

Better Nicely Linked than Poorly Copied.
Historical and Regional Dictionaries of Dutch Digitally United
Tanneke Schoonheim, Instituut voor Nederlandse Lexicologie, Leiden, Nederland
Veronique De Tier, Universiteit Gent, België

The Woordenboek der Nederlandsche Taal (Dictionary of the Dutch Language, WNT) has been freely available online since January 2007 (<http://wnt.inl.nl>). Compared to the original (printed) dictionary, the search facilities have been considerably expanded. For instance, you can now search for a headword using modern spelling, and submeanings and citations can be displayed or omitted on demand.

Another innovation is that headwords in the WNT are now linked to external information, for example, to language maps, pictures and etymological information. At the moment, we add links to the available dialect material, starting with the large dictionaries of the dialects of Flanders, Brabant and Limburg. In this contribution we describe how this is done.

1. Introduction

In the process of dictionary-making, lexicographers use primary sources, reference works and other dictionaries. They often incorporate the information they find into their own dictionary articles, rephrasing it in their own words, not always with satisfactory results. Of course, simple and straightforward information usually causes no problems, but the more subtle distinctions of the original text may get lost in the summary, while a complete quotation of the text, on the other hand, comes with its own difficulties, not the least of which is plagiarism.

Since 2005, the Instituut voor Nederlandse Lexicologie (Institute for Dutch Lexicology) in Leiden has worked at linking information from several external sources to the four large historical dictionaries of Dutch, as part of its dictionary application ‘Historische Woordenboeken van het Nederlands online’ (Online Historical Dictionaries of Dutch).

2. The historical dictionaries of Dutch

The vocabulary of Dutch throughout the centuries has been described in four large historical dictionaries. The *Oudnederlands Woordenboek* (Dictionary of Old Dutch, ONW, completed in 2008) describes the Dutch language passed down from the period of ca 500 to 1200; the *Vroegmiddelnederlands Woordenboek* (Dictionary of Early Middle Dutch, VMNW, completed in 1998) describes the Dutch language passed down from the period of 1200 to 1300; the *Middelnederlandsch Woordenboek* (Dictionary of Middle Dutch, MNW, completed in 1941) describes the Dutch language from the period of ca 1250 to ca 1550; and the *Woordenboek der Nederlandsche Taal* (Dictionary of the Dutch Language, WNT, completed in 2001) describes a large part of the Dutch standard language between ca 1500 and 1976.

3. Digitizing the lexicographical material

In recent years, great progress had been made with the digitization of our lexicographical products. All four INL dictionaries have been made available online, starting with the WNT in January 2007 and followed by the VMNW (May 2007), the ONW (April 2009) and the MNW (December 2009) (woordenboeken.inl.nl). These four historical dictionaries together form the dictionary component of the Geïntegreerde Taalbank (Integrated Language Database, GTB), which allows both separate and combined dictionary searches.

Fig. 1. The search screen of the dictionary application ‘Online Historical Dictionaries of Dutch’

4. External links in the online historical dictionaries

The WNT was the first product of the dictionary application to be made available, in January 2007.¹ In order to improve its user-friendliness, the original text was encoded with structural information, enabling the computer to recognize the principal constituents of each article (headword, header, definition, derivatives) and to present the information in an effective way. The encoding of the structure made it possible, for example, to present the – sometimes highly complex – semantic structure per layer of meaning. As a result, users of the dictionary can follow a very specific track to arrive at the information they are looking for, unhindered by information on all kinds of different usages of the word they may not be interested in. Apart from that, it is possible to view the definitions either with or without their corresponding quotations, improving the perspicuity of the dictionary articles. Users of the dictionary can decide for themselves how much information they would like to be visible.

Other practical matters were taken in hand as well, for example by adding a large number of compounds and derivatives that were lacking in the book and on the CD-ROM; by uniformizing and dating a large number of sources; and by adding a uniformized part-of-speech and a modern Dutch headword. Especially the latter is a crucial element of the application, because the modern headword is the instrument that links interrelated dictionary articles, despite the fact that the headwords of these entries are sometimes spelled completely differently. For example, the modern Dutch headword AAN connects the dictionary entries *aan* in the WNT, *aen* in the MNW, *ane* in the VMNW and *an(a)* in the ONW.

