

Received 10 April 2019.

Accepted 30 May 2019.

THE DICTIONARY OF THE SOUTHERN DUTCH DIALECTS (DSDD): DESIGNING A VIRTUAL RESEARCH ENVIRONMENT FOR DIGITAL LEXICOLOGICAL RESEARCH

Jacques VAN KEYMEULEN, Veronique DE TIER, Roxane VANDENBERGHE & Sally CHAMBERS

Ghent University, Belgium *

jacques.vankeymeulen@ugent.be / veronique.detier@ugent.be /

roxane.vandenberghe@ugent.be / sally.chambers@ugent.be

Abstract

The southern Dutch dialect area consists of four dialect groups: (1) the Flemish dialects, spoken in French Flanders (France), West and East Flanders (Belgium) and Zeeland Flanders (the Netherlands); (2) the Brabantic dialects, spoken in Antwerp and Flemish Brabant (Belgium) and Northern Brabant (the Netherlands); (3) the Limburgian dialects (spoken in the Limburg provinces of Belgium and the Netherlands); (4) the Zeeland dialects, spoken in Zeeland and Goeree-Overflakkee (the Netherlands). The vocabulary of the Flemish, Brabantic and Limburgian group is brought together in three comprehensive thematically arranged dictionaries in a joint international and inter-university effort. The dictionaries are based on their own dialect questionnaires, but the very big - and older - collection of Ludovic Grootaers, father of Willem Grootaers, is added. Thus, the results of the lexicographic efforts of the father of Willem Grootaers, are brought to light at last (after more than 50 years!).

The dictionaries were set up in parallel in order to make possible the aggregation of the data, thus fulfilling the objectives of the founders of the projects (Toon Weijnen in Nijmegen, and Willem Pée in Gent). A project to this effect, the *Dictionary of the Southern Dutch Dialects* (DSDD) was launched in 2017. In our article, we will propose the plan for the aggregation and the state of affairs of the DSDD, the structure of the database and dwell on the different editorial problems that have to be solved. The different dictionaries / database were indeed composed over a very long period of time, at different

* Universiteit Gent, Vakgroep Taalkunde - Nederlands, Blandijnberg 2, 9000 Gent (Belgium).

places (Nijmegen, Leuven, Ghent) and by different editors, hence a great number of inconsistencies arose over time. In order to compose an aggregated DSDD-database, a number of standardization activities have to be carried out.

Keywords

Dutch, dialect lexicography, Flemish, Brabantic, Limburgian

EL DICCIONARIO DE LOS DIALECTOS NEERLANDESES MERIDIONALES (DSDD): DISEÑO DE UN ENTORNO DE INVESTIGACIÓN VIRTUAL PARA LA INVESTIGACIÓN LEXICOLÓGICA DIGITAL

Resumen

El área de los dialectos holandeses meridionales consta de cuatro grupos dialectales: (1) los dialectos flamencos, que se hablan en Flandes francés (Francia), Flandes occidental y oriental (Bélgica) y Flandes Zelanda (Países Bajos); (2) los dialectos brabantes, hablados en Amberes y Brabante flamenco (Bélgica) y Brabante del norte (Países Bajos); (3) los dialectos de Limburgo (que se hablan en las provincias de Limburgo de Bélgica y de los Países Bajos); (4) los dialectos de Zelanda, hablados en Zelanda y Goeree-Overflakkee (Países Bajos). El vocabulario del grupo flamenco, brabantón y limburgués se reúne en tres diccionarios completos organizados temáticamente en un esfuerzo conjunto internacional e interuniversitario. Los diccionarios se basan en sus propios cuestionarios dialectales, pero se agrega la muy grande y antigua colección de Ludovic Grootaers, padre de Willem Grootaers. Así, los resultados de los esfuerzos lexicográficos del padre de Willem Grootaers, finalmente salen a la luz (¡después de más de 50 años!).

Los diccionarios se crearon en paralelo para hacer posible la agregación de los datos, cumpliendo así los objetivos de los fundadores de los proyectos (Toon Weijnen en Nimega y Willem Pée en Gante). Un proyecto en este sentido, el *Diccionario de los dialectos neerlandeses meridionales* (DSDD: por sus siglas en inglés) se presentó en 2017. En nuestro artículo, propondremos el plan para la agregación y el estado de la cuestión del DSDD, la estructura de la base de datos y la atención a los diferentes problemas editoriales que deben ser resueltos. Los diferentes diccionarios / bases de datos fueron elaborados durante un período muy largo de tiempo, en diferentes lugares (Nimega, Lovaina, Gante) y por diferentes editores, por lo tanto, surgieron una gran cantidad de inconsistencias con el paso del tiempo. Para formar una base de datos DSDD agregada, se deben realizar una serie de actividades de estandarización.

Palabras clave

neerlandés, lexicografía dialectal, flamenco, brabantón, limburgués

1. Introduction

Dialectology in Dutch-speaking Belgium (Flanders) has a rich tradition in lexicography. There are two main traditions in that respect: an alphabetical one and an onomasiological one. In what follows, we will firstly dwell on the alphabetical tradition (section 2) and then move on to the onomasiological tradition of *Weijnen's school* (section 3). In section 4 we present the Dictionary of Southern Dutch Dialects (DSDD), a major lexicographic project, aimed at aggregating the databases of the three major dialect dictionaries for the southern Dutch dialects: the *Woordenboek van de Vlaamse Dialecten* (Dictionary of the Flemish Dialects), the *Woordenboek van de Brabantse Dialecten* (Dictionary of the Brabantic Dialects) and the *Woordenboek van de Limburgse Dialecten* (Dictionary of the Limburgian Dialects). In the latter part of this article (section 5), we will present the state of affairs of the DSDD project. In a final section (section 6), we will come to a conclusion and outline a number of directions for the future research.

2. The alphabetical dialect lexicography

The alphabetical dialect lexicography in the southern Dutch area can be divided in two: (a) the comprehensive 19th century / early 20th century lexicography, which was motivated by 'Flemish movement' attitudes and (b) the relatively recent wave of dictionary making in the so-called dialect revival since the 1970s. Alphabetical dialect dictionary making was carried out for the greater part outside the academic world.

