

Kernerman Dutch Dictionary

Truus Kruyt



Truus Kruyt worked on several lexicographical projects, among which are the historical dictionary *Woordenboek der Nederlandsche Taal WNT* (comparable with *Oxford English Dictionary* and *Grimm's Wörterbuch*), the Dutch *PAROLE/SIMPLE* computational lexicon and *Kernerman Dutch Dictionary*. She was head of the Language Database Department at the Institute for Dutch Lexicology in Leiden, and in this position has participated in many European projects in the field of language technology. She has lectured on Computer and Lexicon (Leiden University), Lexicography and Computer Technology (Ivanovo State University) and Dutch Linguistics (Westfälische Wilhelms-Universität Münster and Universität zu Köln), and is a member of national and international boards and scholarly committees. Since 2007 she has been working as a freelance linguist. truuss.kruyt@hotmail.com

Introduction

K Dictionaries (KD) has developed a series of monolingual learners' dictionaries for over twenty languages, which should serve as a basis for bilingual dictionaries. A monolingual Dutch dictionary (12,000 entries) was compiled in 2007, by the lexicographers Rik Schutz, Truus Kruyt, Monique Woltring and Geert Bilsen. The dictionary has been published on *Woorden.org*, which has a collection of 170,000 entries derived from various online lexicographic resources. It is accessible for free. The entries from our dictionary can be recognized by the sentence "*Bovenstaande informatie is afkomstig van Kernerman Dictionaries*" (the information above is originating from Kernerman Dictionaries). Another distinctive feature is the pronunciation in IPA¹ representation. The entries from the other sources are less detailed and sometimes concern outdated words and spellings. The aim for the coming years is to improve and to update these entries, to start with words that are most frequently retrieved by the users of *Woorden.org*. For our dictionary, this ambition will be achieved in 2011 by an extension with another 12,000 entries (see the final section).

The dictionary was translated into French and translation to English, German and Spanish are on-going. They are all available on Mijnwoordenboek.nl.

Target users and headword selection

The monolingual Dutch dictionary was aimed to constitute the basis for bilingual dictionaries with a twofold target group: adults with a different mother tongue who learn Dutch in their own country, and Dutch native speakers who need a concise bilingual dictionary. The 12,000 entries of the dictionary are considered to represent a Dutch basic vocabulary which enables the user to largely understand an elementary text.

Sources for the selection of headwords included a wordlist provided by KD and several Dutch language resources for learners, among which *Van Dale Pocketwoordenboek Nederlands als tweede taal (NT2)* (*Van Dale's concise dictionary of Dutch as a second language*)². Given the target group and the limited size of our dictionary, a major selection criterion was that the words to be included should frequently occur in Dutch daily life. For example, terminology,

less prevalent occupations, exotic animals and plants, rather unknown phenomena outside the Dutch-speaking regions, and non-naturalized international words and abbreviations were deliberately left out. Criteria for the distinction of homonyms were different parts of speech and different pronunciation, rather than etymology. The first two criteria are clear to learners, etymology is a too difficult matter.

The compilers were requested to critically consider the initial list of headwords, and to suggest entries to remove and to add in accordance with the target group. In this respect, texts on the Internet were often helpful, for example to check the current usage of words (printed dictionaries are inherently traditional) or as a source of inspiration for colloquial words to be incorporated in our dictionary.

Structure elements of a special nature

The entries contain usual structure elements, such as headword variants, pronunciation, grammatical information (part of speech, gender, plural, singular past, past participle etc), meaning descriptions, synonyms, antonyms, examples, collocations and fixed expressions (idioms). Here we will pay attention to some special features.