Apart from these improvements, vital to a proper presentation of the WNT on the Internet, a team of volunteers worked on a project called ‘Linking External Information’. The aim of this project, started in 2006, was to improve the usability and topical interest of the WNT by adding relevant information to the entries, primarily etymological information from the *Etymologisch Woordenboek van het Nederlands* (Etymological Dictionary of Dutch, EWN)

¹ Available at <http://wnt.inl.nl>, see also Schoonheim and Van Sterkenburg 2007.

and the *Etimologiewoordeboek van Afrikaans* (Etymological Dictionary of African, EWA); dialect information based on maps from the *Taalatlas van Noord- en Zuid-Nederland* (a linguistic atlas of the Netherlands) by Kloeke and the PLAND (a data bank of plant names in the Dutch dialects); and illustrations from several WNT sources.

Two different linking techniques have been used until now: either from a URL to a web page (e.g. the PLAND) or from a filename to a file on an INL-owned WNT server (e.g. the headword files of the EWA and EWN or the illustrations from the Language Atlas). Via a separate ‘link database’ that contains nothing more than the identification number (ID) of each entry in the WNT and the filename or the URL to be linked to each lemma, these links have now all been added to the WNT online and can be activated from a separate link box in the dictionary entry.



Fig. 2. The link box in the dictionary application ‘Online Historical Dictionaries of Dutch’: links (‘Koppelingen’) belonging to the entry AARDBEI.

After user reviews of the WNT had made it clear that linking external information to the entries would indeed add value to the dictionary, we considered how we could carry this work forward. For example, much of the added information is not only of interest to the users of the WNT online, but would be a valuable addition to other dictionaries as well. After all, users of the ONW, VMNW or MNW may also be interested in the language map of ‘*zeven*’ or in the etymology of ‘*church*’. In other words, this information should be accessible from any of the four dictionaries, which is why in many cases the links will no longer be generated in one specific dictionary, but in a single cumulative list of headwords serving all dictionaries.

5. The regional dictionaries of Dutch

Due consideration was also given to the question which available material would be a useful addition to the historical dictionaries of Dutch. The first thing that came to mind, of course, were the regional dictionaries of Dutch. After all, the current dialects have developed directly from the language as it was spoken throughout the centuries in the Dutch-speaking regions. And although we nowadays make a distinction between dialect and standard language, our dictionaries of Old Dutch and Middle Dutch are actually compilations of historical dialectology. Many words that are currently used only in dialects, actually feature in the ONW, VMNW and MNW. As a matter of fact, even the WNT contains many words that are used only regionally.

The idea of somehow linking the regional and historical dictionaries of Dutch is not a new one. Already in 1998, Magda Devos, as chairman of the ReWo², spoke about a future in which the results from lexicographical research of dialects would be linked to the word material of the INL. Her appeal to the representatives of the regional dictionaries and the INL to work together in adapting the dialectological material to enable its integration into the INL material did not lead to any arrangements towards closer collaboration at the time. The INL and the regional dictionaries are still pursuing their own courses. But that does not mean the words of Magda Devos are not still relevant: ‘We have no good reason to call a language database such as the one being compiled at the INL *the* Language Database of Dutch, unless it gives access to all available information with regard to the dialectical variations within the Dutch language.’³ Now, twelve years down the line, the technical possibilities to link this information to the dictionaries have improved considerably.

The large regional dictionaries of Dutch are the *Woordenboek van de Brabantse Dialecten* (Dictionary of the Brabantic Dialects, WBD, completed in 2005), the *Woordenboek van de Limburgse Dialecten* (Dictionary of the Limburg Dialects, WLD, completed in 2008) and the *Woordenboek van de Vlaamse Dialecten* (Dictionary of the Flemish Dialects, WVD, probably to be completed around 2018). These three dictionaries are thematically arranged and are composed of more or less parallel sets: the vocabulary of agriculture, a variety of professional vocabularies and the general vocabulary. The website of the Radboud University provides all indexes of the WBD (<http://www.ru.nl/dialect/wbd/>) and WLD (<http://www.ru.nl/dialect/wld/>). A couple of years ago, project D-Kwadraat was started to make these dictionaries digitally accessible. All volumes of the WBD and WLD were digitized into text files (agricultural vocabulary) or into a database (General Vocabulary). All this is to be made available through a web interface with a map module that generates maps in Google Earth.⁴

Only a small part of the material of the WVD is digitally available to the public, but in the coming years it will be presented on the Internet in much the same way as the WBD and the WLD. In the meantime, the indexes of all published volumes of the WVD have been made available on www.wvd.ugent.be, with a direct link to the PDF files, which are available for each volume. However, in the coming years the editors of the WVD will give priority to editing the paper dictionaries, because time is running out when it comes to taking stock of dialects. However, they do have Word and PDF files of all published volumes of the WVD at their disposal, as well as database files of the General Vocabulary. Still on the wishlist is the automatic conversion of the old volumes dating from before the computer period into database files. Preliminary research is already being made.

