

CHAPTER 13

Making the Dictionary of the Frisian Language Available in the Dutch Historical Dictionary Portal

Katrien Depuydt^a, Jesse de Does^b, Pieter Duijff^c and Hindrik Sijens^d

^aDutch Language Institute, katrien.depuydt@ivdnt.org, ^bDutch Language Institute, jesse.dedoes@ivdnt.org, ^cFryske Akademy, pduijff@fryske-akademy.nl, ^dFryske Akademy, hsijens@fryske-akademy.nl

ABSTRACT

The main goal of the Wurdboek fan de Fryske Taal/Woordenboek der Friese taal-Geïntegreerde TaalBank (WFT-GTB) project was to publish the monumental Dictionary of the Frisian Language (Wurdboek fan de Fryske Taal/Woordenboek der Friese taal, WFT) in the CLARIN research infrastructure, according to open, CLARIN-compliant standards. This has been achieved by 1) curation of the dictionary data, resulting in a well-structured TEI-conformant encoding and 2) publication of the dictionary in the Dutch Institute for Lexicology (INL) dictionary portal, together with the main historical dictionaries of Dutch. The project was carried out by two CLARIN partners, the Fryske Akademy (FA) in Leeuwarden and the INL in Leiden.

The dictionary has been online for more than six years now and has served many users, assisting both researchers and general users interested in the Frisian language. In this chapter we look back on the project and discuss the use of the dictionary application by analysing the retrieval application logs and by providing a use case, exemplifying how the dictionary can be used for historical lexicographical research. The chapter concludes with some suggestions for the improvement and enhancement of the online version of the WFT.

13.1 Introduction

Her Majesty Queen Beatrix of the Netherlands launched the online version of the Dictionary of the Frisian Language (Wurdboek fan de Fryske taal/Woordenboek der Friese taal, WFT) on 6 July 2010. The presentation of the online WFT was the final step in the Wurdboek fan 'e Fryske

How to cite this book chapter:

Depuydt, K, de Does, J, Duijff, P and Sijens, H. 2017. Making the Dictionary of the Frisian Language Available in the Dutch Historical Dictionary Portal. In: Odijk, J and van Hessen, A. (eds.) *CLARIN in the Low Countries*, Pp. 151–165. London: Ubiquity Press. DOI: <https://doi.org/10.5334/bbi.13>. License: CC-BY 4.0

Taal-Geïntegreerde TaalBank (WFT-GTB) project, a project that was supported by CLARIN-NL. The project involved a data curation and a demonstrator project and was carried out by two CLARIN partners, the Fryske Akademy (FA) in Leeuwarden and the Dutch Institute for Lexicology (INL) in Leiden.

The dictionary has been online for more than six years now and has served many users in their research and quest for knowledge about the Frisian language. In this chapter we look back on the project and discuss the use of the dictionary application on the basis of the log files. Furthermore we present a research use case. The chapter concludes with a short outline for improvement and expansion of the online version of the WFT.

13.2 The Dictionary and its Issues

The Dictionary of the Frisian Language is a scholarly, descriptive dictionary of the vocabulary of the Modern West Frisian language from the period 1800–1975. It has its roots in the 19th-century tradition of large dictionaries, and can therefore be compared with the Oxford English Dictionary, the German Dictionary (Deutsches Wörterbuch) and the Dictionary of the Dutch Language (Woordenboek der Nederlandsche Taal).

The dictionary project started in 1938 with the compiling of a corpus, the first volume was published in 1984, and the 25th and final volume was published in 2011. With this dictionary, more than 10,000 pages of lexicographic information about Modern West Frisian is available to the professional linguist and the layperson interested in the Frisian language, a language of about 400,000 speakers, spoken in the Dutch province of Friesland.

The dictionary contains approximately 120,000 lemmas and the entries provide information on the spelling of the headword, its part of speech and its pronunciation. In addition, information is given about the inflexion and etymology of the headword. The semantic section provides the user with information about the meanings of the headwords by means of definitions or translations into Dutch. All the meanings of a word are illustrated by citations, so the user is able to verify the lexicographer's work.

Idiomatic information is given in the idioms section, which contains collocations, proverbs and figurative meanings. The final section of an entry describes compounds and derivatives belonging to the headword.

Five hundred copies have been printed of each volume, and some 400 subscribers received a copy of one volume every year. These subscribers are language enthusiasts and professional linguists, as well as universities and public libraries. The WFT as a paper dictionary has restricted search possibilities; the alphabet is the only means by which the headwords and their descriptions can be accessed. The goal of making the dictionary available online was to enable extensive exploration of the copious linguistic information in the dictionary and to make the dictionary available to a larger audience.

13.3 Integration

Integrating the WFT into the historical dictionary portal of Dutch in the INL was the obvious means of reaching the goal described above and the WFT-GTB project was therefore initiated. The INL dictionary portal describes 15 centuries of Dutch language. It contains four Dutch dictionaries:

- the Dictionary of Old Dutch (Oudnederlands Woordenboek, ONW, ca 500–1200),
- the Dictionary of Early Middle Dutch (Vroegmiddelnederlands Woordenboek, VMNW, 1200–1300),

- the Dictionary of Middle Dutch (Middelnederlandsch Woordenboek, MNW, ca 1250–1550), and
- the Dictionary of the Dutch Language (Woordenboek der Nederlandsche Taal, WNT, ca 1500–1976).

The portal is freely accessible.

