

The Frisian Language Database as a Tool for Semantic Research

Bouke Slofstra and Arjen Versloot
Fryske Akademy, Leeuwarden

In this paper, the authors present two examples from the field of body parts, illustrating the level of details, very often with a diachronic component, that can be detected in the study of semantics. The bilingual background of Frisian – Dutch has in one way or another been a second language in Friesland already since the late Middle Ages – constitutes an extra trigger in the organisation and restructuring of the lexemes and their meanings. The given examples from Google and the Frisian Language Data Base illustrate that a diachronic and comparative corpus based approach can add several aspects that have remained uncovered so far in the more traditionally conceptualised WFT. This enhanced picture of meanings and their developments are an essential prerequisite for gaining a deeper understanding of the organisation and structure of human semantic concepts and their operationalisation in human speech.

The 2010 Euralex Conference was held in the city of Leeuwarden (*Ljouwert* in Frisian), the capital of the Dutch Province of Friesland. The *Frisian Academy* hosted the event. One of the main projects of the Frisian Academy has been the twenty-five volume ‘Dictionary of the Frisian Language’ (*Wurdboek fan de Fryske Taal*, WFT), the last volume of which came to the light in 2010, on the occasion of that very Euralex Conference.

The WFT has a long history. As a first stage of this huge dictionary project a collection of file cards was created by staff members and volunteers from the very beginning of the Frisian Academy in 1938. Obviously, this collection has ever since served as a major source for this and other lexicographical work. But this is not the only source anymore.

In the first place, written Frisian spread over the internet in the last few decades, which enables the investigator to check the appearance of words and word combinations in modern daily Frisian by means of popular searching engines like Google. In the second place, the Frisian Academy created a digital *Frisian Language Database* (henceforth: FLDB for short), which comprises a broad selection of texts throughout the history of written Frisian, that is to say from approximately 1300 until the present day. The FLDB is still in progress, though a lot of work has already been done. In its final stage, all of these texts will have been POS-tagged. (At this moment, all texts from the period 1550 – 1800 have been POS-tagged already.) By doing so, we are able to link all tokens in the texts to types (grammatical and orthographical variants) and lemmas (the lemmas being represented in modern-day Frisian orthography). The lemmas are stored in the *FLDB lemma-list*, which is a central component of the FLDB. The FLDB thus enables the researcher to look for lemmas, types and tokens in a Keyword-in-Context module. An additional application makes it possible to look for two or more tokens within the same context for example ‘play’ and guitar’. Let us call this the ‘token combination application’.

The FLDB database (www.fa.knaw.nl/tdb) was set up for a number of purposes. It was meant to serve formal linguistic research as well as lexicographical practice.

Moreover, as a result of the Frisian Dictionary (WFT) project, we have at our disposal a large, well-defined and well-described WFT word list that is almost, but not yet fully integrated in the Frisian Language Database (FLDB). So the FLDB will be needing quite a bit of maintenance long after the Dictionary has been completed.

As it is clear now, the FLDB is made up of a number of interrelated modules, at the centre of which is the lemma list of Modern Frisian. An important derivative of this word list is the morphologically enhanced word list (the so-called Morpho-list). This list shows the morphological structure of most items in the LDB word list (for example: *farmer* = *farm* (noun) plus *-er* ending). We will illustrate how these resources can and do support the understanding of linguistic semantics.

The human body is an interesting word field, both from a linguistic and from an anthropological point of view. Recently a dissertation was published by Han Nijdam (2008), discussing the notions of *body* and *honour* in the mediaeval Frisian society, a so-called *feud society* that witnessed no functioning central authority. In this dissertation, a detailed description was given of a pre-modern concept of the former Frisian noblemen's 'physical' and 'political' 'bodies'. Much attention was paid to actual body parts, mainly because of the circumstance that mediaeval feuds used to be settled in the end by means of a system of financial compensations for the loss of 'honour', physical (the functioning of body parts) as well as immaterial. Versloot (1994) published a study of body part names like *foot* and *leg* in a wide area, including West Frisian in the Netherlands and historically related Frisian dialects in northern Germany.