Over the years a lot more has been published in the field of dialect lexicography, of course. However, for the time being we will restrict ourselves to collecting the data of the above-

² The *ReWo* (a permanent consultative body for regional dictionaries) is a cooperation between a number of Dutch and Belgian research institutes and universities where traditional vocabularies are recorded and studied.

³ Devos 1998: 87.

⁴ For further information about D-kwadraat, please refer to the following websites:
<http://www.ru.nl/dialect/d2>
<http://dialect.ruhosting.nl/d2/publications/Digitaalv1.1.pdf>
http://dialect.ruhosting.nl/d2/publications/paper_D2_v1.1.pdf

mentioned regional dictionaries and the historical dictionaries into the application⁵, using the PDF files of the regional dictionaries.

Other thematic dialect dictionaries such as the WALD, the WOD and the WGD can always be included in the project at a later time. These dictionary projects are less extensive than the above-mentioned projects, so it shouldn't be too difficult to include them into the current project, after experience has been gained with the three larger ones. The etymological dictionaries can easily be incorporated at a later time as well.

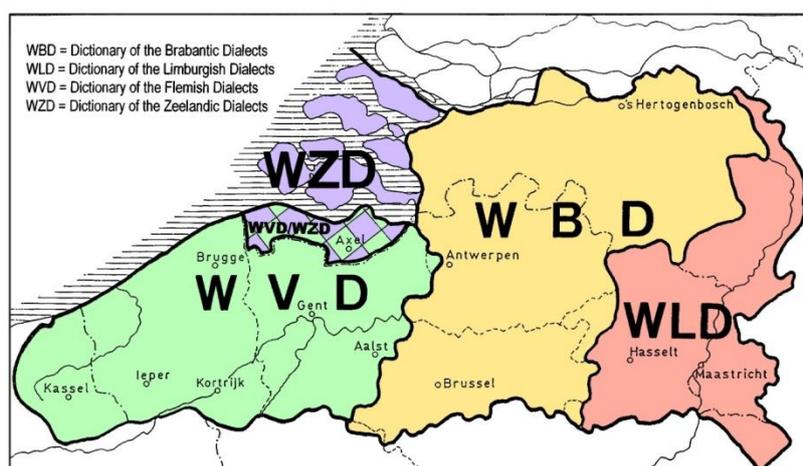


Fig. 3. The regions covered by the WZD, WBD, WVD and WLD (map courtesy of the Department of Dutch Linguistics of Ghent University).

6. Linking the regional and historical dictionaries

When linking information to a dictionary, one option would be to look only in one direction: we have a dictionary and to that dictionary we link whatever may be relevant to it. The way external links are currently implemented in the application serves as an example. However, it would be much more beneficial to look in both directions and make the links as versatile as possible, with two sides profiting from the efforts made. Let me give a few examples of what the results could be if we interlinked the regional and historical dictionaries.

Magda Devos uses the example of the headword *vlinder* ('butterfly') to show how much dialect lexicography could add to the WNT material.⁶ Her wish list is crystal clear: make a connection between the language map of the butterfly and the WNT headword in question, make hyperlinks of the various dialect words for 'vlinder', provide these hyperlinks with the matching information on origin, pronunciation, distribution etc., add sound fragments of

⁵ The following regional dictionaries are also part of the ReWo: the *Woordenboek der Zeeuwse Dialecten* (WZD), the *Woordenboek van de Drentse Dialecten* (WDD), het *Stellingwarfs Woordenboek* (SW), the *Woordenboek van de Achterhoekse en Liemerse Dialecten* (WALD), the *Woordenboek van de Overijsselse Dialecten* (WOD) and the *Woordenboek van de Gelderse Dialecten* (WGD). Not all of these dictionaries have been arranged along the same lines, the main problem being that some of them (WBD, WLD, WVD, WALD, WOD en WGD) have an onomasiological division and others (WZD, WDD and SW) a semasiological one. For the scholarly description of the (dialectal) vocabulary, information from the etymological dictionaries is essential as well, of course. We are referring to the *Etymologisch Dialectwoordenboek* by A.A. Weijnen (2nd edition 2003) and the trilogy on this topic by Frans Debrabandere, the *West-Vlaams etymologisch woordenboek* (2002), the *Oost-Vlaams en Zeeuws-Vlaams etymologisch Woordenboek* (2005) and the *Zeeuws etymologisch woordenboek* (2007).