The WFT and the Dutch dictionaries were developed in the same lexicographical tradition and integration was feasible because of the similarity in their structures. The advantages of linking the Frisian dictionary with the online Dutch historical dictionaries were many. Connecting the dictionaries enhanced the possibilities for synchronic and diachronic analysis of both languages. To give some examples, the following questions could be researched: which words appear in both languages, and which are specifically Frisian or Dutch? What are the phonological and morphological differences between the two languages? And what is the influence of the Dutch language on Frisian and vice versa? An additional value is that etymological information about Frisian words can be derived from one or more of the Dutch linked dictionaries.

In order to integrate the WFT in the portal, a list of search options had to be drawn up. The starting point was the existing application and the possibilities of the tagged WFT data. Because the WFT and the portal dictionaries are similar both in the information categories they encode and in the dictionary entry structures, the options for searching entries, word senses, quotations, and collocations in the dictionary application are also relevant for increasing the accessibility of the WFT. In fact, it was possible to link most of the information categories in the WFT to the application's existing search options, for example variants of the headword, words in collocations, idioms and proverbs, or languages mentioned in the etymology field.

13.4 Data Curation for the Online WFT

13.4.1 *Repair and Optimisation of the Existing Database*

The original data for the print edition of the dictionary were stored in a database. Since the early 1990s, this has been a BRS/Search database. BRS/Search is a full-text database and information retrieval system which uses a fully-inverted indexing system to store, locate, and retrieve unstructured data.¹ The only metadata added to a dictionary entry were *Word* and *Desc*, where *Word* refers to the headword of the dictionary entry, and *Desc* to a section devoted to the description of a particular word sense within the full text of the entry. No other information categories were tagged explicitly. The data were stored in Windows cp1252 format and marked with layout codes that were used by scripts to convert the database text to rtf documents. The entries of the dictionary were accessible with a search and input interface and a simple text editor.

Before the data could be added to the portal, mistakes and errors identified in the printed dictionary had to be corrected. The data had to be optimised in other ways as well; for instance, abbreviations such as *Id.* and *ibid.* for same author and same source had to be resolved. In a set of compounds with a common first part, the abbreviation marks had to be expanded. Another job was to verify the consistency of cross-references between entries.

13.4.2 *Part-of-Speech Mapping*

The part-of-speech information of the headwords needed to be mapped to the tag set used for the Dutch online dictionaries. For instance, a search query for reflexive verbs in the application uses

¹ See: <https://en.wikipedia.org/wiki/BRS/Search> [access date 16 April 2017].

the standardised category label *ww refl.* (werkwoord reflexief, ‘reflexive verb’) whereas the WFT had the label *v.* (verbum, ‘verb’) with the addition of the Frisian reflexive pronoun *jin* (‘oneself’). Linking the Frisian label to Dutch *ww refl.* enables the simultaneous retrieval of both Frisian and Dutch verbs in this category.

13.4.3 Adding Modern Dutch equivalents

All entries in the integrated Dutch dictionaries are linked to a Modern Dutch headword. Thanks to this link, users have access to the Middle or Old Dutch dictionaries, even if they have no knowledge of older Dutch language stages; for instance, a search query for the Modern Dutch headword PAARD (‘horse’) will yield forms like *pert* (VMNW) and *peert* (MNW). The same goes for users who do have a command of (Modern) Dutch but no knowledge of Frisian. For them, it may be difficult to search for a Frisian entry in the application; hence, a Modern Dutch headword was added to a Frisian lemma.

Not every Frisian lemma can be linked to a Dutch cognate; for instance, Frisian words like *heit* (‘father’, Dutch ‘vader’), *hynder* (‘horse’, Dutch ‘paard’) and *sneon* (‘Saturday’, Dutch ‘zaterdag’) do not have a Dutch cognate.

Although Frisian and Dutch are related languages, the differences between them are substantial. One can therefore assume that cognates like Dutch *neus* (‘nose’) and Frisian *noas* are equivalent; therefore, the Modern Dutch lemma NEUS covers both entries. Both languages also have the word *naad* to mean ‘seam’. In the integrated dictionary, Frisian *naad* can thus be mapped to the Modern Dutch equivalent NAAD. Subsequently, all Dutch compounds and derivatives with *naad-* can be translated into Frisian in the same way. Yet, another meaning of *naad* in Frisian is ‘ridge’, which is not recorded for *naad* in the Dutch dictionaries (the Dutch word for this meaning is *nok*). The equivalent of Frisian *naadfoarst* (‘ridge tile’) would therefore be the non-existent Dutch equivalent NAADVORST. No Dutch, non-Frisian user would use this morphologically correct term to search the dictionaries for the concept ‘ridge tile’.

So, when a Frisian lemma has no Dutch equivalent or cognate, another strategy has to be used to find the correct Frisian entry. Since the definitions in the WFT are mostly Dutch synonyms, a user can enter this synonym in the ‘definition’ search field in the application.

Originally it was planned that only the Frisian lemmas with a known Dutch cognate would be linked to a Modern Dutch lemma, with the intention to show the diachronic similarities and relations between Dutch and Frisian. The assignment of Dutch equivalents to the Frisian headwords was done automatically by selecting the Dutch cognates from the etymology field. When no etymology was known or given, the script searched the field meaning for a one-word definition. Often this resulted in a hit, especially with compound words. It was necessary to check the results of this operating procedure manually.