Both studies stress the geographical and historical variability of body part names, the metaphorical (e.g. *head* = 'cup') as well as metonymical aspects of this variability (e.g. *foot* = 'leg', *leg* = 'bone'), and (mainly in Nijdam's work) the cultural aspects (e.g. the importance of the left foot for a horse riding Frisian warrior).¹

With these and other studies in mind, we may presume that 'body parts' is quite an interesting word field for lexicological studies. So let us concentrate on body part names in modern Frisian. Our first observation will be that a lot of mediaeval Frisian words did not survive (e.g. mediaeval *âch-brêd*, 'orbit'), but were replaced by Dutch-like synonyms. As a matter of fact, early-modern Dutch had a fine-grained medical and biological vocabulary, unlike Frisian, which was destined to be confined to everyday speech after the loss of the Frisian political independence in the sixteenth century. In the second place, early-modern and modern Frisian body part names are significantly less standardised than their Dutch counterparts, due to the above mentioned reason. This entails that one designatum (a body part) may be referred to by several words and vice versa (a single word may well be polysemous). A famous example is the word *foet* ('foot') that a growing number of native speakers use to denote the whole leg, which occasionally gives rise to misunderstandings. (See Versloot (1994) among others.)

Let us now concentrate on one particular body part: the Frisian and Dutch synonyms for 'neck' (Slofstra & Nijdam, forthcoming). In Dutch, there are two words that can be translated by English 'neck': (*de*) *nek* and (*de*) *hals* (cf. German *Hals*). The word *nek* refers to the bony back part of the joint, whereas *hals* is used to denote the soft and fleshy front part. *Nek* is in use as a pars-pro-toto for the whole joint that separates the head from the chest. So, modern Dutch makes an anatomically useful distinction between *hals* and *nek*, whereas in English *hals* is non-existent

¹ Variability in the naming of natural phenomena is also a well-known feature in the study of (indigenous) plant and animal names, both of which were studied extensively in the Frisian language area (e.g. Franke & van der Ploeg). A short look at the numerous names for the lady bird in the ALE will suffice to prove the sometimes bewildering variation.

and in German *Neck* is rare. (Note that in this case-study it is irrelevant whether the *hals/nek*-distinction reflects an etymologically ‘correct’ old-Germanic distinction or not.)

The Dutch dictionary user will find that his/her clear-cut semantic distinction is rather recent, inasmuch as *nek* is by now the only word that can be used as a *pars-pro-toto*. Petrified expressions from the old days reveal the older stage: a *halsmisdrijf* (‘neck-crime’) for instance is a capital crime; the word reminds us of the gallows in the middle ages. It is rather improbable that the gallows were conceived as an instrument that exclusively suffocated its victim instead of encircling (and breaking) the whole neck joint. Another example is the famous head-neck-chest-farm (*kop-hals-romp-boerderij*), a type of farm characterised by a tripartite structure that is typically found in the province of Friesland. In present-day Dutch, as is stated before, this extended use of *hals* is outdated. The *hals* is the soft, smooth and fleshy body-part that is visible to the spectator whether or not it is decorated by collars or ornaments like neck-laces. In metaphorical use a *hals* is a part of a bottle or a violin.

Nek, it seems, has become a kind of default expression (‘neck joint’) in modern Dutch, whereas *hals* has a more refined and even more ‘feminine’ meaning. In Google we find 50.000 attestations of Dutch *zijn/haar* (= his/her) *hals* and ten times as many attestations of *zijn/haar nek*; the combination *zijn nek* is much more frequent than *haar nek* (approximately 7:1), which is not so surprising, as the Dutch default possessive pronoun *zijn* can mean both ‘his’ and ‘its’. But *zijn hals* is only slightly more frequent than *haar hals* (5:4), thus corroborating the ‘feminine’ flavour of *hals* (table 1).

haar hals	her H	20.500	4%
	his/its		
zijn hals	H	27.200	5%
sum		47.700	9%
haar%		43%	
haar nek	her N	60.600	11%
	his/its		
zijn nek	N	445.000	80%
sum		505.600	91%
haar%		12%	

Table 1. *hals* and *nek* in Dutch (Google).
The contrast is statistically significant (Fisher’s Exact Test; $p \ll 0.01$)

One of the modules of the FLDB (the Token Combination Application) enables the investigator to search for word combinations.² In table 2, the reader can immediately find out that the proportions of *his* and *her neck* and ‘*hals*’ in the FLDB are different from the Dutch experiment. *Hals*, not Frisian *nekke* is the most frequent word. Note that both *hals* (Old Frisian *idem*) and *nekke* (< Old Frisian *hnekke*) are fine and indigenous words in Frisian. Frisian *hals* doesn’t seem to be as ‘feminine’ in combination with *hals* as it is in Dutch. Frisian *hals* is less exclusively associated with women, bottles and violins than modern-day Dutch *hals* is, which fact can be

² As a matter of fact, the Keyword-in-Context application can do the trick as well.

observed by using the Keyword-in-Context mode of the FLDB. Of course, we are comparing Frisian data from the last century with Dutch data from the last decade or so. (A ‘Google test’ for recent Frisian yields a more Dutch-like result.) But this only means that present-day Frisian mimics present-day Dutch.

har hals	her H	55	22%
syn hals	his/its H	108	43%
TTL		163	65%
har%		34%	

har nekke	her N	26	10%
syn nekke	his/its N	61	24%
TTL		87	35%
har%		30%	

Table 2. *hals* and *nekke* in Frisian (LDB).