⁶ Devos 1998: 84-85.

dialect recordings and make a link to encyclopaedic information on the butterfly. Finally, make all publications on the various words for the butterfly visible and if possible directly accessible by means of a link.

A first step in this direction, the digital word atlas set up by Prof. Dr Jacques Van Keymeulen, can be seen on the website of the ReWo. It shows five language maps, with hyperlinked headwords and sound samples, with an underlying etymological article that provides further information on the hyperlinked headwords. A continuation of this work are a number of language maps on food, made available in a project of the non-profit organisation Variaties⁷. Unfortunately, for lack of financial means and time it is not possible (yet) to further develop these projects.

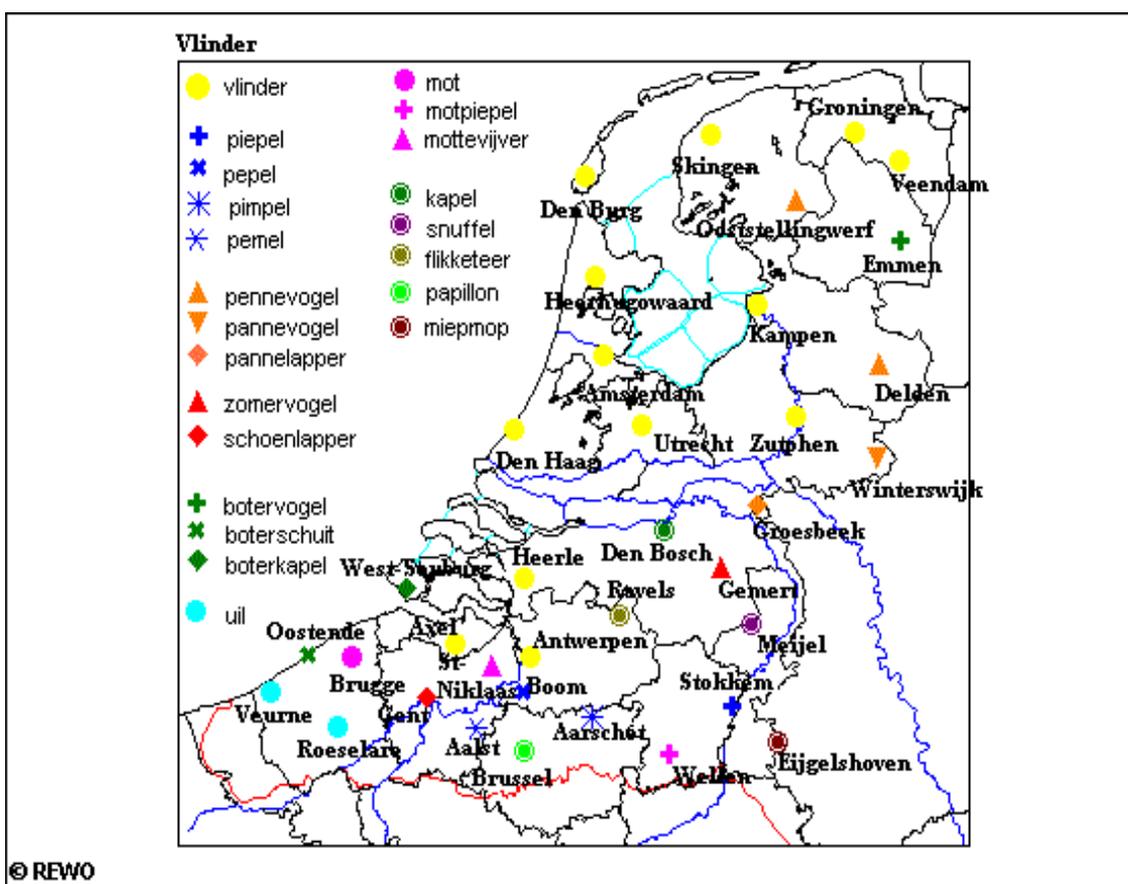


Fig. 4. The ‘butterfly map’ on the website of the ReWo.
(http://fuzzy.arts.kuleuven.be/rewo/Sprek_kaart/Vlinder/vlinderned.htm)