Through this process about 70% of the Frisian lemmas have been linked to a Modern Dutch headword. On further consideration it is doubtful that only cognates have been linked. Users of the GTB application will find it more beneficial when all Frisian lemmas are linked to a Modern Dutch headword.

13.4.4 Sources and References

It is possible to search the list of citation sources used in the dictionaries. Therefore, a list of sources used in the WFT has been linked to the dictionary. Next to that, a list with references to linguistic literature was created and added to the portal.

13.4.5 Parsing

Writing parsing software in order to tag the logical structure of the dictionary entries caused some difficulties, due to the inconsistencies in the structure. For instance, the etymology section on the

heading starts with the label *Etym.* and the etymological information itself consists of references to cognates and equivalents in other languages, or just references to other languages. However, it can also contain morphological information such as *denominatief van noas* ('denominative of nose'), or *dim. van noas* ('diminutive of nose'), or even a cross-reference to another entry: → *nocht* (→ 'companion').

Headword	Field	Etymology
noas	Etym.	→ N. <i>neus</i> , D. <i>Nase</i> , E. <i>nose</i> .
noasje	Etym.:	Fr., Lat.
noaskje I	Etym.:	denominatief van <i>noas</i> .
noaskje II	Etym.:	dim. van <i>noas</i> ?
noflik	Etym.	→ <i>nocht</i> .

In order to support specific queries, further analysis was needed to distinguish morphological and etymological information.

13.4.6 *Enriching the XML Database with TEI Encoding and Incorporating the Dictionary into the Portal*

In order to incorporate the WFT into the portal, the dictionary data had to be converted to the TEI annotation scheme for printed dictionaries. The existing online dictionary application, which is part of the Dutch Language Bank, allows for querying in more than one dictionary simultaneously. At the time the plans were developed, the challenge was not only to give the user optimal access to the dictionary information, but to do so without compromising the uniqueness of each individual dictionary. All Dutch dictionaries were available in digital form, but in a different encoding system and with a different level of encoding. Their structure had similarities, however, with the presence of the headword; the section with linguistic information at entry level; the section with semantic analysis of the headword; and the section with related entries. TEI encoding for printed dictionaries was chosen as a standard because it allows both fine-grained and coarse-grained encoding. Moreover, all encoding needed for the main Dutch historical dictionaries could be converted to TEI without modifying the encoding scheme, which is more than can be said of competing standards like LMF. A basic encoding scheme for the Dutch dictionaries was defined at INL. This scheme defines a minimum level of mandatory encoding for all dictionaries necessary for the integrated retrieval on the dictionary data. Apart from the basic level of encoding which applies to all dictionaries, the additional encoding present in each of the dictionaries has also been converted into TEI. Consequently, there are some retrieval possibilities applicable to all dictionaries, whereas others are applicable to only one, or a smaller group of dictionaries, depending on the level of encoding. The application of the TEI dictionary encoding scheme to the Dutch historical dictionaries is documented in Depuydt (2010).

13.4.7 *Interoperability in the CLARIN Research Infrastructure*

The WFT dictionary has been integrated into the CLARIN research infrastructure in the following ways:

13.4.7.1 *PID, Metadata and Data Category Registry*

A PID for the WFT resource has been reserved at <http://hdl.handle.net/10032/00-B1C8-4476-53DC-4DED-8>. Metadata can be harvested in CMDI and Dublin Core.

The main grammatical part-of-speech categories used in the WFT have been entered into the data category registry.

13.4.7.2 REST Web Service

The dictionary portal retrieval backend is a REST web service. In this way, the dictionary data can be integrated with other applications besides the portal frontend.

The basic use scenario (Figure 13.1) is extremely simple.

The underlying data resources are queried, under the control of a set of control parameters. Results are sent back to the querying client in the form of a result list or an HTML article display.

13.5 The Users

The historical dictionary portal is an often used resource. More than 53 million requests have been sent to the backend engine since it first appeared online in 2007. Since user requests are logged, we can obtain some information about the behaviour of users from these logs.

13.5.1 Robots Versus ‘Real’ Users.

A first issue in analysing the dictionary usage is that the overwhelming majority of requests do not originate from real users, but from robots such as web indexing spiders. Distinguishing these two categories is not obvious. We have combined two software modules to try to identify the ‘bots’: Bitwalker user-agent-utils (<http://www.bitwalker.eu/software/user-agent-utils>) and the perl module HTTP::BrowserDetect, which both rely on the (unwarranted) trustworthiness of the user-agent information sent by the client. Figure 13.2 gives the proportion of bot requests according to these modules.

<http://gtb.inl.nl/iWDB/search?actie=results&wdb=WFT&domein=0&lemma=wurd>
<http://gtb.inl.nl/iWDB/search?actie=article&wdb=WFT&id=117831>