2.1 *The old alphabetical school*

The aim of the early lexicographers, of which De Bo (1873) is a very interesting example, was to demonstrate the richness of the dialect vocabulary and to show that it even had the potential to become the basis of a Flemish standard language. The Flemish, however, in the end adopted the Dutch standard of the Netherlands, which

developed in so-called *Belgian Dutch*, a natiolect¹ within the Dutch standard language. Anyway, the early lexicographers were all motivated by both the romantic and political attitudes of the 'Flemish movement', that wanted to restore Dutch as an official language of Belgium, which had at that time French as the dominating language.² Another important motivation was also that the Dutch words only used by Flemings should not be left out of the *Woordenboek der Nederlandsche Taal* (WNT, Dictionary of the Dutch Language), the major comprehensive dictionary for the Dutch language.³

2.2 The alphabetical school of the dialect revival

From the 1960s onwards, dialect loss gained impetus because of the increase of both geographical and social mobility, the rise of the average schooling level of the population, and the introduction of the mass media. Dialect loss is often felt as loss of regional or local identity — an identity that is under the pressure from globalisation. This feeling, combined with a more relaxed attitude with regard to standard Dutch, led to a so-called dialect revival, which is not so much a revival of the traditional dialect itself, but rather the emergence of an explicit popular interest in it.

The dialect revival resulted in a second wave of dialect dictionary making from the 1970s onwards. The ensuing local dictionaries⁴ are mainly made by dialect enthusiasts, who are less motivated by political or scientific aims, as was the case with the older - and mostly regional (not local) — dialect dictionaries.⁵ In many cases, the

¹ A natiolect is a national variety of a standard language which is spoken in more than one state; in Belgium there are three natiolects: Belgian Dutch (often called *Flemish*), Belgian French and Belgian German. The two latter language varieties do not have a name of their own.

² French was spoken in the south of Belgium (the Walloon provinces) and by the upper class in the towns of the Dutch speaking part of the country. Although French in Belgium was (and is) spoken a large minority (nowadays about 40%), it was the only official language of Belgium until 1898, and dominated culturally very well after that. The city of Brussels turned from Dutch to French in the 19th century, and has nowadays officially a bilingual French/Dutch status.

³ The first volume of the WNT was published in 1864, the last one in 1998.

⁴ By local dialect dictionaries, we mean dictionaries that have a limited geographical scope.

⁵ Dialect dictionaries compiled by amateurs have already existed since at least 1836 with the work of Hoeufft for the town of Breda. Such initiatives, however, remained relatively rare until the 1970s.

making of a local dictionary is the work of amateurs without proper linguistic training. Hence, the quality of them varies considerably.⁶

Nearly all amateur dialect dictionaries focus on the contrastive part of the dialect lexicon; dialect words that lexically correspond with the standard language, and only differ in phonology, sometimes are even not considered as dialect words proper. The alphabetical arrangement of the entries is nearly always taken for granted. Many amateurs render the headwords in a home-made dialect spelling, forgetting that by doing so, only good dialect speakers will be able to retrieve words quickly. Most dictionaries lack a standard language to dialect index, which would make these word collections accessible for non-dialect speakers, nowadays the greater part of any local community. The use of traditional dialects is now reserved for the older generations.

In Van Keymeulen & Oosterhof (2009) a plan is proposed to aggregate the products of the amateur dialect lexicography into one database, and enrich the dialectal entries with 'dutchifications' and translations into standard Dutch (SD). To dutchify a dialect word is to spell it as if it existed in SD, by replacing the dialect phonemes with their SD counterparts e.g. dialect *mottejêr* > SD *mottenaarde*. The translation into SD of *mottejêr*, however, is *potaarde* 'potting compost'. By adding translations, the dialect dictionary is so to speak reversed into a bilingual SD to dialect dictionary. The enrichments serve as search terms, and the database is thus opened up for non-dialect speakers and scientific research.

Since 2009, the whole project is now well under way at the Ghent University, thanks to the many volunteers who prepare the digital texts and add the enrichments (see www.woordenbank.be). The database is still under construction.

⁶ It is striking that nearly all 'amateur' dictionaries are made by men, hardly ever by women. An important exception, however, is the WZD (Dictionary of the Zealandic Dialects), a regional dictionary, which was even the first dictionary to meticulously indicate the location of every word, which since then has become a scientific pre-requisite for dialect lexicography.

Woordenbank van de Nederlandse Dialecten

HOME ZOEKEN DIALECTWOORDENBOEKEN VRIJWILLIGERS LINKS PARTNERS DISCLAIMER CONTACT

Woord in tekst:
 A.N.-zoekterm:
 Vernederlandsing:
 Origineel trefwoord:

☐ Exact zoeken

☒ Kies alle beschikbare woordenboek(en) ☐ Kies provincie(s)

Woordenboek: ?
☒ Alle woordenboeken
☐ Antwerpen (AN)
☐ Asse (VB)
☐ Dadizele (WV)
☐ Erpe-Mere (OV)
☐ Gent (OV)
☐ Hageland (VB)
☐ Izegem (WV)
☐ Leuven (VB)
☐ Oosterzele (OV)
☐ Torhout (WV)
☐ West-Vlaanderen (WV)
☐ Zuid-Oost-Vlaanderen (OV)
☐ Aalst
☐ Antwerpen, Bijvoegsel (AN)
☐ Berlaere (OV)
☐ Diepenbeek (BL)
☐ Frans-Vlaams Woordenboek (FV)
☐ Gooik (VB)
☐ Haspengouw (BL)
☐ Kortrijk (WV)
☐ Okegem (OV)
☐ Roeselare (WV)
☐ Waasland (OV)
☐ Zeeland- Supplement (ZL)
☐ Antwerpen - Aanhangsel (AN)
☐ Anzegem (WV)
☐ Beveren
☐ Diest (VB)
☐ Geel
☐ Hageland - Bijvoegsel (VB)
☐ Iddergem (OV)
☐ Lede (OV)
☐ Oostende (WV)
☐ Schellebelle
☐ West-Vlaanderen - zakwdb (WV)
☐ Zeeland (ZL)

Figure 1. Search window of www.woordenbank.be

In 2015, the Meertens Institute in Amsterdam followed the Ghent example with the eWND (<http://www.meertens.knaw.nl/ewnd/>), which covers the local/regional lexicography for the Dutch dialects of the Netherlands.⁷

2.3 Attempt at a comprehensive alphabetical dictionary by Ludovic Grootaers

In the overview about alphabetical dialect lexicography above, we have not mentioned the efforts of prof. Ludovic Grootaers in this respect. The reason is that Grootaers' attempt never resulted in a real dictionary, but 'only' in one of the largest collections of dialect data for Dutch speaking Belgium, which is kept in the archives of the Catholic University of Louvain.

Ludovic Grootaers (1885-1956) became professor at the Catholic University of Louvain in 1935, after a career of 28 years as a teacher in secondary schools (see Figure 2).⁸ Already in 1922, however, he founded the *Zuidnederlandsche*

⁷ The eWND has in some respects a different approach than the woordenbank.be.

