The Woordenbank van de Nederlandse Dialecten (Wordbase of Dutch Dialects)

Jacques Van Keymeulen, Veronique De Tier

Ghent University
Jacques.vankeymeulen@ugent.be, Veronique.detier@ugent.be

Abstract

Three major regional dialect dictionaries, the *Dictionary of the Brabantic Dialects*, the *Dictionary of the Limburgian Dialects* and the *Dictionary of the Flemish Dialects* inventory the vocabulary of the southern Dutch dialects, i.e. the dialects spoken in Dutch-speaking Belgium, Northern Brabant and Limburg in the Netherlands and French-Flanders in France. The three dictionaries are onomasiologically arranged, according to the lexicographic ideas of A. Weijnen. Because of their arrangement, the dictionaries cannot convey detailed semantic information. They are to be considered atlases, rather than dictionaries. Therefore, in order to get a complete overview of the lexicon of the southern Dutch dialects, professional lexicography has to call in the help of 'amateur' lexicography, i.e. the regional and local dialect dictionaries, made by non-professional lexicographers.

In this presentation a project is described which aims at the creation of a digitized lexicographical database for the alphabetical amateur lexicography of the (southern) Dutch dialects, including both the old alphabetical tradition of the end of the 19th / beginning of the 20th century and the new tradition, rooted in the so-called dialect revival of the 70s and afterwards. The project is still in progress. First, we present the aim of the project; next we go into the details of the database structure and the search engines.

Keywords: database, dialect dictionaries, Dutch, Flanders

1. Introduction

The Dutch language has excellent digital dictionaries, thanks to the *Instituut voor Nederlandse Lexicologie*[1] (Institute for Dutch Lexicology) at Leiden University. The INL is financed by the *Nederlandse Taalunie* (Dutch Language Union), an organisation two-thirds subsidized by the Dutch government and one-third by the Flemish regional government. The main project of the INL was the *Woordenboek der Nederlandsche Taal* (WNT, Dictionary of the Dutch Language), a gigantic dictionary covering the modern Dutch period (i.e. 1500–1976) in 43 volumes (supplement included), totalling over 49,000 pages (95,000 main entries); it is the most comprehensive and detailed dictionary in the world. It has been digitized and is available in open access (see iWNT, the on-line *Woordenboek van de Nederlandse Taal* / Dictionary of the Dutch Language); in recent years other historical dictionaries have been linked to the WNT: the *Vroegmiddelnederlands Woordenboek* (VMNW,

---

[1] For more information on the INL, see http://www.inl.nl

Dictionary of Early Middle Dutch, 1999), the *Middelnederlands Woordenboek* (MNW, Middle Dutch Dictionary by Verwijs and Verdam) covering the 1200–1500 period and the *Oudnederlands Woordenboek* (ONW, Dictionary of Old Dutch, 2009). In 2011the Frisian Dictionary was added. When the ANW[2] is completed, all historical periods of the Dutch language will be covered.[3] All the aforementioned historical dictionaries are, for obvious reasons, based on written sources.

Many words, however, only survive orally in the traditional dialects. In this paper, we use the term *dialect* in its continental sense, i.e. as a term for a *geographically determined language variety*. We even use it in its narrowest – traditional and nowadays old-fashioned – sense: the geographically determined language variety typical for Trudgill's 'NORM'-informant, the Non-mobile Old Rural Male, a person who has as its urban counterpart the unskilled blue-collar factory worker. Traditional dialects are disappearing or have already disappeared, especially since the 60s of the last century, due to changes in the modern world: increase of social and geographical mobility, increase of schooling level and introduction of the mass media.

In what follows, we will discuss the dictionaries for the geographically differentiated vocabulary of traditional Dutch dialects. In section 2, we consider the dictionaries for oral language traditions, both the geographically oriented *onomasiological* ones (§2.1.) and the local/regional *semasiological* ones (§2.2.). It will be made clear that thematically arranged dictionaries should be supplemented with semasiologically arranged ones. The latter dictionaries should be brought together in a database: in section 3 an outline for such a database is presented with the project *Woordenbank van de Nederlandse Dialecten* (WND, Wordbase of the Dutch Dialects). Section 4 is devoted to the structure of the database; section 5 comprises the conclusion.

2. Oral language dictionaries: dialect lexicography[4]

Dialect vocabulary is mainly an oral vocabulary and it is geographically differentiated: both aspects should be documented. The first characteristic requests that the data are to be collected by way of fieldwork; the second that the fieldwork should be conducted in a place-to-place manner in order to be able to draw word maps. To this, one may add that the task is urgent. Since the 60s of the last century,

---

[2] Contemporary vocabulary (i.e. post 1970), finally, will be accounted for in the current INL-project *Algemeen Nederlands Woordenboek* (ANW).

[3] With some reservations with regard to the 14th–15th centuries, which are in principle covered by the Middle Dutch Dictionary (MNW). The MNW, although it certainly is a major achievement, has some flaws due to its relatively restricted text basis when compared to both the chronologically preceding (= Early Middle Dutch Dictionary: 13th century) and following dictionary (Dictionary of the Dutch Language: 1500–1976)

[4] For the most recent history of Dutch dialect lexicography see Goossens & Van Keymeulen (2006) and Taeldeman & Hinskens (*Language and Space: Dutch*, Mouton - De Gruyter, in press).

traditional dialects have been affected by both a functional and a structural reduction: they are spoken by ever fewer people in ever fewer situations. The vocabulary – the least stable language component – disappears first, due to the pressure of the standard Dutch language and the disappearance of the referents themselves, as is for instance the case for traditional agriculture, traditional crafts and trades and many aspects of modern life in general. A large part of the dialect lexicon has indeed already become a historical vocabulary.

Since the end of the 19th century, a considerable amount of work has been carried out in collecting dialect words, with different motivations: romantic and nostalgic feelings for the agrarian past; interest in historical linguistics and language reconstruction in the Neogrammarian paradigm; and recently an interest in cognitive semantics, etc. The linguistic interests of the public and scientists, of course, vary widely. Yet, it will be shown that in the case of lexical dialect research, a fruitful cooperation for both groups can be brought about.