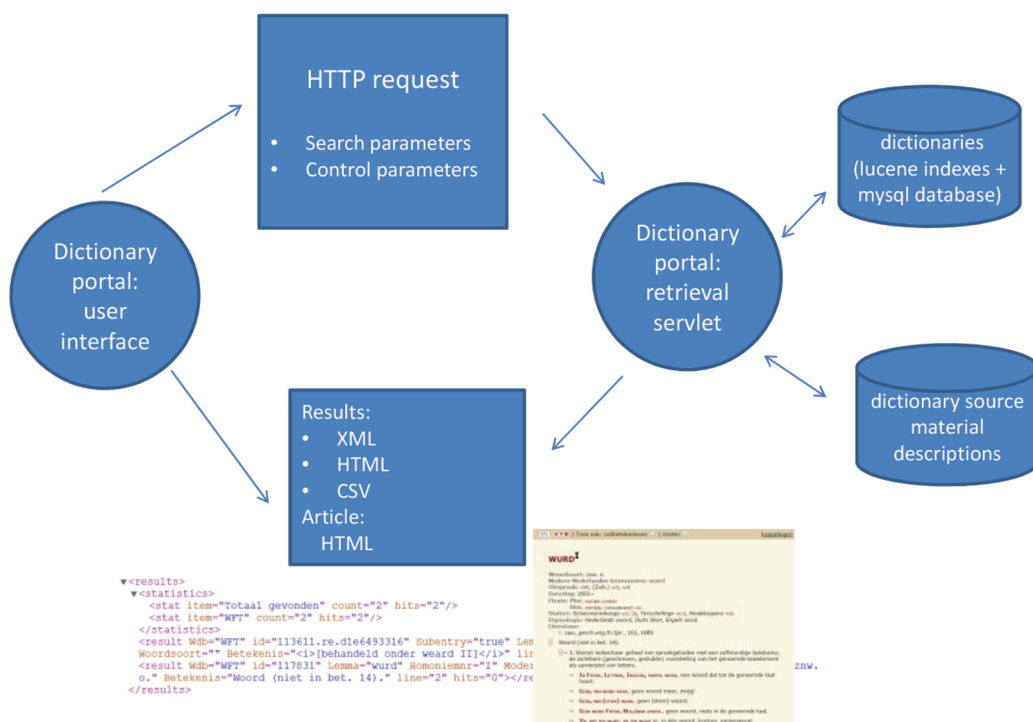


Figure 13.1: Basic use scenario.

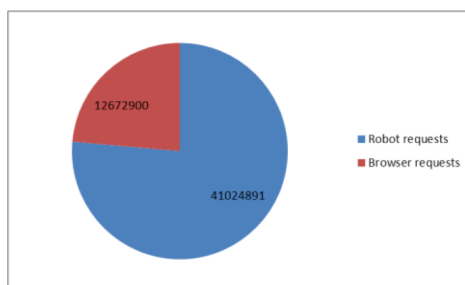


Figure 13.2: Bot requests vs browsers.

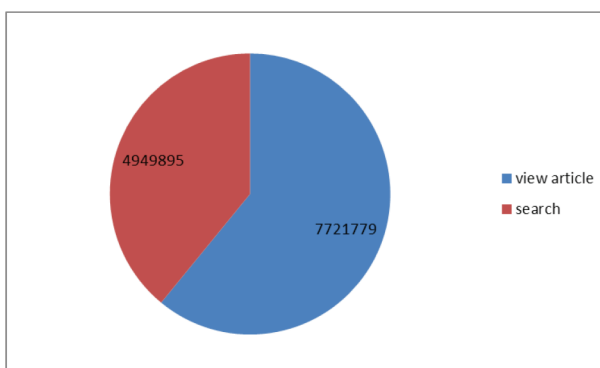


Figure 13.3: Article views vs searches.

13.5.2 WFT Usage

The historical dictionary portal responds to searches by presenting result lists including headword, part of speech and first dictionary sense. In some cases, the user can find the information he/she is looking for in the result lists; however, the user may also jump to article views (cf. Figure 13.3).

13.5.2.1 Article Views

At least 360,196 article views of the WFT can be with some degree of certainty ascribed to real users, and 64,407 different articles have been viewed since 2007.

Most frequently viewed are the lemmas listed in Table 13.1, a lot of which do not have a direct Dutch cognate.

13.5.2.2 Search Requests

The WFT is searchable in combination with the major scholarly dictionaries of Dutch. Hence, it is not obvious how to determine from a search action what the user was really interested in. Analysing the different combinations of dictionaries that have been searched (Figure 13.4 and Table 13.2), the emerging picture is that:

1. Most users do not seem to bother to exclude dictionaries they may not be interested in in their first search; and
2. when not all dictionaries are selected, most users just search a single dictionary (possibly by following links to this option).

omkoal	752	roppich	303	gean	226
oekedakke	720	nuver	300	kapokje	225
pommerant	653	noflik	295	tiid	222
knibbel	603	ramplesant	294	tankewol	220
bargebiten	416	bealch	286	muoike	219
bûter	408	tuike	267	ierappel	216
sjampoepel	388	jaan	257	jeuzel	216
suterich	385	eide	253	fertuten	215
gnob	385	jeits	251	petiele	213
obsternaat	384	hawwe	244	Fries	210
njonkelytsen	379	babbelegûchjes	243	santjin	209
opsokkebalje	367	kakwangen	242	earrebarre	205
futsjefinne	354	tuolle	240	hispel	202
kloet	346	frommes	237	bryk	201
poppeslok	325	heit	230	hynder	201
deunsk	320	kattedagen	230		
wurd	316	oft	229		

Table 13.1: Frequently viewed entries.