⁸ For a more extensive account of the life and scientific meaning of L. Grootaers, see Grootaers (1950).

*Dialectcentrale*⁹ (ZND, *Dialect Institute for the Southern Netherlands*) at the Louvain University. Professor Grootaers will be remembered, not only as a 'perfect gentleman',¹⁰ but also as an important phonetician and dialectologist. His academic assignment also included the interpretation of German authors. He supervised a large number of dialectological PhDs and master theses and was the driving force of the periodical *Leuvense Bijdragen*.



Figure 2. Prof. Ludovic Grootaers (1885-1956) (Catholic University Leuven).

One of Grootaers' scientific aims was the compilation of a comprehensive dialect dictionary for the Dutch dialects in Belgium, together with a word atlas. To that end, he composed numerous questionnaires that were sent out between 1922 and 1954 to a huge number of dialect speakers (see Figure 3 for an example). Part of the data were used for articles and for the *Taalatlas van Noord- en Zuid-Nederland*

⁹ *Zuid-Nederland* (Southern Netherlands) is nowadays a somewhat obsolete term for Dutch-speaking Belgium, although not for the South of the Netherlands. The first — but short-lived — name of the institute was *Leuvensche Dialectcentrale* (*Dialect Institute of Louvain*).

¹⁰ He was characterized that way by his students (Van Coetsem 1957: 100).

(*Linguistic Atlas of the Northern and Southern Netherlands*),¹¹ but Grootaers' dream could never be fully realised due to lack of financial means. The data he collected, however, are to a large extent incorporated in the three regional onomasiological dictionaries that will be discussed below. There is an urgent need to digitise the completed questionnaires as the paper they are printed on is deteriorating rapidly. Grootaers' son, Willem, was one of the founding fathers of Japanese dialectology and adjusted the theory and methods he learned from his father to the Japanese situation.

KATHOLIEKE UNIVERSITEIT
TE LEUVEN
Zuidnederlandsche Dialectcentrale

Lijst 35.

1901-19 0 54

Bijvoegsel bij de Mededeelingen van de Zuidnederlandsche Dialectcentrale, 1941, Nr. 1-2.

a) Naam van de gemeente waar dit dialect wordt gesproken: *Bambrugge*

b) Provincie: *Oost-Vlaanderen*

c) Naam van den dialectspreker: *Gees Amaat*

d) Leeftijd en beroep: *21/ onderwijzer*

e) Geboorteplaats: *Bambrugge*

f) Adres van den inzender: *Gees Amaat
Heuweg 23
Bambrugge*

N. B. — Lees eerst de aanwijzigingen op de ommezijde.

Hieronder vertaling in dialect:

1. a) «Dat is echt goud».	1. a) <i>da is echt goud</i>
1. b) «Het is echt gebeurd».	1. b) <i>'t is echt gebeurd</i>
2. Hoe heet de sterk geurende reseda (Lat. <i>reseda odoranta</i>) die in vele tuinen voorkomt? Indien het dialect het woord «reseda» gebruikt, wordt dan geen andere naam gebruikt, misschien nog enkel door ouderen van dagen? Duidelijk aangeven welke naam oud is en welke nieuw, indien er twee bestaan.	2. <i>alleen: reseda</i>
3. «Eigenlijk heet hij Frans».	3. <i>eigentlich iet a Frans</i>
4. a) «Hij zit in de ellende».	4. a) <i>e zit in de misere d'ellengne</i>
4. b) «Ik ben ziek van ellende».	4. b) <i>ik ben ziek van ellegne</i>
5. Hoe heet een kanten tusschenzetel in een gordijn, een schort, een kleed, enz. (Fr. <i>entre-deux</i>).	5. <i>en buersel</i>

Figure 3. Example of a ZND-questionnaire (1941) for the East Flemish village of Bambrugge.

¹¹ *Taalatlas van Noord- en Zuid-Nederland* (*Linguistic Atlas of the Northern and Southern Netherlands*), later renamed as *Taalatlas van het Nederlands en het Fries* (*Linguistic Atlas of Dutch and Frisian*), was a major project of the Meertens Institute in Amsterdam.

3. The onomasiological dialect lexicography: Weijnen's school

The southern Dutch dialect area consists of four dialect groups: (1) the Flemish dialects, spoken in French Flanders (France), West and East Flanders (Belgium)¹² and Zeeland Flanders (the Netherlands); (2) the Brabantic dialects, spoken in Antwerp and Flemish Brabant (Belgium) and Northern Brabant (the Netherlands); (3) the Limburgian dialects, spoken in the Limburg provinces of Belgium and the Netherlands; (4) the Zeeland dialects, spoken in Zeeland and Goeree-Overflakkee (the Netherlands).

The dialect vocabulary of the Flemish, Brabantic and Limburgian dialects is collected in three parallel regional dictionaries. The *Woordenboek van de Brabantse Dialecten* (WBD, *Dictionary of the Brabantic Dialects*, 1960-2005) and the *Woordenboek van de Limburgse Dialecten* (WLD, *Dictionary of the Limburgian Dialects*, 1960-2008) are set up according to a plan, conceived by prof. Toon Weijnen (Catholic University of Nijmegen) (see Figure 4): they are onomasiologically arranged and published in thematic fascicles.¹³ The *Woordenboek van de Vlaamse Dialecten* (WVD, *Dictionary of the Flemish Dialects*, 1972 -) started 12 years later than its sister projects; the initiator was prof. Willem Pée of the Ghent University (see Figure 5), who adopted the same methods of collection and presentation as WBD/WLD.¹⁴ Together with the already existing, but alphabetically arranged *Woordenboek der Zeeuwse Dialecten* (WZD, *Dictionary of the Zeeland Dialects*, Hendrika Ghijsen 1964), the three dictionaries cover the whole southern Dutch dialect area (see Figure 6). They describe in principle the vocabulary of the traditional dialects of the first half of the 20th century in the southern part of the Dutch language area, in a joint international and inter-university project. The minimal aim of every dictionary is to record at least once every dialect word / dialect meaning in the area of investigation (lexical aim); the maximal aim is to have one good data for every locality (geographical aim). The last aim, however, is

¹² West and East Flanders are the names for the western and eastern part of the old county of Flanders, not the west and east of present-day Flanders, which eventually became the name for Dutch-speaking Belgium as a whole.

¹³ In the 1990s, an additional editorial board for WBD and WLD was created at the Catholic University of Louvain.