2.1 Onomasiological dialect dictionaries and dialect atlases

Scientific Dutch dialect lexicography is not carried out at the Institute for Dutch Lexicology, as one might expect, but in fact began at the Catholic University of Nijmegen (today: Radboud University) by A.A. Weijnen, professor in Dutch, Indogermanic linguistics, Dialectology and Onomastics. Weijnen introduced the systematic, onomasiological arrangement of Dutch dialect lexicography. In doing so, he applied the lexicographical ideas of his predecessor Van Ginneken. According to Van Ginneken, the macrostructure of a dictionary should be presented in such a way that the clustering of semantically related words reveals the everyday life of the dialect speakers: this position resulted in a thematic arrangement. Next to the choice for a thematic arrangement, the lexicographical initiatives of Weijnen were also indebted to the dialectological paradigm of his day: i.e. word geography.

In the 1960s, Weijnen started the *Woordenboek van de Brabantse Dialecten* (WBD, Dictionary of the Brabantic Dialects) and the *Woordenboek van de Limburgse Dialecten* (WLD, Dictionary of the Limburg Dialects).[5] In 1972, prof. W. Pée started the *Woordenboek van de Vlaamse Dialecten* (WVD, Dictionary of the Flemish Dialects),[6] along the lines set out by Weijnen. The WBD was completed in 2005; the WLD in 2008. The Dictionary of the Flemish Dialects is still being compiled. The three dictionaries combined (and together with the 'mappable' WZD)[7] cover the

---

[5] In the 1990s an editorial board for the two dictionaries was opened at the KULEUVEN (Catholic University of Louvain).

[6] The term 'Flemish' is often misunderstood. The Flemings have Dutch as their standard language. Flemish is used as a colloquial term for 'Belgian Dutch'; in dialectology it denotes a dialect group in the west of Dutch-speaking Belgium.

[7] The *Woordenboek der Zeeuwse Dialecten* (WZD, Dictionary of the Zeeland Dialects, 1964) by Mrs. H.C.A. Ghijsen is the first regional dictionary which gave detailed geographical information for every single dialect word.

whole of the southern Dutch language area, i.e. French-Flanders (France), Dutch-speaking Belgium, and the three southern provinces of the Netherlands (Zeeland, Northern Brabant and Limburg) (see Figure 1 below for the dialect landscape of the southern Dutch dialects).

All the dictionaries of Weijnen's school[8] are arranged in the same way: every fascicle deals with a certain 'theme' (e.g. 'birds', 'the miller') pertaining to one of three parts: I. Agricultural Vocabulary; II. Technical and Crafts Vocabulary; III. General Vocabulary. Every fascicle is onomasiologically arranged and consists of a row of concepts, headed by a standard Dutch 'title', followed by a description and the heteronyms[9] which can be used to refer to the concept. Every dialect word is followed by (general) indications as to frequency and location[10] (see Figure 2 below for an example).
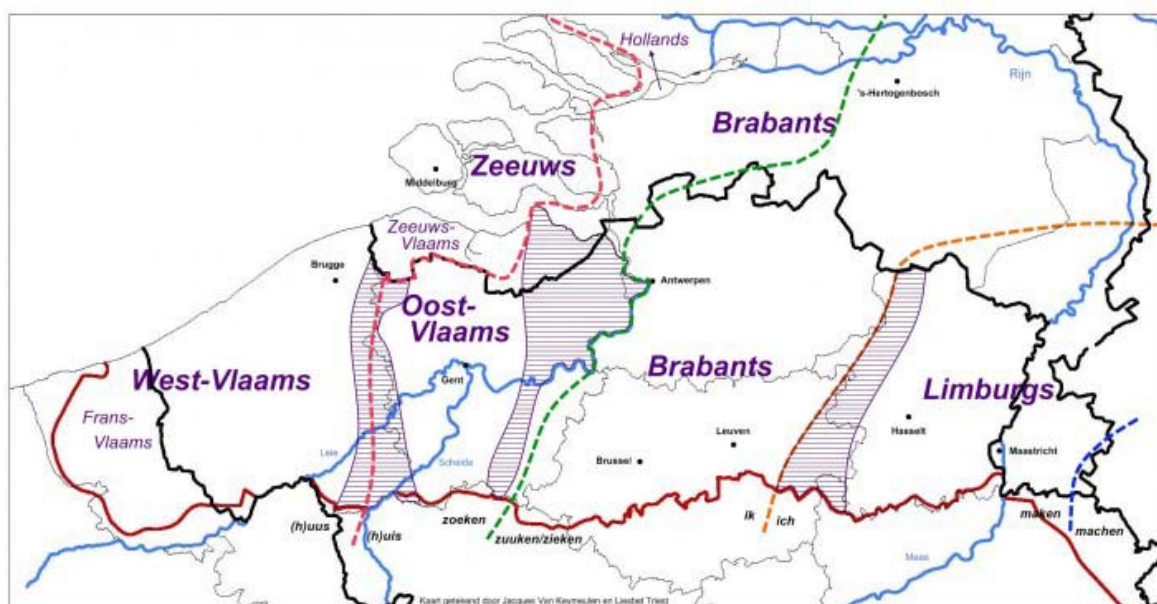

Figure 1: Dialect landscape of the southern Dutch dialects (according to Taeldeman, 2001: 8)

The printed texts of the three above-mentioned dictionaries for the southern Dutch dialects were based on automated databases from the late 80s onwards. For WBD and WLD the program Filemaker was used; for the WVD a specialized program (a relational database under Oracle) was developed by InfoService Belgium (ISB), a Ghent software company. The databases are provided with cartographic tools. The Filemaker database was already used for sophisticated geographical analyses (see

---

[8] The WBD/WLD also served as an example for the WALD, WOD and WGD afterwards.

[9] Weijnen coined the useful term *heteronym* for (dialect) words which mean the same in different (dialectal) language systems.