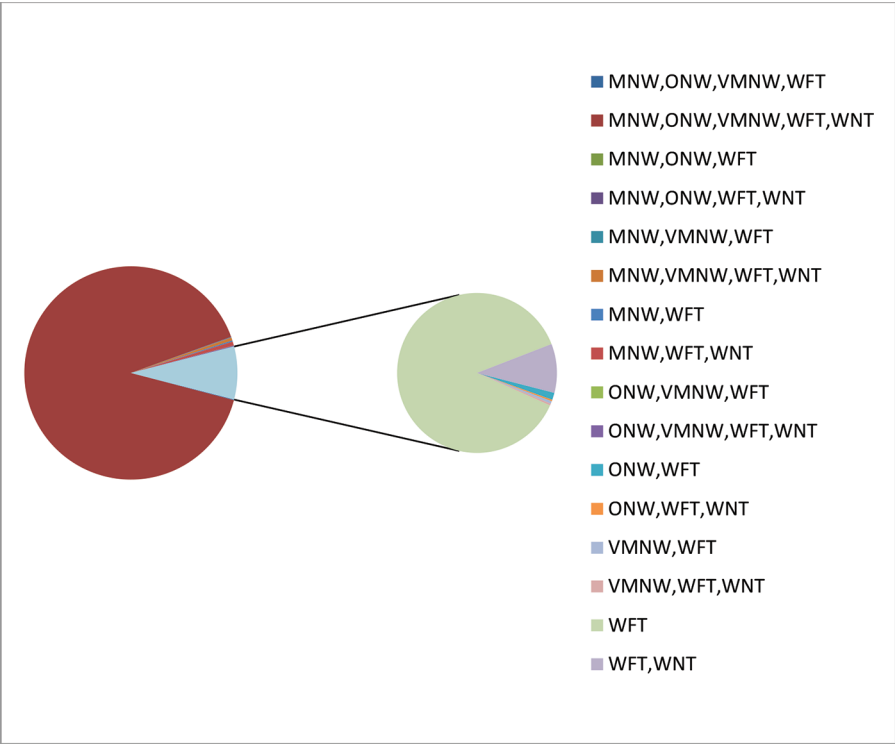


Figure 13.4: Relative frequency of consulted dictionary combinations involving WFT.

MNW,ONW,VMNW,WFT,WNT	1,748,488
WFT	135,462
WFT,WNT	15,318
MNW,WFT,WNT	11,465
MNW,VMNW,WFT,WNT	8,821
MNW,WFT	3,460

Table 13.2: Top consulted dictionary combinations involving WFT.

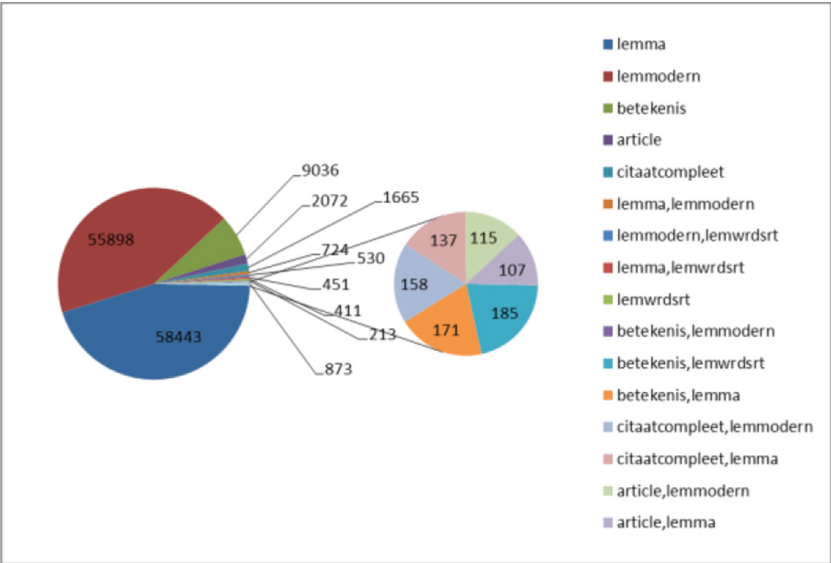


Figure 13.5: Frequency of search field combinations (WFT-only searches).

13.5.2.3 Search Types

The search application has many options (searchable field combinations). As can be seen from Figures 13.5 and 13.6, the overwhelming majority of users search for lemmas, using either the Frisian headword form or the Modern Dutch lemma field.

There is a high proportion of users searching for ‘betekenis’ in the WFT-only searches, as compared to the WFT-in-combination searches. This probably shows that people are trying to find Frisian equivalents for Dutch words.

It is striking that the many advanced options to explore the corpus of quotations are seldom used. Also, the beautiful set of collocations that has been encoded explicitly appears to be neglected.

13.5.2.4 Search Results

Table 13.3 shows to which extent users find what they look for, as measured through counts of either lemma queries grouped by lemma or ungrouped lemma queries (number of lemma search requests).

Most of these are Dutch words; others are spelling variants (*great, jim*), diminutives (*húske, brochje*) or inflected forms (*is, weagen, hie, froulju*) of existing headwords.