However, a beginning has been made with the linking of the dialect vocabulary to the WNT. At the moment there is actually a link from the article *vlinder* (I) in the WNT to the Speaking Dialect Map of *vlinder* on the website of the ReWo. A link to this map has also been made from the following dialect words that occur in the WNT as (sub)entries: *boterkapel*, *botervogel*, *kapel* (III), *mot* (I), *panne(n)lapper*, *pannevogel*, *pennenvogel*, *pepel* (I), *pimpel* (IV), *schoenlapper*, *uil* (I) and *zomervogel*.

Because users are sent straight to the original website, they can use all functionalities offered there. Apart from that, hyperlinks are provided from the headword *vlinder* and all other headwords under *vlinder* to the PDF files of the WBD and the WLD and there is a link to

⁷ See www.variaties.be.

<http://www.vlindernet.nl/> for encyclopedic information on and illustrations of various types of butterflies. These links have also been added to other relevant entries that have a definition containing the word ‘vlinder’, from *antiek* (I) to *zwijn* (I).⁸ Finally, a link was added that leads to the BNTL⁹, where an overview is given of all publications on the various words for *vlinder*.

Headword(s) GTB	VLINDER (I) # APOLLO # ATALANTA # AVONDROOD # BOTERKAPEL # BOTERVLIEG # BOTERVLINDER # BOTERVOGEL # FLIEFLODDER # enz.
ID-no(s) GTB	M080989 # S003928 # S004735 # M004653 # M010646 # M010669 # M080989 # M010631 # A006768 # enz.
Lemma title ReWo	VLINDER
Illustrations	http://www.vlindernet.nl/
Encyclopedic information	http://www.vlindernet.nl/
Link BNTL	http://picarta.pica.nl/DB=3.4/SET=2/TTL=1/REL?PPN=076565041
Language map	http://fuzzy.arts.kuleuven.be/rewo/Sprek_kaart/Vlinder/vlindernet.htm
WBD	WBD-vlinder.pdf
WLD	WLD-vlinder.pdf
WVD	WVD-vlinder.pdf

Fig. 5. Link table with data belonging to the *vlinder* words in GTB and ReWo.

The VMNW and the MNW contain several headwords that may benefit from a connection with the dialect dictionaries. After the dialect dictionaries have been finished and digitized, the next step would be to provide links from the lemma titles in the regional dictionaries to the relevant (sub)entries in the historical dictionaries. Existing links in, for example, the WNT to the encyclopaedic information and secondary literature can be used as well, implementing the same link table as was used for the linking from the historical dictionaries.

The lemma *vlinder* (‘butterfly’) is an example of a lemma with links to the dialect map. However, survey maps of the three regional dictionaries are not available for every lemma. For that reason, the work is continued by linking not only to the dialect maps on the ReWo website, but also to the lemma titles of the three regional dictionaries. At first an attempt was made to add links for all dialect entries in the WVD, WBD and WLD, but to do that manually would hardly be feasible. Therefore plans are being developed to have automated processes generate the links, at least partly.¹⁰ Among the possibilities are automatic comparison of headword lists or searching in definitions for hyperonyms. The first attempts to achieve this had disappointing (i.e. slow) results. However, the linking to dialect dictionaries should not be postponed indefinitely, so a compromise has been found. Not all dialectal words will be linked, but only the lemma titles of the three dialect dictionaries themselves. The editors of the WVD will provide semi-automatically linked lemma lists of the regional dictionaries. These links will show immediately whether the word occurs in the three various dictionaries

⁸ In this case, the search words *vlinder*, *dagvlinder* and *avondvlinder* (‘butterfly’) had been entered into the search box ‘woord in betekenis’ (word in definition) in the WNT online. This resulted into the following headwords: *antiek* (I), *apollo*, *atalanta*, *avondrood*, *distelvink*, *gehoornd*, *harlekijn*, *hermelijn*, *molenaar*, *non* (I), *oleanderpijlstaart*, *oog*, *paap*, *pannekoek*, *pellerijnvogel*, *piempampoentje*, *piepauw*, *piepeling* (II), *roos*, *rouwmantel*, *stengeldoorn*, *vleek*, *vlichelter(e)*, *vliemel*, *vliervleugel*, *wapendrager*, *weeskind*, *witje* (I) and *zwijn* (I).