[10] See Van Keymeulen & De Tier (2010) and Kruijsen & Van Keymeulen (1997) for more details.

Foldert de Vriendt, 2012 for the D² project). A combination of the databases[11] of the three dictionaries will hopefully start in the near future.

KOPPIG

▲ hoofdig
▬ koppig
○ steeg
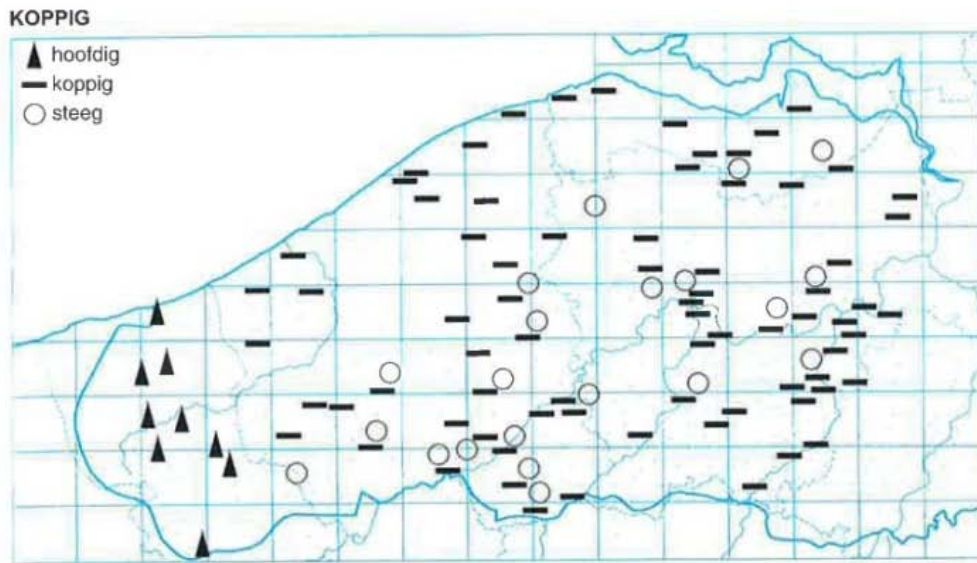
KOPPIG

Niet van eigen wil of inzichten af te bren-
gen.

De Bo vermeldt voor 'koppigheid' **kop-
achtigheid** i.v. kop.

WVD 129 (2001), 75. ZND 1 (1923), 50e;
ZND 28 (1938), 31; Willems (1886), 145.

bokkerig [*bukkerig*]: ♦ Veurne.

bokkig [ook: *bukkig*]: Lovendegem.

♦ Schoonaarde. Wdb: Loquela: bukkig,
Kortemark.

dikkoppig: ♦ Heist-aan-Zee.

driekante: ♦ Beveren (IJzer).

eenwillig: Wdb: Lievevrouw-Coop-
man.

hoofdig: f r e q . FV.

♦ s p o r . FV, ook Nazareth.

keikoppig: Roeselare.

kopachtig: Wdb: De Bo; Teirlinck:
kopachteg i.v. kop.

koppig: f r e q . WV, ZV en OV.

♦ a l g . WV en OV, ook Steenvoorde, Aar-
denburg en Zuiddorpe. Wdb: Teirlinck:
koppeg i.v. kop.

kopziende: Sint-Michiels.

obstinaat [ook: *opsternaat*]: Oostende,
Kooigem, Nieuw-Namen.

Wdb: De Bo; WZD: opste(r)naot,
opste(r)naet, obstinaot: WZV (Breskens,
Groede, Retranchement), L. v. Hulst.

obstinatig [ook: *opstenadig*]: ♦ Elsegem.
Wdb: Desnerck: opsenoadig, opstenoadig.

opienig: Geraardsbergen.

Wdb: De Bo: opijnig; Teirlinck: opieneg
i.v. opiene.

ossekoppig: ♦ Heist-aan-Zee.

steeg: f r e q . WV zuidoost; s p o r . str.
v. Ieper-Poperinge, ZV grensstr. en OV.
Wdb: Joos.

stijf: ♦ Sint-Goriks-Oudenhove.

stijfhoofdig: ♦ Wachtebeke.

stout: ♦ s p o r . OV zuid, ook Koksijde.

KOPPIG PERSOON

Iemand die niet van eigen wil of inzicht af
te brengen is, die tegen alle rede en goede
raad in aan zijn eigen mening vasthoudt.

Voor een koppig kind geeft WZD **steen-
vreter** (stêênfreter) in WZV (Retranche-
ment) en L. v. Axel (Axel, Hoek). De Bo

Figure 2: A page from WVD III, 4: 195 'stubborn'

[11] All three dictionaries have printed fascicles dating from the pre-computer era. Their data still have to be digitized.

The geographical aspect of the southern Dutch dialect landscape is covered by the three dictionaries of Weijnen's school. The semantic aspect of the dialect vocabulary, however, is poorly accounted for. Because of their onomasiological arrangement, the dictionaries cannot render detailed semantic descriptions. Overall, they have a very poor microstructure. Only in a dictionary where semantic detail (together with other microstructural information) is added to a headword (normally alphabetically arranged), is it possible to describe 'meaning', since in order to do so one needs to start from the lexeme (and not from the concept as is the case in onomasiological dictionaries).[12] Thus, the onomasiological dictionaries of Weijnen's school are to be considered as highly-structured geographical inventories of word usage: they are – notwithstanding the names of the publications – atlases, not dictionaries.[13]

2.2 Semasiological dialect dictionaries: two traditions

2.2.1 The old alphabetical tradition

The last half of the 19th century / first half of the 20th century witnessed some important works in the Netherlands, such as Molema (1887) for the province of Groningen and Boekenoogen (1897) for the dialect of the Zaan district (North Holland). The major regional alphabetical dialect dictionaries in Dutch-speaking Belgium were compiled because of romantic motivations and the search for linguistic identity by the Flemings.[14] De Bo (1873), for instance, collected West-Flemish words in order to contribute to a Flemish standard language of its own (a plan that did not work out). All those regional dictionaries, which sometimes cover whole provinces, have one important flaw: they do not exactly locate the dialect words. The geographical scope of the words is not indicated.

