

A standardized wordlist and new spellchecker of Frisian

Anne Dykstra, Pieter Duijff, Frits van der Kuip and Hindrik Sijens

1. Introduction

In 2015 the Frisian Language Web saw the light of day online (FLW, <http://taalweb.frl/>). It has been developed by the lexicography department of the Fryske Akademy and consists of a newly created standardized wordlist of Frisian, an online spellchecker, a machine translator and a dictionary portal. These four applications together and individually offer a unique language tool to help Frisian native speakers and learners to write proper Frisian. The FLW is the first step towards a Frisian language platform that makes use of modern web technology. The standardized wordlist of Frisian is an important part of FLW and will not only be a firm base for future dictionaries but also the backbone of a special spellchecker. In this article we will give a brief overview of the compilation process of the standardized wordlist and how it is applied in the spellchecker.

2. Standard Frisian

Frisian has two major dialects and a number of smaller ones. Traditionally, variant pronunciations from minor dialects are not considered standard and thus are not included in dictionaries (anymore)¹. Nevertheless there is not a fully-fledged standard yet. Today's standard is mainly based on the two major dialects: Clay Frisian and Wood Frisian. In practice the present dictionaries do not always make a clear and well-reasoned choice between the two. In daily practice it appears that the variant that in the dictionaries is given first, is and has always been regarded as the standard or preferred variant by users. Yet the dictionaries' ambivalent attitude towards standard Frisian causes some uncertainty with those very same users. It appeared that educators, authors and civil servants had a strong wish for more guidance as to the choice of preferred variants². This doubt regarding correct usage is rooted in a lack of education and routine in writing Frisian. Frisian only became a compulsory

- 1 Only the scholarly *Wurdboek fan de Fryske taal* (<http://gtb.inl.nl/>) provides data from all dialects.
- 2 It must be said that there has also been resistance towards further standardization of Frisian, the main reason being fear of losing dialect variation.

school subject in the second half of the last century. In addition, because this obligation was applied only to primary schools and because written Frisian plays a minor role in daily life, most Frisians are not proficient in writing their own language. Language learners, as well as native speakers, are insecure in their language use, fearing to make mistakes. Even language professionals like journalists, editors, translators and novelists experience such problems.

Since the lack of a standard was felt to be a major obstacle in furthering the position of written Frisian, the Province of Friesland asked the Fryske Akademy to develop a standard for Frisian. It should be stressed here that the Provincial government does not have the legal means, nor the wish, to make the wordlist of Frisian a mandatory language standard. It is merely meant to be an aid to those who seek guidance when it comes to the choice of preferred variants.

3. Standard wordlist

The selection process for the wordlist has been guided by different criteria, though they have not been applied simultaneously:

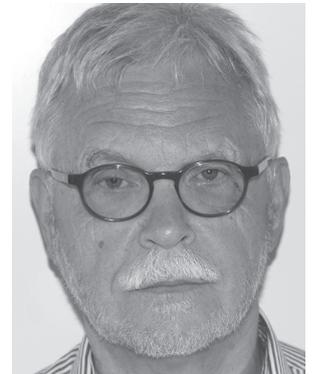
(a) tradition

Though not fully-fledged, there is a certain standard Frisian language, which has been codified in Frisian dictionaries over the years. Though the dictionaries have most of the time not selected variants from peripheral dialects, they do not always make clear which variant from the two major dialects should be prioritized. Yet it has been common practice for a long time to consider the variants given first in the dictionaries as the preferred ones. That is why, for instance, Clay Frisian variants in *-om* or *-omme* (*rom* 'spacious, large, wide', *tomme* 'thumb') are given preference over Wood Frisian variants in *-ûm* or *-ûme* (*rûm*, *tûme*).

(b) distancing

One element in the standardization of a language can be distancing, which implies moving away from another (dominant) language. In the case of Frisian, that other language is Dutch. An illustrative example is the preferred variant *hiel* 'whole', though it is only used in a small part of Friesland. In by far the larger part of Frisians use *heel*, which is identical to the corresponding Dutch word and therefore disqualifies as a preferred variant.

Another example is the deletion of *e* in words like *kiste* 'box' and *mûtse* 'hat, cap'.



Anne Dykstra has recently retired from the Fryske Akademy, where he has taken part in various Frisian language projects, including the scholarly *Wurdboek fan de Fryske taal* (Dictionary of the Frisian language). He was the editor of the *Frysk-Ingelsk wurdboek* (Frisian-English Dictionary), has worked on the Frisian Language Database, and was project manager of the Frisian Language Web. Currently, he is editor of the *International Journal of Lexicography*.
dykstraanne@gmail.com

Pieter Duijff is lexicographer at the department of linguistics of the Fryske Akademy. He was editor of a legal dictionary (2000), co-editor of the *Wurdboek fan de Fryske taal / Woordenboek der Friese taal* (1984-2011), one of the editors-in-chief of the monolingual *Frysk Hânwurdboek*, and editor of word lists of Town Frisian dialects (1998) and modern Frisian (online, 2015). Since 2014 he is one of the editors of a new online Dutch-Frisian dictionary.