¹⁴ The WBD/WLD also served as an example for the WALD (*Woordenboek van de Achterhoekse en Liemerse Dialecten*), WOD (*Woordenboek van de Overijsselse Dialecten*) and WGD (*Woordenboek van de Gelderse Dialecten*) afterwards.

seldom met, because the editorial boards had to rely on voluntary cooperation for the data collection.¹⁵



Figure 4. Prof. Antonius Angelus Weijnen
(1909-2008) (Catholic University Nijmegen)



Figure 5. Prof. Willem Pée
(1903-1986) (Ghent University)

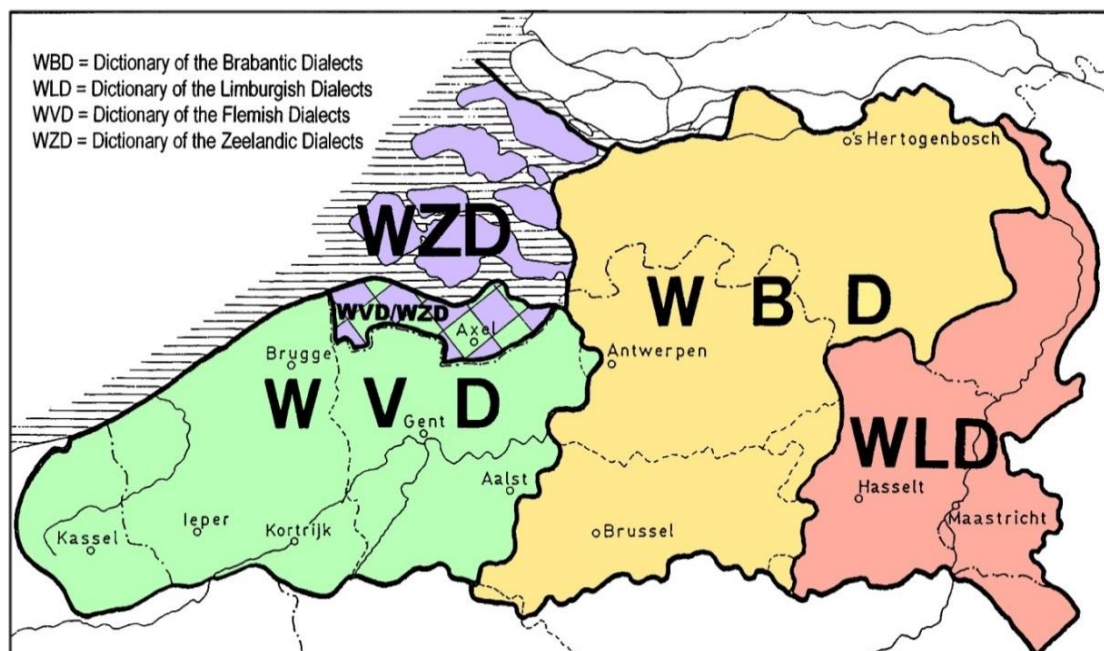


Figure 6. Areas of investigation of WVD, WBD and WLD (+ the alphabetically arranged WZD).

¹⁵ The very ambitious goals of the dictionaries were never provided with sufficient long-term financing.

Weijnen perhaps did not introduce, but he certainly popularised the systematic arrangement in Dutch dialect lexicography. The three above mentioned onomasiological dictionaries are published in a series of volumes, each of them devoted to a certain conceptual field within three parts: I. Agricultural Vocabulary (e.g. 'the farmhouse', 'poultry' etc.); II. Technical Vocabulary ('the cooper', 'the miller' etc.); III. General Vocabulary (e.g. 'plants', 'birds', 'human character' etc.).

The lexicological data in the dictionaries are derived from two main sources: direct investigation by the editorial boards by way of written (thematically organised) questionnaires and existing written material (published or unpublished). The collection of Grootaers is the main part of the 'older' material that is incorporated in the dictionaries.¹⁶ Grootaers collected his data some 50 years earlier than the editorial boards of the dictionaries, hence it is possible to base chronological comparisons on his collection.

As figure 7 below shows, each dictionary article¹⁷ consists of a standard Dutch 'title', followed by an encyclopedic / semantic description of the concept denoted by it, a source list and the different dialect words (in bold) which can be used to refer to the concept.¹⁸ Every dialect word is followed by (general) indications as to its frequency and geographical distribution (details with regard to phonetics and location are stored in a database). More details as to the history and the arrangement of the three dictionaries, and to the way the phonetic component of the data is handled, can be found in Kruijsen & Van Keymeulen (1997).

¹⁶ For an overview of the sources of the three dictionaries, we refer to the introductions of the dictionaries (and the introduction of each volume).

¹⁷ In the tradition of the three dictionaries, a dictionary article is called a *lemma*, which normally is the term for an entry / headword.

¹⁸ Weijnen coined the useful term *heteronyms* for (dialect) words which mean the same in different (dialectal) language systems; synonyms only occur within the same language system.

schoteltje springen [*scheuteltje* —]:
Wdb: De Bo: schutteltje springen i.v.
schuttel.

sletsje springen: Div: De Cock en
Teirlinck I: sletske-springen, Meerbeke.

— 2.6. Toestelspelen

— 2.6.1. Grote toestellen

SCHOMMEL

Speeltuig bestaande uit een plankje dat
tussen twee neerhangende touwen
bevestigd wordt. Men kan op het plankje
heen en weer zweven door zich af te zetten,
of doordat men geduwd wordt. Zie ook het
lemma SCHOMMELEN, blz. 323. Zie afb. 107.
Teirlinck geeft de benaming *wip* [*wiep*]
(wiepe) voor een schommeltouw. In Rum-
beke vermeldt men het woord *renkoord*,
een touw waaraan men zich heen en weer
kan slingeren, meestal bevestigd aan een
hoge boomtak.



Figure 7. Flemish words for the swing (WVD III,6:319).

WVD 123 (2000), 6; WVD 156 (2005), 90;
WVD 166 (2006), 46; DC A'dam 19 (1951),
3a-b-c; SND 2007 (2007), 1; Bruylant
(2004), 281; Vanmaelsaete (2006), 313.
ZND B1 (1923), 130; ZND 14 (1926), 28;
ZND 32 (1939), 104. Biekerf 81 (1981), 148;
't Wingheroen 20 (1996), 41; Biekerf 97
(1997), 323. De Cock en Teirlinck IV (1902-
1908), 165-166; Anzegemse spelenkroniek
(1992), 64.

bak: Roeselare.

balançoire [*Fr. uitpr., balançoire*]:
freq. FV oost en str. v. Ieper-Poperinge;
omg. Geraardsbergen, ook Rekkem en De
Panne.