The above-mentioned dictionaries were not produced by an institute; hence we have called them 'amateur' dictionaries. This qualification perhaps does an injustice to the high scientific quality of many of them. The *Zuid-Oostvlaandersch Idioticon* (Southeast Flemish Dictionary) of Teirlinck (1908–1924), for instance, is a marvellous example of completeness and semantic detail. Some amateurs of the 'old school' certainly deserve admiration for the scientific excellence of their work.

2.2.2 Dialect 'revival' and amateurs

The standard Dutch language has spread to large sections of the population since the 1960s, hence using a dialect is nowadays regarded as a matter of 'choice' and not as a sign of backwardness. The new positive attitude towards these 'endangered' varieties, considered as 'immaterial heritage', has resulted in the production of amateur dialect dictionaries, meant for a local population.

---

[12] For a discussion of this matter: see Weijnen 1961, 1963, 1967 and De Tollenaere 1960, 1968.

[13] The WZD proves that geography can have its proper place in an alphabetical dictionary, together with detailed information as to meaning.

[14] The most important ones being: De Bo (1873), Joos (1900), Tuerlinckx (1886), Teirlinck (1908–1922), Rutten (1890), Cornelissen-Vervliet (1899–1903).

The table below (Table 1), taken from Oosterhof & Van Keymeulen (2009), visualizes the production of amateur dictionaries from 1835 until the beginning of the 20th century. The lexicographic effect of dialect revival becomes clear after the 1980s.

	FLANDERS					THE NETHERLANDS												
	LI	AN	VB	OV	WV	GR	FR	DR	OV	GD	FL	UT	NH	ZH	ZL	NB	LB	TOT
1835–39																1		1
1850–54																	1	1
1880–84																	1	1
1895–99									1									1
1900–04														1				1
1910–14																	1	1
1915–19																	1	1
1925–29															1		1	2
1935–39			1				1									1		3
1940–44										1								1
1945–49														1				1
1950–54				1														1
1955–59									1							1	1	3
1960–64																		0
1965–69							1		1					1			1	4
1970–74					1					1						1	2	5
1975–79			1	1			1			1	1					1	2	8
1980–84	3		1	1			1			5			1	2		2	3	19
1985–89	3	1	2	4	1					3				2		4	7	27
1990–94	3	1	4	3	2		1		2	5	2		1			8	9	41
1995–99	5	2	5	7	5		2	1	2	4		2	1	2		7	3	48
2000–04	6	2	2	3	6		2		5	3		1	4	2		8	8	52
2005–		1	1	4	1				1	3			1	4		7	11	34
TOT	20	7	17	24	16	0	9	1	13	26	3	3	8	15	1	41	52	256

Explanation of abbreviations: LI = Limburg (Belgium); AN = Antwerp; VB = Flemish Brabant; OV = East Flanders; WV = West Flanders; GR = Groningen; FR = Friesland; DR = Drenthe; OV = Overijssel; GD = Gelderland; FL = Flevoland; UT = Utrecht; NH = North Holland; ZH = South Holland; ZL = Zeeland; NB = North Brabant; LB = Limburg (the Netherlands)

Table 1: Local dialect dictionaries by lustrum and province
(Van Keymeulen & Oosterhof, 2009)

Many things may go wrong when amateurs engage in writing dictionaries. There is indeed a huge variety, both in quantity (i.e. number of entries) and quality (notably of the semantic descriptions) between amateur dictionaries, and in the macro- and microstructural options. It goes without saying that the best of such dictionaries are made by amateurs with a linguistic schooling, although some authors with no training command admiration for their perseverance in detailed observation and the ensuing lexicographical result.

Nearly all amateur dictionaries copy, to a certain extent, the macro- and microstructural options of the standard language or bilingual dictionaries with which they are acquainted. Few amateurs know that there is such a thing as professional dialectology or professional lexicography. Nearly all local dictionaries are alphabetically arranged, mostly using a home-made dialect spelling for the headwords. Few amateurs are aware that this way of presenting the macrostructure of a local dialect dictionary will frustrate the intended user.

The interpretation of an entry word, rendered in a home-made dialect spelling, indeed presupposes a good knowledge of the dialect, on both lexical and phonological levels, in order to be able to look up a word. The form of the headwords is very often in blatant opposition to the needs of the user as envisaged in the introduction of many a dictionary: namely the user in a dialectless future. The example set by alphabetical standard language dictionaries seems to be very strong indeed, and inventing a spelling system for a dialect is a codifying activity with a high symbolic value.[15] Many authors are inclined to appropriate the dialect by creating spelling for it. Members of the local population buy the dictionary not to use it, but to 'possess' it, as a symbol of their local identity.[16]

3. Towards a Woordenbank van de Nederlandse Dialecten[17] (WND, Wordbase of Dutch Dialects)

A comprehensive collection of the dialect vocabulary of a vast area is a gigantic task, often beyond the financial reach of an institution. Although amateur lexicography obviously has its flaws, there are many amateur lexicographical products of good quality. They give at least a word form and an indication as to locality or region. Since they are alphabetically arranged, they are in principle able to describe meaning. Many of them also include collocations of all types: idiomatic expressions, proverbs, etc. Many contain example sentences in which meaning is illustrated. In short, they come as very welcome additions to the onomasiological dictionaries of the WVD-type, which are not capable of conveying semantic details and are altogether very poor in microstructure in general.

In September 2009, a pilot project[18] was launched at Ghent University, funded by the Flemish Ministry of Culture, which envisages the creation of a digital database for

---

[15] Van Keymeulen (1993) and Cajot (1995) are lexicographical manuals, meant for amateurs.

[16] Delarue (2009) devoted a master thesis to the evaluation of a number of amateur dictionaries.

