What is 'normal'? Exploring semantic shifts in the medical domain

Merijn Beeksma[1]

[1]*Radboud University*

m.beeksma@let.ru.nl

This study aims to shed light on the semantic shifts from standard language that can be observed in the medical domain. We investigate the correlation between semantic change and word frequency, as well as the nature of semantic shifts in terms of narrowing, broadening, elevation, degeneration and metaphorical extension.

To study shifts from standard language in the medical domain, we compare clinical notes (±4 million words) taken from a Dutch corpus of electronic medical records to a corpus of Dutch Wikipedia pages (±17 million words), and detect semantic shifts with the aid of word2vec.

The first question this study addresses, is: *To what extent are shifts in meaning associated with frequency of use?* Because repetition is a precursor to conventionalization, we hypothesize that words which are frequent in the medical domain (relative to the general domain), are more likely to shift semantically than words that are relatively infrequent in the medical domain. Our findings illustrate that semantic shifts are indeed stronger for words that are used relatively often in the medical domain. The effects are most prominent for the categories of nouns and adjectives, and less so for verbs.

Additionally, we explored the nature of the semantic shifts for the most dissimilar words of each part of speech category, to answer the question: *How can semantic shifts from standard language in the medical domain best be characterized?* We notice that compared to the general domain, words in the medical domain seem to have shifted mostly towards more specific or more metaphoric uses. Verbs and nouns for example contain many metaphoric words, such as fight-related metaphors to describe how patients handle diseases: *battle**, *fight*, *struggle*, *attack, conquer*, and *lose* are strongly associated with terminally ill patients, while these words are associated with concepts such as *war*, *invasion* and *allies* in the Wikipedia corpus. Other metaphors include nature-related phenomena: in the medical corpus, it is pain that comes in waves, while it are mostly fluid things such as water that come in waves in the Wikipedia corpus.

Furthermore, we notice that meaning shifts result from semantic narrowing rather than semantic broadening: *positive* is associated with words such as *honest* and *logical* in the general domain, but is related to test outcomes in the medical domain. Similar observations can be made for *normal*, which is used to qualify observations objectively rather than subjectively in the medical domain.

In this talk we will provide a detailed description of the methods we propose to detect semantic shifts in unlabeled corpora, and will elaborate on our findings and conclusions, illustrated with examples.

* All examples are translated from Dutch to English.