Pronunciation

As mentioned above, the pronunciation is represented by use of the International Phonetic Alphabet (IPA). Some sounds are not pronounced by all Dutch speakers, for example the *n* at the end of verbs (e.g. *eten* / to eat) and the *t* in a word like 'informatie' (information). We represented cases like these as follows: 'etə(n) and ɪnfər'mɑ(t)si (stressed syllable marked by a preceding ')'. This representation is in accordance with other Dutch reference works and with some bilingual dictionaries produced under the authority of the *Committee for International Lexicographical Resources*, subsidized by the *Nederlandse Taalunie (Dutch Language Union)*.³

Meaning description / definition

For the determination of word meanings, the Internet corpus proved to be very useful. Of course, the output of searches is to be critically judged by an experienced lexicographer. But apart from meanings already recorded in dictionaries, the Internet corpus revealed new meanings

and new meaning extensions. For example⁴, dictionaries describe *geheugencapaciteit* (a compound of *memory* and *capacity*) as a computer term only, whereas the Internet also reveals the meaning ‘how much or how long you can remember things’, in contexts like ‘de geheugencapaciteit van je hersenen verbeteren’ (to improve the *memory capacity* of your brain) and ‘Een goudvis heeft een geheugencapaciteit van drie seconden.’ (A goldfish has a *memory capacity* of three seconds.).

For the meaning descriptions we adopted a concise and colloquial style with wordings as simple as possible. Some examples: *schuld* (*debt*) is defined as ‘geldbedrag dat je nog moet betalen’ (sum that you still have to pay), *sekse* (*gender*) as ‘feit dat je een man of een vrouw bent’ (fact that you are a man or a woman), *understatement* as ‘uitspraak waarmee je iets (opzettelijk) zwakker uitdrukt dan het in werkelijkheid is’ (statement by which you (deliberately) express something weaker than it actually is). In fact, the ambition of defining this way was one of the most challenging ones in writing the dictionary. We avoided idiomatic expressions in the definition and the use of a synonym as the sole definition was only allowed for words for which a common alternative is available. Sometimes a paraphrase rather than a definition was given. For example, the pronoun *jullie* (plural *you*) is described as ‘je zegt dit woord als je tegen twee of meer mensen praat’ (you say this word when you speak to two or more people), and *aha* has the description ‘uitroep bij positieve gevoelens’ (exclamation going with positive feelings).⁵

Examples

There should be an example of usage (at least one) for each sense of polysemous entries. Monosemous entries have examples as well, as considered appropriate by the editors, and compositional phrases may also serve as examples in certain cases. The main function of the example is to demonstrate current lexical and grammatical collocations, frequent inflections or a characteristic grammatical construction. Furthermore, it serves to clarify the concise meaning description, in particular in the case of polysemous words for which the difference between meanings might be difficult to grasp for the target group. Examples can also give encyclopedic, cultural, social or other functional information to clarify the word’s meaning and usage. In general, the aim was to show the word in its natural context.

The example is preferably not a full sentence but rather a short phrase, or even a characteristic compound. For example,

‘een slagvaardige ondernemer’ (a decisive entrepreneur) (lemma underlined). We preferred examples that immediately come into mind hearing the word. In other cases, the Internet was helpful as a source of inspiration. We deliberately avoided the special features of many dictionary examples, such as overspecificity, use of the past tense, little relationship with the current social reality, etc.

Working method

Preparatory work (headword list, author’s instructions, customized XML editor, etc) was done in consultation with the coordinating team and technical staff at KD. The lexicographers worked at home and sent their files by e-mail to each other for a quality check and feedback. Although useful in itself, this supervision was needed due to the high time pressure. The coordinator was Rik Schutz, who also did much correction work on the files after the project.

Future perspectives

As mentioned in the introduction, our dictionary has so far been published online with other lexicographic resources, not primarily oriented towards learners. User statistics show that 10,000 out of our total of 12,000 entries were looked up in the first half year. This is a fairly high percentage given the basic character of the vocabulary.

In February 2011, a follow-up started, that is to be finished in autumn this year, extending the dictionary with another 12,000 entries. A main criterion for the selection of new headwords is frequency of retrieved words by the users of *Woorden.org*. These words often differ in nature from the headword list of the initial product. For example, the amount of intricate words and specialist terms appears to be much larger (e.g. *oormerken/earmark*, *synergie/synergy*, *concipiëren/conceive*, *verdisconteren/discount*, *agiol/premium*). It was also decided to extend the headword list with word categories that were left out in the first phase, for example abbreviations and acronyms, a broader range of international words in the Dutch language, such as *scouting*, *sealen* (*to seal*) and *schwung* (*dash*), words for phenomena outside the Dutch-speaking regions, such as *tsunami* and *gletsjer* (*glacier*), and names of exotic animals and plants. As a consequence, the target group has shifted towards a somewhat higher level than before. Still, the accessible style of the definitions and the example sentences is maintained. For example, *kubisme* (*cubism*) is defined as ‘kunststroming waarbij alle vormen met rechte lijnen worden weergegeven’ (trend in art that represents all shapes by straight lines).