[17] We thank J. Kruijsen for his suggestion to name the database *Woordenbank* instead of *Woordenboek*.

[18] The first tentative results of the pilot study were reported in Van Keymeulen & De Tier (2010).

alphabetically arranged regional and local dialect dictionaries.[19] The project consists of two phases: 1) creating a digital database on the basis of the dictionary texts; and 2) annotating the database in order to be able to perform efficient search operations. Software is being developed by the firm Info Service Belgium (Ghent). The overall purpose of the project is to combine the products of the 'amateur' dialect lexicography, and organize them in a digitized database in such a way that many types of automated searches can be done. In what follows, the present state of affairs of the *Woordenbank* (demo on www.woordenbank.be) will be presented. In 2012 the idea was copied by the Meertens Institute in Amsterdam; it was agreed that Ghent University would take care of the amateur lexicography in Dutch-speaking Belgium and the province of Zeeland; the Meertens Institute would collect the material of the dictionaries in The Netherlands.

3.1 The old alphabetical tradition

The dictionaries of the 'old alphabetical tradition' (end 19th / beginning 20th century) have a relatively bad printing quality, which creates problems for OCR-procedures. Much time is needed to correct the text files, and prepare texts for the input. Luckily, the project can rely on a number of volunteers to do the job.

Old dictionaries usually cover fairly large areas. Since they had to cope with a wide geographical differentiation in matters phonological, they were obliged to 'lift' the orthographical form of the headword to a relatively high level, i.e. they were obliged to normalize dialect word towards standard Dutch. In the West-Flemish dialects, for instance, standard Dutch *sch*- [sχ-] as in *school* can be represented by [sk]-, [ʃ]-, [ʃχ]-, [sʔ]. All this variation is implied and summarized in the normalized form <sch> in the headword *school* in the West-Vlaams Idioticon of De Bo (1873). The dialectal headwords are, so to speak, 'dutchified', i.e. written as if the word were a standard Dutch word. For the older dictionaries, the headwords should be modernized in spelling (e.g. *bosch > bos* 'wood'; *keeren > keren* 'turn'; *roozewied > rozewied* 'cornflower'). Since this activity is relatively easy, it can be carried out by volunteers.

3.2 The dictionaries of the 'dialect revival'

As explained above, a few hundred amateur dialect dictionaries have been written since the 1980s. Since the late 90s, many amateurs have created lexicographic databases, and have produced digital texts. Many of them were willing to contribute a copy of their final texts, and even volunteered to prepare them for input in the WND. In many cases, however, corrections of the proofreading still had to be transferred to the textfile. As the dictionaries were very often sold out in their printed form, most authors were willing to cooperate with the WND, without requesting financial or other compensation.

---

[19] Financial support was given to the organization 'Varieties vzw. Koepelorganisatie voor Dialecten en Oraal Erfgoed in Vlaanderen', based at Ghent University.

Most amateur dictionaries are meant for a local population. Even in 'local' dictionaries, phonological (and even lexical) differentiation may occur. The dictionary of Cools (2000) on the dialect of Beveren-Waas (East Flanders), for instance, tries to account for seven former villages (Doel, Haasdonk, Kallo, Kieldrecht, Melsele, Verrebroek, Vrasene) which were joined with the town of Beveren in the 1970s. Because of this geographical scope, however small it may be, the author was forced to normalize the headwords towards standard Dutch. The word *lopen* 'to run', e.g., is pronounced in three different ways ([ly.ᵊpm], [lu.ᵊpm] and [li.ᵊpm]) in the different villages. So as to avoid frustrating a dialect-speaking community, the 'dutchified' headword *lopen* is used, thus avoiding having to make an unwanted choice between the dialects. Since the normalizing of dialect words towards standard Dutch presupposes etymological insights, the dutchifications by amateurs without linguistical training cannot always be trusted. They should be evaluated and, if necessary, corrected before input into the WND.

Most dictionaries of the dialect revival have a very small geographical focus and use home-made spellings. Dutchification is applied on the basis of etymology and the correspondence rules between standard Dutch and the dialect. This is performed by volunteers, who receive in-job training, and is corrected afterwards by linguistically skilled persons.

A few examples of normalization / dutchification of headwords are:

dialect spelling	>	dutchification
buttersjhuute ('butterfly') (Desnerck, 1972)	>	*boterschuit*
kraaisj ('cross') (Pletinckx, 2003)	>	*kruis*
trênink ('training') (Pletinckx, 2003)	>	*training*
ouverttoeegd ('convinced') (Wellekens, 1994)	>	*overtuigd*
(h)euneenk ('honey') (Pieters, 1995)	>	*honing*

The normalization of the original spelling of the dialect word is undoubtedly the most essential addition to the database.[20] In this way, the word collections will be opened up for non-dialect speakers and for scientific research. As a rule of thumb, the orthography of the headwords (and variants of them) in the WNT will be taken as a guide, since the WND will eventually be linked with iWNT in the Integrated Language Database.

---

[20] Normalization is not always easy, hence the database allows for the input of more than one normalized form.

At present, records in the database contain fields for: 'original headword', 'dutchified headword' and 'search term in standard Dutch' (boiling down to a one-word translation). It is evident that many more additions or refinements are possible: the different elements of the microstructure may be inputted (or added) into different fields, in order to facilitate many types of research. The fully-fledged enrichment of the database (including the addition of audio files) is postponed until a later phase of the project. The present record structure is as follows:

Original dictionary article:

> Beddezièèkre, (nen) A.w.v. Paardebloem. (Taraxacum officinale). Samengesteldbloemigen met een gele bloem en gepluimde zaadjes. Men zegt ook wel Nen beddepissere. [21]

Field structure in the WND-database with annotations:

Beddezièèkre (= *original headword*)
beddezeiker, beddezeker (= *dutchification of the headword*)
paardenbloem (= *search term in standard Dutch*)

(example taken from Clinckemaillie, 1996)

4. Procedures

4.1 Software

The software of the Wordbase of Dutch Dialects operates under an Oracle platform. The software has been developed by the firm Info Service Belgium (Ghent). The database has been compiled from existing amateur dialect dictionaries on paper (see §2.). The first step is to digitize these dialect dictionaries by scanning and ocr-ing, if they are not available in a digitized version.[22]

When a dictionary is scanned and ocr'd, a Word file is used as output. In most cases this text is not without mistakes, especially if the headwords are written in a self-made dialect spelling that is not recognizable by a computer. The Word files of these dictionaries need to be corrected, in our case by volunteers. At this moment a new campaign has started to increase the number of volunteers. Most volunteers are not skilled lexicographers. They are mostly male persons between 50 and 70 years of age, willing to do some cultural work.

