to leave traditional ideas of dictionary making behind us, if necessary. If we apply the onomasiological principle outlined in 3.2 to the type of words I have been focussing on in (1)-(7), viz. adjectives naming human characteristics, the picture could look as follows:

- (13) *be-jaar-d* 'elderly'
> [bejaarde man/vrouw] 'elderly man/woman':
a. *bejaard-e*
b. *ouder-e*
c. *oud-je*
d. *senior*
e. *ouden-van-dagen*

The dictionary entry for Du. *bejaard* would of course first specify its morphological make-up: *be-jaar-d*. In addition one could simply list *bejaard-e* as the only derivation from this adjective. But then the reader would miss the relevant information of the status of that derivation in its onomasiological paradigm. Listing all names in Dutch for elderly people, even if they are morphologically unrelated to the base word, as in b.-e. in (13), seems to be a better solution. By clicking on each of these words under a.-e. in the e-dictionary the user would then generate a comprehensive definition of the items listed, and of course information on usage of these items, compared to their (near) synonyms. The user would in this case learn from the dictionary

- 1) that Dutch has names for elderly people which are gender neutral: there are several words to be used for both elderly men and women;
- 2) that derivation with *-e* is possible, both with the base adjective *bejaard* (a.) and with the base adjective *oud* in its comparative form *oud-er* 'older' (b.);
- 3) that there is an alternative derivation with the suffix *-je* (c.), but clicking on the suffix would inform the user of the patronizing, belittling connotation of names on *-je*;
- 4) that *senior* (both with Dutch and English pronunciation) is the new euphemism about to replace all the other names, as is attested by many new compounds or expressions with *senior*, often replacing the older equiva-

lents with *bejaarden-*: *seniorenpas*, *seniorenwoning*, *seniorenraad*, *seniorenflat*, etc. ('pass/house/council/flat for elderly people').⁷

- 5) that *ouden-van-dagen* (lit. 'old-of-days') really is corny now, and is absolutely 'out'.

By listing all these possible names for elderly people with the adjective *bejaard* 'elderly', the user will learn that *bejaarde* is part of a constellation of terms in which each has its own history, its own connotation, and that each term is subject to particular conditions of use.

In (14) we give two more examples of a combined morphological/onomasiological approach to word formation in electronic dictionaries:

- (14) *optimist-isch* 'optimistic'
 > [optimistische man/vrouw] 'optimistic man/woman':
 a. *optimist*
 b. *positiv-o*
- pessimist-isch* 'pessimistic'
 > [pessimistische man/vrouw] 'pessimistic man/woman':
 a. *pessimist*
 b. *zwartkijker* (lit. black-watch-er)
 c. *hypochonder*

This is a special case as the adjectives – *optimistisch* and *pessimistisch* – may have been derived from the name for the person with that particular characteristic feature – *optimist/pessimist*, or it may be a matter of paradigmatic morphology whereby (also) the nouns *optimism/pessimism* are involved. If a dictionary would only specify which derivations and compounds have been formed with either *optimistisch* or *pessimistisch* as a base word, nothing would have to be specified in this case, since there are no formations with these base words. From an onomasiological perspective on word formation on the other hand, there is every reason to list the words that are used to name the concept of 'optimistic/pessimistic man/woman' even if they are not directly derived from the adjectives under consideration, such as *optimist*, *positivo*, etc.

Positivo illustrates a relatively new word formation pattern in Dutch; it is used to create informal, trendy words for people, sometimes a little derogatory, as in this case: 'someone who will always look at the bright side of life, even if there is hardly anything bright to it ...'.

⁷ *Senior* is so popular now because it's the only word without (as yet) the connotation of 'little, old people who can't help themselves any more, and who are just good enough for living a passive life in an old people's home'.

For the notion ‘pessimistic man/woman’, apart from *pessimist*, there are two onomasiological alternatives, one synthetic compound⁸ (*zwartkijker*), and one formed with classical ingredients (*hypochonder*).

4. Conclusion

Of course it is an interesting aspect of electronic dictionaries that so much information on word formation can be specified with any lexical item. So of course it may be interesting to add derivations and compounds of a base word to the lexical entry of that base word. But it would be even more *relevant* if this information is embedded in a wider onomasiological context, providing the user with all naming alternatives even if – as is often the case – these alternatives are not morphologically related to that base word. By complementing the most obvious derivations (e.g. *bejaard* > *bejaard-e*) with relevant naming alternatives (e.g. *senior*) the dictionary helps the learners of a language to acquire a vocabulary in a particular domain, in all its denotational and connotational nuance. In addition, it enriches the insight of the native speaker into the structure and dynamics of the lexicon.

Before we can implement this onomasiological principle in electronic dictionaries, decisions will have to be made on both (1) the selection of the onomasiological/morphological material to be incorporated in the dictionary, and (2) the structure for the presentation of this material. Research on the naming needs of native speakers and language learners will inform dictionary makers on the optimal make-up of their dictionary format, or formats, for both types of user groups may have quite different needs and expectations with respect to onomasiological and morphological information in dictionaries.

5. References

- ANS 1997 = Haeseryn, Walter/Romijn, Kirsten/Geerts, Guido/De Rooij, Jaap/Van den Toorn, Maarten (eds.) (1997): *Algemeene Nederlandse Spraakkunst*. Second, completely revised edition. Groningen/Deurne.
- Booij, Geert (2002): *The morphology of Dutch*. Oxford.
- Booij, Geert (2010): *Construction morphology*. Oxford.
- Booij, Geert/Van Santen, Ariane (1998): *Morfologie. De woordstructuur van het Nederlands*. Second, completely revised edition. Amsterdam.
- De Caluwe, Johan (1990): Complementariteit tussen morfologische en in oorsprong syntactische benoemingsprocedures. In: De Caluwe, Johan: *Betekenis en produktiviteit*. Gent, p. 9-23.

⁸ Cf. Booij/Van Santen (1998, p. 171ff.) on the notion of synthetic compound in Dutch.

- De Caluwe, Johan (2010): Woordvorming in onomasiologisch perspectief. Een case study uit het Nederlands. In: *Voortgang, jaarboek voor de nederlandistiek* 28, p. 71-91.
- De Caluwe, Johan/Taeldeman, Johan (2003): Morphology in dictionaries. In: Van Sterkenburg (ed.), p. 114-126.
- De Haas, Wim/Trommelen, Mieke (1993): *Morfologisch handboek van het Nederlands: een overzicht van de woordvorming*. 's-Gravenhage.
- Koefoed, Geert (1991): Morfologie en pragmatiek: produktiviteit en de act van benoeming. In: *Forum der Letteren* 32, p. 161-172.
- Koefoed, Geert (1993): *Benoemen. Een beschouwing over de faculté du langage*. Doctoral dissertation Utrecht University. Amsterdam.
- ten Hacken, Pius/Abel, Andrea/Knapp, Judith (2006): Word formation in an electronic learners' dictionary: ELGIT. In: *International Journal of Lexicography* 19, 3, p. 243-256.
- van Dale (2005) = Den Boon, Ton/Geeraerts, Dirk (eds.) (2005): *Van Dale Groot woordenboek van de Nederlandse taal*. Fourteenth, completely revised edition. Utrecht/Antwerpen.
- van Sterkenburg, Piet (2003): Onomasiological specifications and a concise history of onomasiological dictionaries. In: Van Sterkenburg (ed.), p. 127-143.
- van Sterkenburg, Piet (ed.) (2003): *A practical guide to lexicography*. Amsterdam/Philadelphia.