♦ zeldz. WV zuid, ook Deinze en
Geraardsbergen. Wdb: Lievevrouw-Coop-
man: balançoire.

balanceur: ♦ Poelkapelle.

balansier: ♦ Geraardsbergen.

bies: Baasrode, Dendermonde, Herder-
sem.

♦ spor. Waas. en str. v. Dendermonde,
ook Baasrode en Herdersem.

biezebijs [ook: *bijzebijs, biezebijs*]:
freq. OV.

♦ Gent, Aalst, Haasdonk. Div: De Cock en
Teirlinck: Belsele, Herdersem.

bijs: freq. OV, L. v. Hulst en ZV
grenastr.

♦ alg. OV. Wdb: WZD i.v. touter: L. v.
Hulst; WZD (sup.): L. v. Hulst, OZV.
Div: De Cock en Teirlinck: Geraardsber-
gen, Denderhoutem, Okegem, Liedekerke,
Denderleeuw, Welle, Moorsel, Baardegem,
Herdersem, Lebbeke, Massemen-Westrem,
Sint-Lievens-Esse, Mariakerke, Gavere,
Semmerzake, Sint-Martens-Latem, Zulte,
Hamme (Waas), Eksaarde.

bijskop: ♦ Mere.

bijzer [ook: *bijzder*]: Eede.

An onomasiological arrangement of the word collection has the advantage that the material for word maps is immediately at hand.¹⁹ The arrangement bears witness to the scientific interests of Weijnen (and of dialectology as a discipline at his time). To put it bluntly, the three dictionaries, notwithstanding the name of the projects, are not dictionaries, but highly ordered collections of word usage fit for geographical mapping; they are in fact atlases. It is indeed the case that the semantics (e.g. polysemic structures) can at best be described in alphabetical dictionaries, which start from 'words' and not 'concepts'. Weijnen's approach also entailed that the focus of the dictionaries is on content words like substantives, adjectives and (full) verbs.

¹⁹ Another advantage is that it is inspiring for etymological research, because very often different word forms for the same etymon are brought together under same concept.

4. Aggregating three databases: The Dictionary of the Southern Dutch Dialects (DSDD)

Since the end of the 1980s / beginning of the 1990s, the editorial boards of the three dictionaries moved to digitisation, instead of processing all the data manually. Thus, three comprehensive databases came into existence.²⁰ Thanks to the parallel set up of the dictionaries, it must be possible to aggregate the three databases in order to be able to draw word maps of the whole of the southern Dutch dialect area (minus Zeeland, that is).²¹ For an example of it, see below the word map for 'swing'.

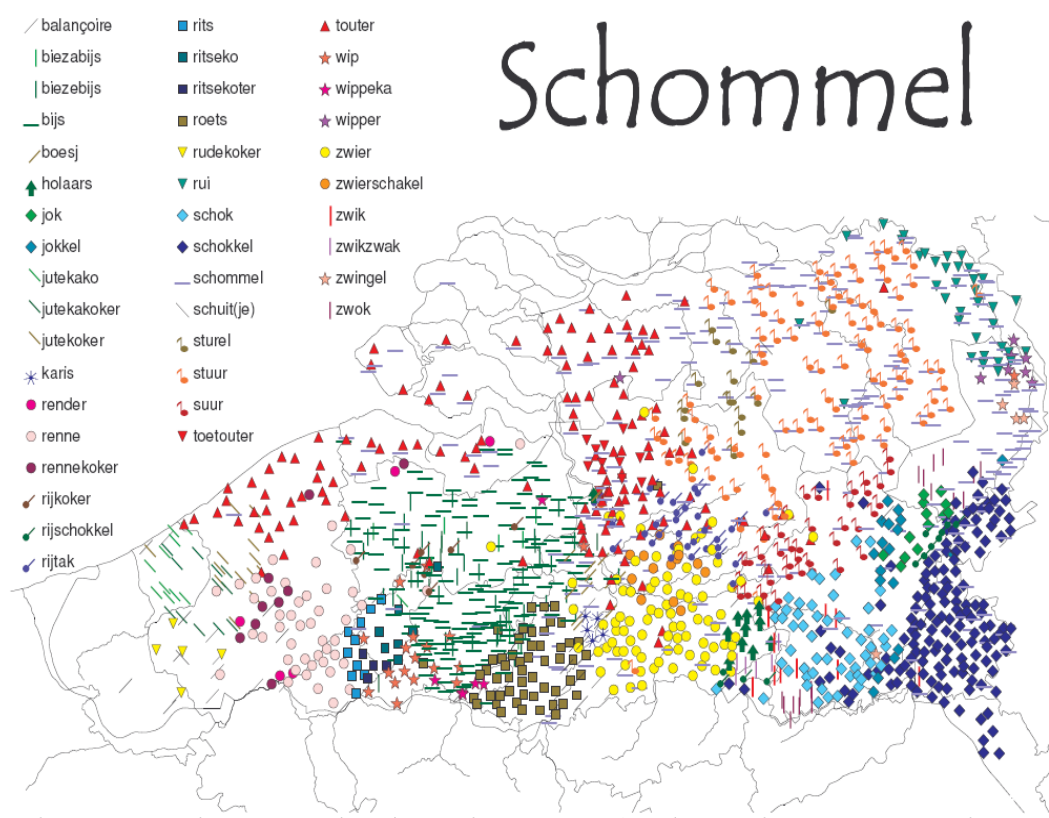


Figure 8. Word map for 'swing', based on the data of WVD, WBD, WLD (+ WZD), drawn by Tineke De Pauw (2007), co-editor of the WVD.

²⁰ The data of the 'manual' period were all retro-digitised and put into the different databases.

²¹ The incorporation of the data of the WZD is in principle possible because the words have been geo-located. Since the dictionary is arranged alphabetically, it has to be re-arranged. This will be done — hopefully — after the aggregation of WVD-WBD-WLD.