---

[21] Translation in English: 'dandelion'.

[22] The final text file version of a dictionary is normally in the editor's possession (and not the author's).

In order to import the Word file into the database, the text must comply with some standards. Each dictionary article has to begin with a bold headword (it being the original headword) and end with two hard returns. Once a Word file is corrected, and once the headword is put into bold, and two hard returns are inserted after each dictionary article, the file is ready for input in the database (see Figure 3).

Fleur(e)pietje [ˈfløːr(ə)pitʃə], zn. o. 1 Het allerbeste, allermooiste, allersterkste. - 2 Goede tol (speelgoed).

Fleures [ˈfløːrəs], zn. o. Longontsteking, pneumonie. Bij Kiliaan al: *pleuris/pleurisije.* Uit *pleuris/ pleuritis* met wisseling van *pl/fl,* zoals in *flerecijn,* Wvl. *prut/frut* 'cichorei', *Plutol/Flutol, perplex/ perflex.* Ook *vliegend fleures.* Zie ook *waterfleures.*

Fliflouder, zie Fifouter.

Figure 3: Word file corrected by volunteers as input for the database

(example from Debrabandere, 1999)

4.2 Input in the database

The Word document is converted into a standard XML-file by means of a script. The XML-file can be imported into the database by means of a purpose-built application for this database. For some dictionaries, it is necessary to write an adapted custom script, which generates the standard XML-file for the application. These scripts will generate the XML-file by means of the already mentioned typographical conventions. Once the XML-files are uploaded, the database of the Wordbase of Dutch Dialects can be made. Each bold headword + the microstructure of it[23] will be put into the database of the Wordbase of Dutch Dialects as one entry (see Figures 4 and 5).

4.3 Annotation

The editors or volunteers (most of whom speak, or are acquainted with, the dialect of the specific dictionary) may then annotate the database with dutchifications, search terms in standard Dutch, comments and (later on) thematic markers and location (see the example in Figure 6). Sometimes it is possible to add some of these annotations automatically (e.g. when the headword is already a dutchification in the original work), but mostly the editor has to do this part half automatically or manually.

Because volunteers are helping to add the dutchifications and the search terms in standard Dutch, we decided two years ago to introduce a second, easier, way to put the data into the wordbase. While volunteers are correcting the ocr'd text, they can

---

[23] Different dictionaries vary widely as to their microstructure (semantic definitions, phonetic notation, grammatical information, example sentences ...).

already insert the dutchifications and the search terms in the Word file itself. This is performed by putting codes at the end of the dictionary article. The dutchification is placed between $$ $$; the search term is put between ££ ££ (see Figure 4).

Fleur(e)pietje [ˈfløːr(ə)pitʃə], zn. o. 1 Het allerbeste, allermooiste, allersterkste. - 2 Goede tol (speelgoed).
$$fleurepietje, fleurpietje$$ ££allerbeste; tol££

Fleures [ˈfløːrəs], zn. o. Longontsteking, pneumonie. Bij Kiliaan al: *pleuris/pleurisije*. Uit *pleuris/ pleuritis* met wisseling van *pl/fl,* zoals in *flerecijn,* Wvl. *prut/frut* 'cichorei', *Plutol/Flutol, perplex/ perflex*. Ook *vliegend fleures*. Zie ook *waterfleures*.
$$fleuris$$ ££longontsteking££

Figure 4: Word file corrected and annotated by volunteers as input for the database

(example from Debrabandere, 1999)

Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fientig	fijntig	tenger, frêle
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fiesshow	fichau, fisjauw	bunzing; sluwe rappe persoon
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fiester'n	fisteren	zeuren, zaniken
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fietjefatjerieë	fitjefatjerie	waardeloze spullen; charcuterie
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fiette	fietje	sofietje; dommevrouw; ziezo
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fikke	fikje	klein, levendig kind/persoon
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fikkel'n	fikkelen	onhandig werken/snijden
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	fleet'aor	fleet haar	dun sluik haar
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	flêitre	fleiter	oorveeg
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	flièèws	fleeuws	flauwe fletse smaak
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	flieflottre	flieflotter	vlinder
Het dialect van Midden-West-Vlaanderen - J. Clinckemaillie Import: F-G (421)	flikk'n	flikken	werkje opknappen; bedriegen

1) ▲button to visualize the lemma ▲2) importfile ▲3) original headword ▲5) search term
▲4) dutchification

Figure 5: List of lemmata uploaded in the database (by clicking the button on the left, one accesses the specific dictionary article) (example from Clinckemaille, 1996)

Macros are employed to change typical dialect or orthographic characteristics in order to make the adding of dutchifications and search terms a bit more comfortable for the volunteers. In this way, it is possible to suggest dutchifications or search terms. The volunteer knows what alterations are made by means of the macro and when correcting the text file and reading the annotations, he may correct and complete suggestions simultaneously. These macros are of course not always the same for each dictionary, so for each new dictionary the macrostructure has to be analyzed.

Figure 6: Example of the lemma *flieflottre* ('butterfly') (*flieflottre* = original headword) in the Wordbase of Dutch Dialects, annotated with dutchification (= vernederlandsing) (*flieflotter*) and search term (= A.N.-zoekterm) (*vlinder*) (example from Clinckemaille, 1996)

This procedure is easier for volunteers, because they do not have to use the online database for correction. The work can thus be performed on a stand-alone computer or without internet. The input text for the computer program is the corrected Word file, with the addition of dutchification and search terms between codes (see Figure 4). This coded text is then converted into a standard XML file by means of a script, and this file can be imported into the database by a purpose-built application. The fields with dutchification and search terms will therefore already feature in the database and do not need to be added online after the import.