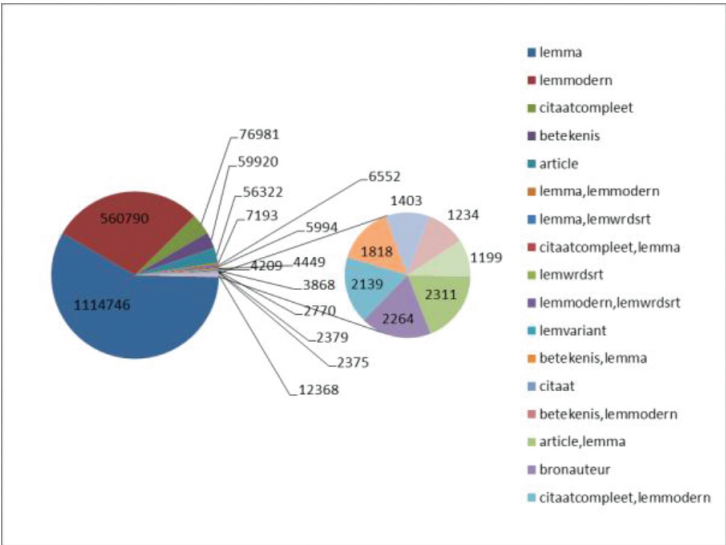


Figure 13.6: Frequency of search field combinations (searches where WFT is selected, possibly with other dictionaries).

Distinct lemma queries (29,893)	Found	20,201
	Not found	9,692
Lemma queries (58,561)	Found	45,972
	Not found	12,589

Table 13.3: Number of successfully or unsuccessfully retrieved lemmas.

13.5.3 Usage Through Time

There is a clear decline in the number of users of both WFT and the dictionary portal as a whole from 2014 (Figures 13.7 and 13.8). The main reason for this is that the portal application, developed in 2006, relies on Flash and is thus unsuitable for tablets and phones.

13.6 An Example of Research

Searches in the Frisian lemma field without results (queried more than six times) are listed in Table 13.4.

In the WFT every dictionary article describes all the meanings of a lemma, and every meaning is illustrated with citations, if available, for the complete described period (from 1800 until 1975). Accordingly, the dictionary can function as a source to shed light on the development of Frisian words and word variants. Using the search functions in the dictionary portal and the citations in the WFT, we wanted to investigate the historical development of Frisian word forms. The portal is a better tool than the paper dictionary for this type of research, since searching the paper dictionary is limited to looking up articles. We illustrate this type of investigation with research on the word *hynder* for English ‘horse’. Every meaning distinction in the dictionary articles is based on and illustrated with citations from two different card index systems that were built up between 1940 and 2011. The oldest card index system was more or less randomly put together, and the latest one was built up more systematically. The lexicographers needed the latter because the random system

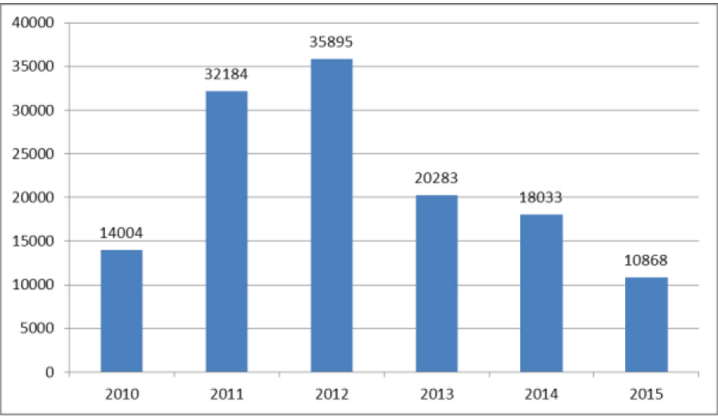


Figure 13.7: Usage (number of queries) through time, WFT-only.

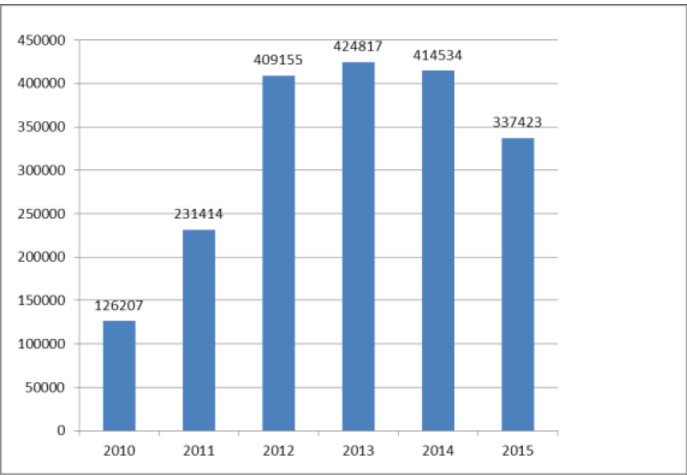


Figure 13.8: Usage through time, WFT-with-other dictionary.

Word	#Queries
gefeliciteerd	20
tot ziens	19
met	18
great	16
jij	16
wie	15
bedankt	13
jullie	13
kleinkind	12
koe	10

is	10
huske	10
jim	10
paard	9
zijn	9
samen	9
z	9
weagen	9
gaat	8
brochje	8
moeilijk	8

fierljeppen	7
groeten	7
froulju	7
meiinoar	7
hie	7
kinderen	7
weze	7
trouwen	7
merel	7

Table 13.4: Frequent lemma searches without result.

did not contain enough function words such as pronouns. In the first half of the 19th century, Frisian was not often written; this was also visible in the first card index system, in which this period was not well represented. When putting the dictionary together, a lot of work was put into collecting citations that would properly represent the entire described period. This means, for example, that if a word form or a collocation with the word form is attested across a period of 150 years, citations with that word form are given across the duration of that period. Because of this, it is possible to make statements about the development of written Frisian in the described 176 years on the basis of the citations. The development of the word *hynder* serves as an example.