To this end, a medium-scale research infrastructure project of the Flemish Research Foundation was launched in 2016. A consortium was formed that consisted of dialectologists, geomaticians²² and computer scientists. At Ghent University, the Department of Dutch Linguistics, the Ghent Centre for Digital Humanities, the Internet Technology and Computer Science Lab (IDLab) and the CartoGis Research Group are involved. In particular, dialectologists from Ghent University work closely with the Ghent Centre for Digital Humanities (GhentCDH) to prepare the ground for the aggregation of the databases and their exploitation via a Virtual Research Environment (VRE) for digital lexicological research. The CartoGis Research Group is sharing their cartographic knowledge. The Ghent team will work closely with the *Instituut van de Nederlandse Taal* (INT, Dutch Language Institute) with regard to the technical and linguistic sustainability of the DSDD. Through this collaboration, interoperability with CLARIN²³ will also be ensured. The DSDD is additionally a pilot project of DARIAH-BE Belgium.²⁴

Other members of the consortium are: the Catholic University of Leuven, the Radboud University in Nijmegen, the University of Groningen and the Meertens Institute. The project is led by prof. Jacques Van Keymeulen (Ghent University) and is carried out by Veronique De Tier, Roxane Vandenberghe en Lien Hellebaut (Ghent University), with the help of Sally Chambers (GhentCDH), together with Katrien Depuydt, Jesse de Does and Tanneke Schoonheim of the Dutch Language Institute.

The aims of the DSDD project are:

- (1) to aggregate the three regional dialect dictionaries (WVD, WBD and WLD) into one single database;
- (2) to provide access to the aggregated DSDD dataset via a user-friendly website;
- (3) to enable the export of the aggregated dataset for analysis using existing research tools;

²² Geomatics can be described as the collection, interpretation, integration and management of geographic data in general.

²³ CLARIN, the *Common Language Resources and Technology Infrastructure*, is a European research infrastructure which “makes digital language resources available to scholars, researchers, students and citizen-scientists from all disciplines, especially in the humanities and social sciences, through single sign-on access”. Further information is available on the CLARIN website: <https://www.clarin.eu/>.

²⁴ DARIAH is a pan-european infrastructure for arts and humanities scholars working with computational methods. It supports digital research as well as the teaching of digital research methods (see www.dariah.eu).

(4) to demonstrate the potential of the DSDD for digital scholarship with 2-3 research use cases. The research questions that come to mind and could be answered on the basis of the aggregated database are (amongst others):

(a) What systematic lexico-geographical patterns do the southern Dutch dialects show? Do they coincide with the traditional ones, based on phonology?

(b) How can the different levels of heteronymy in different conceptual fields be explained?

(c) Are there geographical patterns in semantics?

(d) What will be the results of cluster analysis and exploration of the linkage (and visualisation) of linguistic data with synchronic and diachronic extralinguistic data of all kinds?

5. The state of affairs of the DSDD

The DSDD project, which started in January 2017, is organised in three phases: 1) design and preparation, 2) implementation and 3) exploitation. As part of the first phase, the majority of the work focuses on data preparation, in particular selecting the concepts for standardisation and alignment.

5.1 Concept selection and standardisation

Despite the aim to develop the dictionaries according to the same approach, as they were compiled over a period of several years, at different places (Nijmegen, Leuven, Ghent) and by different editors, a large number of inconsistencies arose over time (Van Keymeulen 2004). In order to create the aggregated DSDD database, a number of standardisation activities had to be carried out before the selected concepts could be aligned. To investigate these problems, 1500 concepts from different thematic volumes of the dictionaries, e.g. fauna and flora, human body, children's games, school etc. were selected, to make an inventory of the issues to be resolved.

The principle behind the alignment is to select a concept (lemma title e.g. swing, known as the ‘DSDD concept’), which will be used as the title of the dictionary article (lemma) in the aggregated dataset. All the related keywords and lexical variants can then be aligned with the relevant ‘DSDD concept’. For some concepts, the mapping between the concepts, keywords and lexical variants is quite straightforward, for example, where the concepts and the keywords are the same in all three dictionaries. However, for some concepts, different editorial choices had been made between the different dictionaries. For example; in one dictionary, two separate lemmas have been created, e.g. “black beetle” and “blue beetle”, whereas in another dictionary there is one single concept, e.g. “beetle”.

The challenge is how to combine these concepts within the DSDD. For example, the “beetle” could be chosen as the ‘DSDD concept’ and then the “black beetle” and “blue beetle” (and their related dialect words) could be simply linked with “beetle”. However, it could be argued that this reduces the granularity of the aggregated dataset and therefore constitutes ‘data loss’. Where inconsistencies appear at the ‘lower’ levels of the dictionary (e.g. keyword, lexical variant), the situation becomes even more complex. For example, where different choices between what is a ‘keyword’ and what is a ‘lexical variant’ have been made.

The aim of this inventory was to identify and document the different types of issues that occur in the data, including details of how they had been resolved. In this way, if similar issues arise at a later stage of the alignment, the editorial choices made can be referred to and reused, to increase the consistency across the dataset. Following this evaluation and standardisation phase, the ‘real’ concept alignment could begin.

5.2 Data import and concept alignment

In the original project proposal, it was anticipated that a concept-alignment tool would need to be developed. However, thanks to the collaboration with the Dutch Language Institute via CLARIN, their existing Lex’it tool could be adapted for the DSDD project. Lex’it is a tool for creating links between different entities (e.g. concepts,

keywords, lexical variants) between lexicographical datasets (in the case of DSDD, between the three dialect dictionaries).

For the concept alignment phase, firstly, the lexicographical data needed to be exported from the three databases. These exports (two of them as csv-files, the third one as an oracle database dump), or data-dumps, then needed to be 'imported' into Lex'it for alignment. To ensure the integrity of the underlying (source) datasets, they are stored separately within Lex'it and then 'merged' into an integrated 'DSDD dataset'.

At the beginning, only 100 concepts were chosen at random from the three different dictionaries to explore the different needs and possibilities for the alignment. Then the list of concepts was enlarged to another 1500 concepts. These 1500 concepts were used to link the dictionary articles. When we select a concept, e.g. swing, then the programme searches automatically for dictionary articles with the name 'swing' in the three dictionaries. These separate articles can then be linked to the matching DSDD lemma 'swing'. If however, the titles of the dictionary articles are not similar, you may have to extend your search in other fields, such as definitions or questions, to find the same concepts in the different dictionaries.

At the time of writing more than 2500 concepts (the target), have been aligned to create new 'DSDD concepts'. Currently, the underlying alignment, at the level of keywords and lexical variants, needs to be tackled. In order to do this, it is sometimes necessary to solve some editorial problems. As said before, the dictionaries or databases were composed over a very long period of time (1962-2018), at different places (Ghent, Leuven, Nijmegen), by different editors.