4.4 Front and back matter

Both the front matter (such as copyright page, imprint, foreword, list of abbreviations, user's guide etc.) and the back matter (lists, index, etc.) of each dictionary are incorporated in the database as pdf-files. Next to that, the bibliographical reference is added (together with a pdf for title page and cover) and the state of affairs with regard to correction and annotations. The illustrations are imported either at the level of the dictionary article, or at a higher level. Addition of sound samples is not yet possible (partly because of the high time investment it would require).

4.5 Website

The database is connected to a website with search facilities and the following search terms: 'original headword', 'dutchification', 'search term in standard Dutch' and 'word in article'. The dutchifications and general search term 'word in article' really disclose the word collection as a whole. In each search facility, one can use wildcards.

The list of dictionaries is growing. Most users are looking for a particular word in one region. That is why there is also an option to search in all or some dictionaries or in one dictionary. One can also choose the dialect dictionaries of one province.

Recently a contact form has been added. With this form, users can let us know if they see mistakes, or if they want to give a new explanation which is not yet in the dictionary. In the next addition we intend to visualize a comment-line. Therefore, when there are obvious mistakes or questions, when an author wishes to add something, or when the editors wish to provide additional information, adding and visualizing these comments will be possible. Currently, the database itself includes a comment possibility. This is intended for the volunteers to offer questions or comments. The comment-field is not displayed on the website. A list of the dictionaries that are completely or partially annotated and imported is also available. Under the information button, one can find information on the state of affairs for each dictionary.

4.6 Preliminary results

The search facilities in the database are already operational although the dictionaries in the database have not been completely imported and annotated as yet. The search facilities still have to be evaluated and maybe adapted to the needs of the user. To evaluate the usability, it would be good to start with a selected user group in the near future.

The import of some dictionaries has been completed (e.g. the *Kortrijks Woordenboek* of Debrabandere, 1999). In the case of other dictionaries, it remains to import the front and back matter and the annotations. We rely heavily on volunteers, but hope to be able to present a database within a few years which will contain at least 30 dictionaries. The total number of entries (dictionary articles from all dictionaries incorporated in July 2013) is 226,788. Twenty-six dictionaries have been (partly or completely) imported in the database, but the regional distribution has to be improved in the near future.

The demo version of the website with the first results can be seen on www.woordenbank.be (see a list of search results in Figure 7).

Figure 7: Search results for search term = vlinder ('butterfly')
in the Wordbase of Dutch Dialects

5. Conclusion

Compilation of the *Woordenbank van de Nederlandse Dialecten* is under way, thanks to volunteers, and coordinated by Veronique De Tier at Ghent University. Mrs Silvia Weusten collaborated for approximately one year, especially for Limburg dictionaries (and the Ghent dictionary). The workload is indeed gigantic, because of the lexical and semantic richness of the dialects and because of the sheer number of good dictionaries. Thanks to the different search terms, both the semantic and encyclopedic information of the dictionaries are easily accessible. The results of a search term in the microstructure reveal not only lexicographic data, but also show that the *Woordenbank* can be used as an ethnographic database.

6. References

Boekenoogen, G.J. (1897). *De Zaansche volkstaal. Bijdrage tot de kennis van de woordenschat in Noord-Holland.* Leiden: Sijthoff.

Cajot, J. (1995). *Hoe maak ik een dialectwoordenboek. Een handleiding voor Limburgers en anderen die dialectwoorden willen spellen, verzamelen en beschrijven.* Mededelingen van de Vereniging voor Limburgse Dialect- en Naamkunde nr. 78/79. Hasselt: Vereniging voor Limburgse Dialect- en Naamkunde.

Clinckemaillie, J. (1996). *Ool koett'n en ool doen. Het dialect van Midden-West-Vlaanderen.* Aartrijke: Decock.

Cornelissen, P.J. & Vervliet J.B. (1899-1903). *Idioticon van het Antwerpsch dialect (stad Antwerpen en Antwerpsche Kempen).* Gent: Siffer.

Debrabandere, F. (1999). Kortrijks Woordenboek. Kortrijk/Brugge: De Leiegouw/Uitgeverij Van de Wiele.

De Bo, L. (1873). *Westvlaamsch Idioticon.* Brugge: Gailliard.

Delarue, S. (2009), *Van woordenlijst tot woordenboek. Lexicografische beschrijving en evaluatie van de Vlaamse amateurdialectlexicografie in de 19de, 20ste en 21ste eeuw.* (unpublished master thesis, Ghent University).

De Tier, V. & Van Keymeulen, J. (2010). Software demonstration of the dictionary of the Flemish Dialects and the Pilot Project Dictionary of the Dutch Dialects. In: Dykstra, A & T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress.* Fryske Akademy, Leeuwarden. 620-627 (issued on CD-ROM).

De Tollenaere, F. (1960). *Alfabetische of ideologische lexicografie? Bijdragen tot de Nederlandse Taal- en Letterkunde.* Leiden: Uitgegeven vanwege de Maatschappij der Nederlandse Letterkunde I.

De Tollenaere, F. (1968). Problemen van het dialectwoordenboek. Theorie en praktijk. In: *Tijdschrift voor Nederlandse Taal- en Letterkunde 84.* 197-212.

De Vriendt, F. (2012). *Tools for computational analyses of dialect geography data.* Radboud University Nijmegen (Ph.D. dissertation).

Frisian Dictionary = *Woordenboek van de Friese Taal / Wurdboek fan de Fryske Taal.* Leeuwarden : Fryske Akademie.

Goossens, J. & Van Keymeulen J. (2006). De geschiedenis van de Nederlandse dialectstudie. In: *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie 78.* 37-97.