The Dutch dictionary core is part of the new global series of K Dictionaries. It was compiled by Rik Schutz (chief editor), Truus Kruyt, Monique Woltring and Geert Bilsen, and is currently being revised and expanded by Rik Schutz, Truus Kruyt, Wilfried Dabekaussen and Hanne Bussels. It is used both in monolingual form and as a base for developing various bilingual versions.

New headwords have been inserted between the existing articles. The latter are being optimized during the compilation of the new headwords. In particular, many more synonyms and antonyms are added. This time of the compilation is assisted by data from Opentaal Foundation (including speech indicators, inflected forms of nouns and verbs, and word frequency information) and the Sketch Engine corpus of Dutch. The present lexicographers are Rik Schutz, Truus Kruyt, Wilfried Dabekaussen and Hanne Bussels.

Registered user needs may guide future extensions of the dictionary.

Notes

- 1 International Phonetic Alphabet.
- 2 Van Dale Lexicografie, Utrecht-Antwerpen. 2003, 2006.
- 3 <http://taalunieversum.org/taalunie/>.
- 4 Example from the current work.
- 5 Example from the current work.

Anthropological and linguistic fundamentals of lexicographic work

Miguel Eduardo Montoro

1. Introduction

I got the opportunity to work in dictionary compilation in quite an unusual way. Although I had studied lexicography, the reason it was included in my studies and research differed radically from the way I used it over the last six years, in which it has become my livelihood.

My biggest passion is philosophical research, and before I started to work on dictionaries I had the privilege of being guided by one of the great masters in philosophy, Andrea Di Maio (Professor at the Pontificia Università Gregoriana, Rome), who steered me towards lexicography.

The reason that philosophy led me to lexicography is that every thought uses words as a vehicle for expression, and to understand the thinking of a philosopher in a profound way requires understanding precisely the corpus of words he or she uses, their polysemy, and the total structure of his/her language understood as a semantic constellation of hierarchically inter-related words—where some emerge in importance over others and determine the total structure of his/her thought. Ultimately, to understand an author thoroughly implies, at least mentally, ‘to make a dictionary’ of the author.

To do this, we face the textual corpus of an author as if we were facing an unknown language for which we need a dictionary in order to understand what he or she is actually saying. We study the words used by the author statistically, make a word frequency list, select (by contrast with the common language) those terms considered as *keywords*, and try to understand the author’s ‘own language’, constructing a

semantic map with all word meanings that are used and their close inter-relationship. This is carried out by using particularly arduous techniques, an explanation of which is beside the point here, but which requires a microscopic study of words.

Such a study is so deep that, in fact, by studying a single keyword from a particular author we could write a full thesis, which leads to the paradox that the dictionary of an author can become a very rich series of such theses.

For example, in my case, I started to study the word *resolution* in St. Thomas Aquinas, and after completing a synchronic study about this word within the Thomistic corpus, my supervisor asked me for a brief diachronic introduction to the word as found in Aquinas. That is to say, I had to write the history of how various influences had contributed in the course of time to Aquinas’s particular semantic nuances. So it was not enough to analyse the term as a lexicographer does, but rather as a historian of word semantics. I started this arduous research following the historical course and the clues all led me to about 17 centuries before Aquinas, to Plato. Since Plato wrote in Greek, I obviously could not find the very same word in his work, but an equivalent, *διαλεκτική* (dialectic). As a result, what should have been a minor aspect of my thesis turned into a thesis in its own right and I ended up doing a lexicographic study of *διαλεκτική* from Plato.

This is my background relating to lexicography, which eventually I started calling *high lexicography*, in contrast to *hard lexicography*, that is, dictionary making,