In the header of the WFT article HYNDER three form variants appear: *hynder*, *hynsder* and *hynzer*. According to the dating of these three variants the last is the oldest (1808), and, according to the source material, the other two date back to 1851. The editor of the article did not choose the oldest form *hynzer* as the head entry – this is the right choice because in today's standard Frisian *hynder* is the only form that is used (Dykstra et al., 2014).² Incidentally, in 20th-century spoken Frisian one still encounters both the forms *hynzer* and *hynsder* (Van der Veen et al., 2001:59). The etymology of the word is best represented in the form *hynsder*, as it is originally a combination of *hynst* ('stallion') and *dier* ('animal').

With the help of advanced search options in the dictionary portal, each of the three word forms of *hynder* that appear in the header of the dictionary article can be selected. We also took into account the written word form *hynser*, which is not attested as a variant form of the lemma in the header of the article, but does occur in dictionary citations.

In the period 1819–1829, five unique citations with the written variant *hijnder* are attested. We have considered that form as a spelling variant of the word form *hynder*. By using the 'extended search' ('uitgebreid zoeken') function with every desired word form filled in in the 'word in citation text' ('woord in citaattekst') field in the 'citations' ('citaten') section, and with a time period selected in 'source data' ('brongegevens'), a list with citations is retrieved. (To select only citations, one needs to select the 'citations' ('citaten') search request in the 'give as results a list of' ('geef als resultaat een lijst met') field.)

In the WFT, not only are citations from literature, magazines and oral records given, but word lists and dictionaries are also used as sources. In the hard copy of the WFT the citations from these last sources are spaced; in the portal, they are tagged as such but are not yet defined as search fields in the extended search functions. The result of this is that the spaced words are missing in the search results.

Because the word *hynder* and its variations can appear as the first part of a construction and derivation and above all can be preceded by one or more parts of words, the wild card function '*' is used before and after the desired root word form when searching. For this example, searches were split into 10-year periods. Only the last period, 1970–1975, is shorter than ten years since the dictionary only includes entries up until 1975.

Searching for occurrences of citations with the four aforementioned forms of *hynder* gives hit totals with a distorted picture of the occurrence of search words, because the same citations can be used in different articles. In order to get a proper picture, unique citations were selected from the total hits. The search functions of the portal do not provide this functionality. For example, compare the next two equivalent citations, presented in two different ways in the dictionary articles SPREKKE (a) and SWAAIE (b):

- (a) It fjûrige, krêftúterjende hynder...(hat) altyd folle mear ta it forstân en it herte fen ús foarâlden spritsen, as de slûge kou.
- (b) It fjûrige krêftúterjende hynder, mei syn stirt yn 'e biezem, syn swaeijende moannen.

² <http://taalweb.frl/foarkarswurdlist>

In Figure 13.9 one can see the development of the use of the variant *hynder*. Up until 1860 there were almost no single unique citations found with this form. Afterwards, numbers increase slightly, but only as of 1930 can we clearly see a rise in numbers.

The form *hynsder* also appears at the beginning of the 19th century, as shown in Figure 13.10. Between around 1900 and 1950 the form appears relatively frequently, but afterwards there are hardly any single hits.

The distributions of the forms *hynzer* (Figure 13.11) and *hynser* (Figure 13.12) developed in the opposite direction to that of *hynder* and *hynsder*; both the *hynzer* and *hynser* word forms actually only appear in the 19th century and disappear completely from the citations after 1950.

Internally, the FA has a database containing all known Middle-Frisian (ca 1550–1800) language material, in which *hynder* and other spelling variations with the consonant <d> do not appear at all, although the form and variations thereof with the consonant <s> do. So, the search results in the online WFT show a clear continuing development in the usage of the word in written Frisian. At the cost of the other three variations, the word form *hynder* is the most used after around 1930, and the forms *hynzer* and *hynser* have now already been out of use for a number of decades.

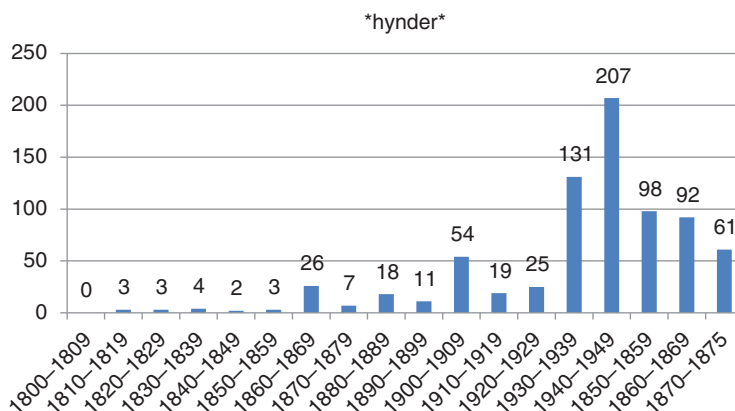


Figure 13.9: WFT *hynder*: number of quotations per decennium.