Joos, A. (1900). *Waasch idioticon.* Gent/St.-Niklaas: Siffer/Strijbol.

Kruijsen J. & J. Van Keymeulen (1997). The Southern Dutch Dialect Dictionaries. In: *Lexikos 7.* 207-228.

Molema, H. (1887). *Woordenboek der Groningsche volkstaal.* Winsum: Mekel.

Pieters, M. (1995), *Woordenboek van het Lokers Dialect.* Lokeren: Uitgeverij Oelbrandt.

Pletinckx, L. (2003). *Woordenboek van het Asses. Bijdrage tot de studie van de West-Brabantse streektaal.* Asse: Koninklijke Heemkring Ascania.

Rutten, A. (1890), *Bijdrage tot een Haspengouwsch Idioticon*. Antwerpen, Zuidnederlandsche Maatschappij van Taalkunde; Boucherij.

Rys K. & Van Keymeulen J. (2009). Intersystemic correspondence rules and headwords in Dutch dialect lexicography. In: *International Journal of Lexicography* 22. 129-150.

Taeldeman, J. (2001). De regenboog van de Vlaamse dialecten. In: Devos, M., J. De Caluwe, & J. Taeldeman (eds.). *Het taallandschap in Vlaanderen.* Wetenschappelijke Nascholing UGent.

Taeldeman, J. & F. Hinskens (in press). *Language and space: Dutch.* Berlin : Mouton - De Gruyter.

Teirlinck, I. (1908-1924). *Zuid-Oostvlaandersch Idioticon.* Gent: Siffer.

Tuerlinckx, J.F. (1886). *Bijdrage tot een Hagelandsch Idioticon.* Gent: Zuidnederlandsche Maatschappij van Taalkunde.

Van Dale = *Van Dale Groot Woordenboek van de Nederlandse Taal* (2005).

Van Keymeulen, J. (2003). Dialectwoorden verzamelen. Een praktische handleiding. In: *Handelingen van de Koninklijke Commissie voor Toponymie en Dialectologie 75.* 383-506.

Van Keymeulen, J. (2003). Compiling a dictionary of an unwritten language. A non corpus-based approach. In: *Lexikos 13.* 183-205.

Van Keymeulen, J. (2004). Trefwoorden en lexicale varianten in de grote regionale dialectwoordenboeken van het zuidelijke Nederlands (WBD, WLD, WVD). In De Caluwe J., G. De Schutter, M. Devos & J. Van Keymeulen (eds.). *Taeldeman, Man van de Taal, Schatbewaarder van de Taal.* Gent: Vakgroep Nederlandse Taalkunde / Academia Press. 897-908.

Van Keymeulen, J. (2009). Volkslinguïstiek en dialectlexicografie in de zuidelijke Nederlanden. In: *Lexikos 19.* 314-339.

Van Keymeulen, J. & A. Oosterhof (2009). Local dialect dictionaries and a proposal for a dictionary of the Dutch dialects. In: Gooskens, C., A. Lenz & S. Reker (eds.). *Low Saxonian dialects across borders: Synchrony and diachrony.* Stuttgart: Franz Steiner. 109-124.

Van Keymeulen, J. & V. De Tier (2010). Pilot Project: A Dictionary of the Dutch Dialects. In: Dykstra, A. & T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress.* Fryske Akademy, Leeuwarden. 754-763 (issued on CD-ROM).

Van Keymeulen, J. & V. De Tier (2010). Towards the completion of the Dictionary of the Flemish Dialects. In: Dykstra, A. & T. Schoonheim (eds.), *Proceedings of the XIV Euralex International Congress.* Fryske Akademy, Leeuwarden. 764-773. (issued on CD-ROM).

WALD = Schaars, L. (1984 -). *Woordenboek van de Achterhoekse en Liemerse Dialecten.* Doetinchem: Staringinstituut.

WBD = Weijnen, A. e.a. (1967-2005). *Woordenboek van de Brabantse Dialecten.* Assen/Maastricht: Van Gorcum; Groningen/Utrecht: Gopher.

Weijnen, A. (1961). De semantische en syntactische problematiek van het dialectwoordenboek. In: *Tijdschrift voor Taal- en Letterkunde 78.* 81-95.

Weijnen, A. (1963). Het dialectwoordenboek. In: *Woordenboek en Dialect. Bijdragen en Mededelingen der Dialectencommissie van de Koninklijke Nederlandse Akademie van Wetenschappen te Amsterdam 24.* 34-52.

Weijnen, A. (1967). De waarde van een dialectwoordenboek. In: *Mededelingen van de Nijmeegse Centrale voor Dialect- en Naamkunde 6.* 5-11.

Wellekens, W. (1994), *Diksjonêr van 't Leives.* Academie voor het Leuvens Dialect: Leuven.

WGD = Giesbers, C./Scholtmeijer, H. (2005 -). *Woordenboek van de Gelderse Dialecten.* Utrecht: Matrijs.

WLD = Weijnen, A., J. Goossens e.a., (1983-2008). *Woordenboek van de Limburgse Dialecten*. Assen/Maastricht: Van Gorcum; Groningen/Utrecht: Gopher.

WOD = Scholtmeijer, H. (2000-). Woordenboek van de Overijsselse Dialecten. Kampen: Stichting IJsselacademie.

WVD = Devos, M. e.a. (1972-). *Woordenboek van de Vlaamse Dialecten.* Gent-Tongeren: Michiels.

WVD III, 4 = De Pauw, T. & M. Devos (2005). *Woordenboek van de Vlaamse Dialecten. Algemene Woordenschat. Aflevering 4: Karakter.* Gent-Tongeren: Michiels.

WZD = Ghijsen, H.C.M. (1964). *Woordenboek der Zeeuwse Dialecten.* Den Haag, Van Goor. [Fraanje, K. e.a. (2003), *Supplement Woordenboek der Zeeuwse Dialecten.* Krabbendijke: Van Velzen]

WNT = Woordenboek der Nederlandsche Taal (1864-1998). See: http://gtb.inl.nl/?owner