Some examples of inconsistencies are e.g. 1) diminutives, 2) binders between two parts of compounds or 3) different insights in dutchifying keywords, such as:

1) two of the dictionaries make a difference between a simplex and the diminutive; the third one combines both forms into one keyword: *schommel* / *schommeltje* (swing / little swing) <--> *schommel(tje)* ([little] swing)

2) the three dictionaries do not always use the same form when using binders in a compound noun (-s-, -n-, -e-), e.g. binder -en- or -e-: *paarde(n)bloem* <-->

paardenbloem ‘dandelion’; *hond(s)draf* <--> *honddraf* / *hondsdrاف* ‘ground ivy’ as separate keywords; *keel-*, *kelebeier* <--> *keelbeier*, *kelebeier* ‘Adam’s apple’

3) sometimes one editor has chosen a particular word as keyword, while another editor has chosen this word to be a lexical variant, e.g. *duvel* (keyword) <--> *duivel* (keyword) with *duvel* as a lexical variant.

examples of inconsistencies			
	<u>diminutives</u>	<u>binders</u> -s- /-n- /-e-	<u>different insights</u> in <u>dutchifying</u>
WVD	<i>schommel(tje)</i>	<i>paarde(n)bloem</i> <i>hond(s)draf</i> <i>keel-, <u>kelebeier</u></i>	<i>duvel</i>
WBD	<i>schommel</i> <i>schommeltje</i>	<i>paardenbloem</i> <i><u>honddraf</u></i> <i>hondsdrاف</i> <i>keelbeier</i> <i><u>kelebeier</u></i>	<i>duivel</i>
WLD	<i>schommel</i> <i>schommeltje</i>	<i>paardenbloem</i> <i><u>honddraf</u></i> <i>hondsdrاف</i> <i>keelbeier</i> <i><u>kelebeier</u></i>	<i>duivel</i>

Figure 9. Some examples of inconsistencies in the aggregated database

5.3. Implementation and exploitation

Now that the concept alignment phase is underway, the preparation for the exploitation phase can begin. The aim of this phase of the project is to develop a Virtual Research Environment (VRE) for digital lexicological research, which will: a) make the newly aggregated DSDD available via a user-friendly website and b) enable the DSDD for digital scholarship. To facilitate this, an Application Programming Interface (API) will be created to enable the export of the integrated DSDD for analysis using existing digital research tools.

In the near future a user-friendly website will be created with search facilities and cartographic tools. As to the cartographic tools, they should be able to automatically generate linguistic maps: an API will feed the data from the database into the mapping tool.

The DSDD project has expressed some linguistic needs with regard to cartography. Dialect maps should be highly flexible: it must be possible for researchers to customise the map according to their research needs. Researchers should be able to select and provide their own base maps with different regions (e.g. a province, larger or smaller dialect areas, ...) and to use orientations such as cities and rivers as they prefer. It must also be possible to manually cluster keywords, and to add some interpretation to the dialect map. This also includes the possibility to change the legend of the map and to select the colours of the symbols.

Using for example Google maps or other systems for a website is the easiest way to visualise the dialect words, but researchers also need a print-quality version of the maps e.g. for publications and presentations.

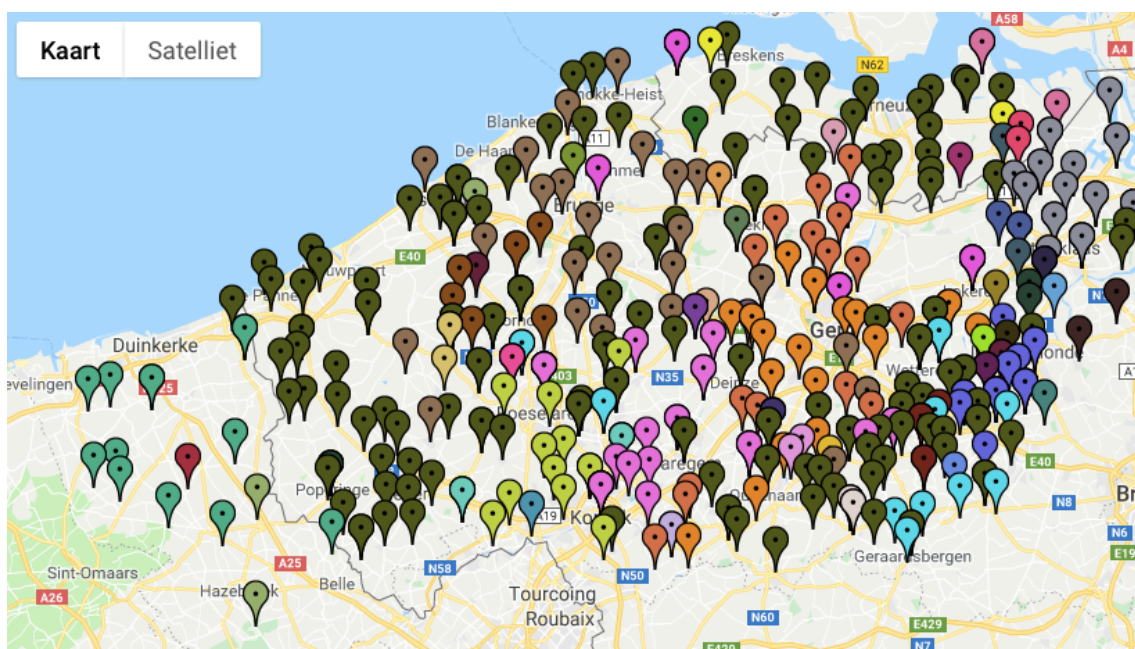


Figure 10. Google map visualising dialect words for 'tadpole' in the digital Dictionary of the Flemish dialects: e-wvd.be.²⁵

²⁵ The legend of the map is absent here, but can be consulted at e-wvd.be with the search term *kikkervisje* 'tadpole' and see Fig. 11 below.

For the last decade, the editors of the three separate dictionaries have been using the commercial software Mapinfo in an adapted form: symbol maps are automatically generated from different base maps on basis of place codes (see the maps for ‘tadpole’ and ‘swing’). This worked perfectly well for the offline creation of dialect maps, but the current set-up does not meet online needs of 21st century dialectology. Therefore, the editors are looking for a new solution to geovisualise the integrated dataset of the Southern Dutch Dialect area. For this we are exploring a number of solutions. As for Google maps, the disadvantage is that it cannot easily be adapted by end users. The DSSD will therefore explore the different possibilities concerning linguistic maps that already exist in other dictionary projects.