⁹ Bibliography of Dutch Language and Literature Studies.

¹⁰ See Landman and Weigand (2007) on this.

or not. Volunteers will then provide a link from these lemma titles to the corresponding headword in the online WNT.

Expectations are that the dialect dictionary lists that are currently available will be ready to be linked in the course of 2010, as some volumes already are. At this moment, virtually the complete general vocabulary has gone through this process. The agricultural vocabulary and the professional vocabularies will follow in due course. After that, if there is still some money and manpower left, the semi-automatic linking to the dialect headwords should also be looked into.

7. Conclusion

Many enthusiastic responses of the users of the dictionary application ‘Online Historical Dictionaries of Dutch’ have made it clear that the concept of linking external information to the various headwords is effective, that providing this kind of information does indeed add value to the dictionaries and that the information often finds its way to an audience that probably otherwise wouldn’t have made the effort to find it but does benefit from it. In order to achieve a balanced system, it is advisable to also link *back* from the regional dictionaries to the historical dictionaries. That is not possible yet, but as soon as the databases of the regional dictionaries have been made available online, it could be quite easily achieved.

Ideally we could even think of one single, large lexicographical application, containing historical dictionaries, the regional dictionaries and the etymological (dialect) dictionaries of the Dutch language, supplemented with encyclopaedic information, visual material – either animated or still images – and maybe even sound recordings. It’s still in the future, but not unachievable. With combined efforts and with the help of technical progress, it should be possible over the next years to develop such a versatile application. Those who will be gaining most, of course, will be the users of such an application, for whom a wealth of information is going to be – literally – a few clicks away.

Bibliography

Dictionaries

- EWA = G.J. van Wyk et al. (eds.). *Etimologiewoordeboek van Afrikaans*. Stellenbosch: Buro van die WAT. 2003. *Supplement*. 2007.
- EWN = M. Philippa et al. (eds.). *Etymologisch Woordenboek van het Nederlands*. 4 Vols. Amsterdam: AUP. 2004-2009. www.etymologie.nl.
- ONW = T.H. Schoonheim et al. (eds.). *Oudnederlands Woordenboek*. Leiden: Instituut voor Nederlandse Lexicologie. 2009. <http://onw.inl.nl>.
- TNZN = L. Grootaers and G.G. Kloeke. *Taalatlas van Noord- en Zuid-Nederland*. Leiden: E.J. Brill, 1939.
- VMNW = W.J.J. Pijnenburg et al. (eds.) *Vroegmiddelnederlands Woordenboek*. Leiden: Instituut voor Nederlandse Lexicologie/Gopher. 2001. <http://wmnw.inl.nl>.
- WBD = A.A. Weijnen et al. (eds.). *Woordenboek van de Brabantse Dialecten*. Assen/Maastricht: Van Gorcum; Groningen/Utrecht: Gopher. 1967-2005.
- WLD = A.A. Weijnen et al. (eds.). *Woordenboek van de Limburgse Dialecten*. Assen/Maastricht: Van Gorcum; Groningen/Utrecht: Gopher. 1983-2008.
- WVD = M. Devos et al. (eds.). *Woordenboek van de Vlaamse Dialecten*. Gent-Tongeren: Michiels. 1979 -.
- WNT = M. de Vries et al. (eds.). *Woordenboek der Nederlandsche taal*. Leiden: Instituut voor Nederlandsche Lexicologie. 1885-2001. <http://wnt.inl.nl>.

Other references

- Devos, Magda. (1999). 'Woordenboeken en dialect'. In: A. Moerdijk; R. Tempelaars (eds.). *Van A tot Z en verder ... Lezingen bij de voltooiing van het WNT*. The Hague/Antwerp. 81-87.
- Landman, Ed; Weigand, Hans. (2007). 'De toekomst van het iWNT'. In: Moerdijk, Fons; Ariane van Santen; Rob Tempelaars. (eds.). *Leven met woorden. Afscheidsbundel voor Professor Piet van Sterkenburg*. Leiden. 169-177.
- Schoonheim, Tanneke; Van Sterkenburg, Piet. (2005). 'Inleiding WNT online'. In <http://wnt.inl.nl> > Help & Info > Over het WNT > Algemeen.
- Schoonheim, Tanneke. (2007). 'Regionale en historische woordenboeken van het Nederlands: samen sterk'. In *Taal en Tongval 20 (Themanummer dialectlexicografie)*. 35-46.