No statistically significant contrast (Fisher’s Exact Test; $p = 53,5\%$)

A closer look at the data in the FLDB corpus and/or the WFT reveals to us that Frisian *hals* still had the pars-pro-toto meaning in the twentieth century and that the name of the visible body part *hals* was also used to denote the invisible throat that is covered by it, a metonymic word use that is quite common cross-linguistically as Nijdam points out in his dissertation, citing sources from comparative linguistic literature. In Frisian you can break your *hals* (as you still can in German) as well as your *nekke* (as you would say in Dutch nowadays). In Frisian a person can ‘have hunger in his *hals*’, (= be hungry) where *hals* is used as a synonym to the word for ‘throat’. This would sound rather odd to a speaker of Dutch. In the morphologically enhanced wordlist (the aforementioned Morpho-list) one can find both *dûkhalzich* and *dûknekkich* (‘dive-neck-y’) meaning ‘head-down’ (literally: ‘with forward bent neck’). So the denotations as well as the connotations of these words are shifting, and the Frisian situation resembles the outdated Dutch situation.

Another nice example from the field of body parts is the word-field of *thigh* in Frisian (Slofstra & Nijdam, forthcoming). Mediaeval Old Frisian *thiâch* is almost identical with English *thigh*. Its New Frisian successor *tsjea* is a rather literary word; it is almost never spoken.³ People prefer to say *dij*, which is clearly borrowed from Dutch, just like e.g. *heup(e)* (‘hip’), a borrowing that was already in use in the early-modern period. The WFT dictionary doesn’t make mention of this ‘unfrisian’ borrowing, but *sub voce* ‘bil’ (‘bottom, buttock’) it is said that this item also means ‘thigh’. The examples make clear that *bil* is probably used in a pars-pro-toto fashion: a person who has fat (or skinny) buttocks obviously has fat (or skinny) thighs attached to them. But *bil* is not found in the context of strong or weak thighs, nor is it associated with beautiful or less beautiful thighs. The funny thing is that these latter kinds of thighs are always dutchified *dijen*, even in the work of very consciously correct writers of Frisian.

³ Not that also words such as *tsjeaf* ‘thieve’ and *tsjeak* ‘cheek’ are obsolete in Modern West Frisian and replaced by *dief* and *wang* (both loans from Dutch). One could assume a pattern where lexical items that show outrageous phonological developments, compared to Dutch, tend to become replaced by more moderate word forms. An intriguing question is whether it is the lexical contrast alone that stimulates the borrowing or the word’s complex phonological shape as well.

These two examples from the field of body parts may suffice to illustrate the level of details, very often with a diachronic component, that can be detected in the study of semantics. The bilingual background of Frisian – Dutch is in one way or another a second language in Friesland already since the late Middle Ages – constitutes an extra trigger in the organisation and restructuring of the lexemes and their meanings. The given examples from Google and the FLDB illustrate that a diachronic and comparative corpus based approach can add several aspects that have remained uncovered so far in the more traditionally conceptualised WFT. This enhanced picture of meanings and their developments are an essential prerequisite for gaining a deeper understanding of the organisation and structure of human semantic concepts and their operationalisation in human speech.

Literature

- ALE = Saramandu, N. e.a. (1975-2007): *Atlas Linguarum Europae*. Assen: Van Gorcum & Rome: Poligrafica. Volume I, 'fascicules' 1 – 7.
- Franke, D. & Ploeg, D.T.E. van der (1985). *List fan offisjele Fryske plantenammen*. Leeuwarden/Ljouwert: Fryske Akademy.
- Frisian Language Database (FLDB). Leeuwarden: Fryske Akademy. www.fryske-akademy.nl/tdb
- Nijdam, H. (2008). *Lichaam, eer en recht in middeleeuws Friesland. Een studie naar de Oudfriesse boeteregisters*. Hilversum: Verloren.
- Slofstra, B. & Nijdam, H. (forthcoming): *Body part names in Frisian*.
- Versloot, A. P. (1994). *'Fuß' und 'Bein' in den modernen friesischen Mundarten, eine strukturell-
semantische Untersuchung*. Fach friesische Philologie Christian-Albrechts-Universität, Kiel.
- WFT = Van der Veen e.a. (eds.). (1984-2010). *Wurdboek fan de Fryske Taal / Woordenboek der Friese Taal*. ('Dictionary of the Frisian Language', Comprehensive Frisian Dictionary.) Twenty-five volumes. Leeuwarden/Ljouwert: Fryske Akademy.