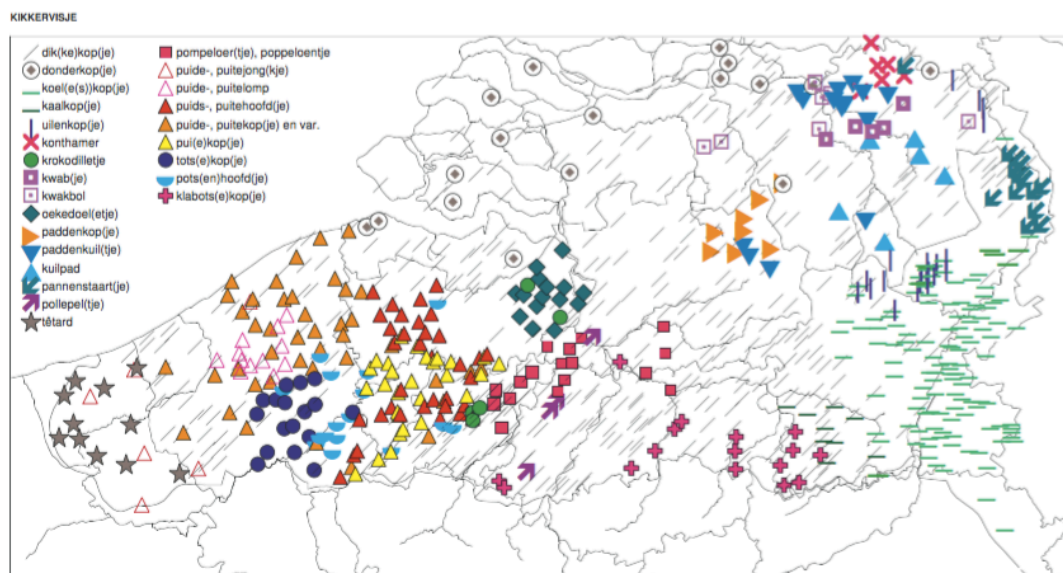


Figure 11. Mapinfo map visualising dialect words from the three dictionaries for ‘tadpole’.

To facilitate the design of the research use cases, an interdisciplinary workshop will take place in March 2019 to bring together researchers from linguistics, digital humanities and cartography to explore how recent advances in geo-visualisation, geo-humanities, web-mapping and digital cartography, can be applied to dialectology. It is intended that as a result of this validation phase, the newly integrated DSDD, as a state-of-the-art linguistic dataset can be opened up, enabling data-driven research (and teaching) in dialectology.

6. Conclusion and directions for future research

The sustainability of the DSDD dataset and the related website has been considered from the outset. It is intended that the DSDD will be made available as a sustainable service beyond the lifetime of the project. To facilitate this, a data management plan for the DSDD is in the process of being created in liaison with the Ghent University Faculty Library of Arts and Philosophy. This will be regularly reviewed and where necessary updated throughout the project. The DSDD website and related research tools will be hosted at the Dutch Language Institute as part of the CLARIN infrastructure.

The DSDD is also a DARIAH-BE affiliated project. Once developed, it will be offered as a Belgian contribution to DARIAH-EU and as a result made available via the DARIAH Marketplace, which will considerably increase the visibility at the international level.

The DSDD should be seen as a first phase in the inter-linking of Dutch dialect data. Subsequent phases include: the integration of the *Woordenboek der Zeeuwse Dialecten (WZD)* into the DSDD (phase 2), the integration of the DSDD into the planned cumulative database for all onomasiological dictionaries of the entire Dutch area (phase 3), which could be linked with the alphabetical dialect dictionaries (Woordenbank and eWND) to create a complete overview of the dialect vocabulary of the Dutch speaking area (phase 4).

References

- DE BO, L. (1873) *Westvlaamsch idioticon*, Brugge: Gailliard ([Reissued and enlarged by J. Samyn (1890-1892)] Gent: Siffer; Reprinted in 1970, 1976, 1984, Handzame: Familia et Patria).
- GHUISEN, H. (1964) *Woordenboek der Zeeuwse Dialecten*, Den Haag: Van Goor. [Fraanje, K. e.a. (2003), Supplement *Woordenboek der Zeeuwse Dialecten*, Krabbendijke: Van Velzen.]
- GROOTAERS, L. (1950) *Album L. Grootaers. Album aangeboden aan Prof. Dr. L. Grootaers, hoogleraar aan de Katholieke Universiteit te Leuven bij zijn vijf en zestigste verjaring*, Leuven: s.n.

- HOEUFT, J.H. (1836) *Proeve van Bredasch taal-eigen of Lijst van eenige in de stad en den lande van Breda gebruikelijke en in sommige oorden van ons vaderland min gewone woorden en spreekwijzen, verzameld en toegelicht*, Breda: Sterk.
- KRUIJSEN J. & J. VAN KEYMEULEN (1997) "The Southern Dutch Dialect Dictionaries", *Lexikos*, 7, 207-228.
- VAN COETSEM, F. (1957) "Ludovic Jean Joseph Grootaers (Tongeren, 9 augustus 1885 - Leuven, 12 oktober 1956)", *Jaarboek van de Maatschappij der Nederlandse Letterkunde*, 100-105.
- VAN KEYMEULEN, J. (2004) "Trefwoorden en lexicale varianten in de grote regionale woordenboeken van het zuidelijke Nederlands (WBD, WLD, WVD), in J. De Caluwe, G. De Schutter, M. Devos & J. Van Keymeulen (ed.), *Taeldeman, man van de taal, schatbewaarder van de taal*, Gent: Vakgroep Nederlandse Taalkunde Universiteit Gent – Academia Press, 897-908.
- VAN KEYMEULEN, J. & A. OOSTERHOF (2009) "Local dialect dictionaries and a proposal for a dictionary of the Dutch dialects", in C. Gooskens, A. Lenz & S. Reker (eds.), *Low Saxonian dialects across borders: Synchrony and diachrony*, Stuttgart: Franz Steiner, 109-124.

Glossaries

- 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; electronic edition: <http://e-wbd.nl>
- WGD = Giesbers, C./ H. Scholtmeijer (2005 -) *Woordenboek van de Gelderse Dialecten*, Utrecht: Matrijs.
- WLD = Weijnen, A., J. Goossens e.a., (1983-2008) *Woordenboek van de Limburgse Dialecten* (1983-2008), Assen/Maastricht: Van Gorcum; Groningen/Utrecht: Gopher; electronic edition: <http://e-wld.nl/>
- 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; electronic edition: <https://e-wvd.be>

WVD III = De Pauw, T. M. Lefebvre en J. Van Keymeulen (2008) *Woordenboek van de Vlaamse Dialecten, Deel III, 6, School en Kinderspelen*, Gent: Academia Press.

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

Websites

www.dialectloket.be

www.e-wvd.be

www.e-wbd.nl

www.e-wld.nl

www.wvd.ugent.be

www.woordenbank.be

www.meertens.knaw.nl/ewnd/