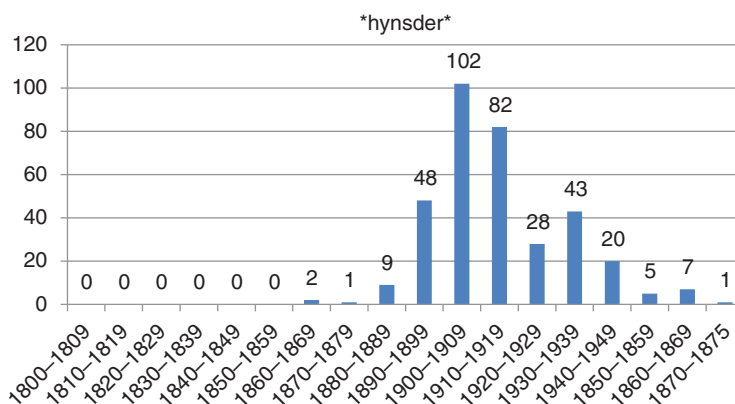


Figure 13.10: WFT *hynsder*: number of quotations per decennium.

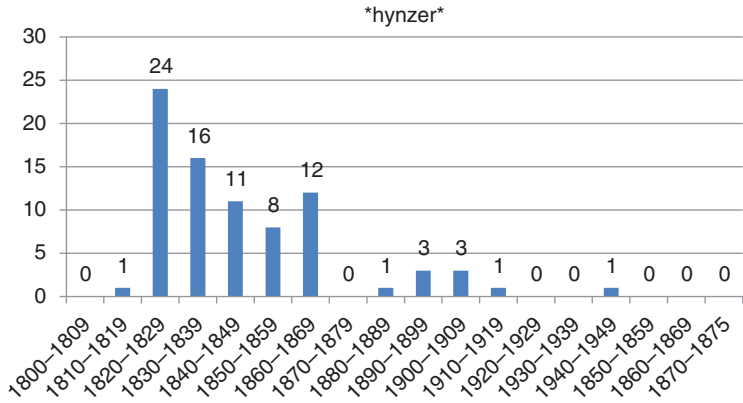


Figure 13.11: WFT *hynzer*: number of quotations per decennium.

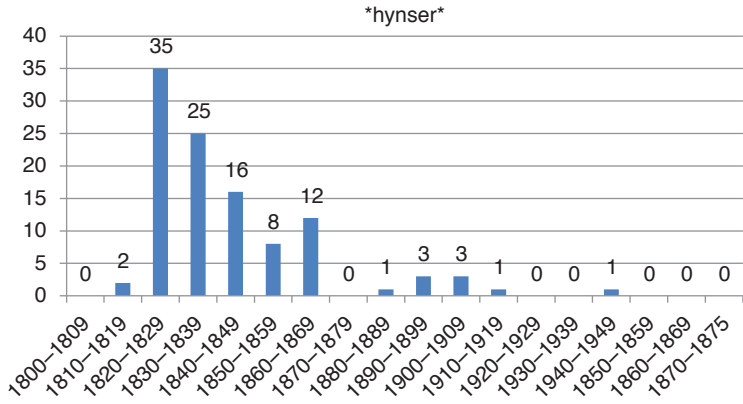


Figure 13.12: WFT *hynser*: number of quotations per decennium.

Apart from providing data on the usage of the Frisian word for ‘horse’, the ‘extended search’ function also showed in the article for *hynder* that the dating of the word form *hynder* to 1851 is incorrect, because in six different dictionary articles a source from 1836 is cited in which the simplex word form *hynder* appears, and in four different dictionary articles a source from the year 1819 is cited in which the simplex word form *hijnder* appears. If we accept the form *hijnder* as a written variant of the form *hynder*, the oldest dating of the word in the heading of the article *hynder* is thus 1819 instead of 1851.

13.7 Future Developments: Improvement and Expansion

There are two main ways in which the online WFT can be improved.

On the one hand, the portal application, which dates from 2010, needs to be updated and in some respects redesigned. The current interface uses Adobe Flash, which is rapidly becoming obsolete. Furthermore, in order to enable users to further explore the rich dictionary content, both the query interface, which should provide better guidance, and the general application design ought to be improved.

On the other hand, the WFT data itself can be enhanced. Frisian and Dutch words that are etymologically related are now linked through the same Modern Dutch equivalent. One wish is to enhance the Dutch–Frisian mapping where no plausible ‘etymological’ equivalent exists.

Furthermore, better tagging of the logical structure of the dictionary is required, especially with respect to the encoding of the field with etymological information in the heading and of the cross-references in the compound and derivation sections.

In order to broaden the possibilities of use, additional information can be added to the dictionary – for instance by linking sources mentioned in the ‘literature’ field to the relevant pdf files and by implementing links to dialect maps in the ‘dialect’ field. It should also be possible to link Dutch–Frisian cognates with the Etymological Dictionary of Dutch (Etymologisch woordenboek van het Nederlands, EWN).

Finally, one can imagine various ways in which to use the WFT data as published by means of the REST web service (either as part of the main portal or independent of it), such as the analysis and visualisations of the temporal and regional distribution of search results.

References

- Depuydt, K. (2010). ‘TEI-structuurcodering van de woordenboeken in de woordenboekencomponent van de Taalbank Nederlands’ (internal project document; not published).
- Dykstra, A.; P. Duijff; F. Van der Kuip; H. Sijens (2014). *Taalweb Frysk* [on-line] Ljouwert: Fryske Akademy. <http://taalweb.frl/foarkarswurdlist> [access date 6 June 2016].
- Veen van der, K.F.; A. Versloot; W. Rypma (2001). *Dialektgeografyske oantekens fan J.J. Hof*. Ljouwert: Fryske Akademy.
- Wurdboek fan de Fryske Taal/Woordenboek der Friese Taal. Ljouwert/Leeuwarden: Fryske Akademy, 1984-2011.