pduijff@fryske-akademy.nl



Frits van der Kuip has been a lexicographer at the Fryske Akademy since 1986. He was one of the editors of the 25-volume *Wurdboek fan de Fryske taal* (Dictionary of the Frisian Language) and one of the editors-in-chief of the *Frysk Hânwurdboek*. In 2003 he completed his Ph.D. thesis on 17th century proverbs. More recently he was involved in the Frisian orthography and spellchecker project, and at present he is preparing an online Dutch-Frisian dictionary. fvdkuip@fryske-akademy.nl



Hindrik Sijens is a lexicographer at the Fryske Akademy, and an editor of the *Frysk Hânwurdboek* (monolingual Frisian dictionary). hsijens@fryske-akademy.nl

In a part of the language area it is perfectly normal to say *kist* or *mûts*, yet these forms are not selected as preferred variants because of their similarity to Dutch.

(c) uniformity

The principle of uniformity can also determine the choice of the preferred variant. Words like *died* ‘deed’, *ried* ‘advice’, *sied* ‘seed’, *goed* ‘good’ or *paad* ‘path’ are often pronounced and written in Frisian without the final *d*. However, the *d* always emerges in inflected and plural forms and in compounds, both in spoken and in written language (*dieden*, *riedsgearkomste*). For that reason *died*, *ried*, etc. are written with the final *d*.

Another example is the Northern variant *baarne* ‘burn’, which up till now is the form that is given first in most dictionaries, but which has to give way to the Southern *brâne* because of the uniformity criterion: *brân* ‘fire’, *brâne* ‘burn’, *brânfersekering* ‘fire insurance’.

(d) common usage

Frequency is a natural selection criterion when compiling a list of preferred variants. Highly frequent adverbs like *eigentlik* ‘in fact, actually’ and *eins* (idem) were included in the wordlist, while the less frequent *eigenlik*, *einlik* and *eink* were not.

(e) etymology

Sometimes the spelling reflects a less careful pronunciation: *abslút* ‘absolutely’, *bibleteek* ‘library’, *bommedearje* ‘to bomb’. The wordlist only contains the full forms: *absolút*, *biblioteek*, *bombardearje*. Exceptions to the etymological principle are lexicalized forms such as *knyn* ‘rabbit’ and *plysje* ‘police’ (instead of *konyn* and *polysje*) and word pairs having different meanings like *krupsje* ‘a/any disease’ and *korruptsje* ‘corruption’.

4. Spellchecker

It stands to reason that the standardized wordlist is the core of the spellchecker that is part of FLW. As explained, the wordlist is meant to be a practical tool to help those who want to use preferred variants in their written text. The term ‘preferred variant’ entails that variants that have not been included in the wordlist are not to be considered wrong. To further facilitate use, the wordlist has been incorporated into a database underlying the spellchecker in which non-standard words are linked to

preferred variants. This feature makes it possible for the user to focus on the use of the preferred variant.

The following example will demonstrate how the spellchecker works:

Hy skamme sich faor syn tiim.

‘He was ashamed of his team.’

The spellchecker will return the sentence with three underlined suggestions, each in a different colour:

Hy skamme sich faor syn tiim.

Red [*faor*] indicates a typo, green [*sich*] a Dutchism, and blue [*tiim*] a variant. The user can decide what he or she wants the spellchecker to do. For instance, to avoid seeing variants underlined the user can tick the blue box and get the following spellchecked text:

Hy skamme sich faor syn tiim.

Faor is a typo of *foar*, which is a mistake that any spellchecker would find. Since the wordlist usually lists English words in their original spelling, the user is advised to opt for *team*. A regular spellchecker might have come up with this suggestion as well.

Sich is quite a different case, clicking it renders the following list of alternatives:

har
harren
him
himsels

The correct Frisian word for the Dutchism *sich* in this case is *him*. The user will be able to replace *sich* by *him* by a simple click. This is where the FLW spellchecker stands out from regular ones, for which it would have been impossible to go from *sich* to *him*.

5. Conclusion

The Frisian Language Web has many functions, not the least of which is language maintenance. The first ever standardized wordlist of Frisian is supposed to encourage Frisians to write their native language with more confidence. When writers apply the spellchecker, they do not have to consult the wordlist directly, the spellchecker will make clear what the preferred variant of a particular non-standard form is. It is important to note that users have a choice in what they want the spellchecker to do. The Fryske Akademy will be happy to share its spellchecker technology with other (small) languages. Those interested are welcome to contact us.

Fryske Akademy was founded in 1938 in Leeuwarden, the capital of the Province of Friesland in the Netherlands, and concentrates on fundamental and applied academic research into the Frisian language, culture and society. While the results of its activities are primarily of academic interest, they have significant social meaning and relevance to Friesland and beyond, particularly in the context of minority and “small” languages. The academy’s

major lexicographic accomplishment is the *Woordenboek der Friese Taal* (Dictionary of the Frisian Language, <http://gtb.inl.nl/>). Its Linguistics Department has developed the Frisian Language Database (<http://fryske-akademy.nl/tdb/>) and the Frisian Language Web (<http://taalweb.fr1/>). Currently, a Dutch-Frisian online dictionary is in preparation. <http://fryske-akademy.nl/>