
Body Parts in Norwegian Books

Lars Gunnarsønn Johnsen

yoonsen@gmail.com

National Library of Norway, Norway

Introduction

The goal of the present study is to look at body parts within the class of published Norwegian books. Our main question is to look at the difference between referencing a body part directly from making a reference via a possessive pronoun. The body is important in our culture and is studied both within and outside the digital realm, e.g. the papers in Christopher E Forth; Ivan Crozier (2005) studies the body in culture, while Mahlberg (2013) uses corpus methods in a literary study.

We report on a pilot study that considers body parts across the whole book collection without breaking the collection up into different genres or time periods. Although the pilot study is limited to an across-the-books analysis, it is part of effort to study the effect of different genres, as well as newspapers and journals.

Method and data

Words for body parts are taken from Norwegian books in the period 1810 up to 2000, using the digitized books made available through the Norwegian National Library, approximately 460 000 books.

The contexts we consider for nouns describing body parts fall into three types as described in Lødrup (2011) and Delsing (1998). Norwegian may express possessives parallel to the English pattern, like “hans arm (his arm)”, alongside the definite version like “armene hans (arms.PL.DEF his)”. Modern Norwegian seems to prefer the definite plus possessive, in particular for body parts, which therefore will be the construction we focus mostly on here.

A slight complicating factor in Norwegian possessive construction is that sometimes the possessor may not be expressed e.g. Lødrup (op.cit.), in contrast to English, where the possessor cannot easily be replaced

with “the” in “John had a pain in his arm” while in Norwegian this is the norm if the possessor is the subject “John hadde vondt i armen/John had pain in the arm”.

Each possessor phrase, pronoun plus body part gets a collocation graph as described in Brezina et.al (2015) where each edge in the graph is weighted with PMI (Pointwise Mutual Information, e.g. Lewandowska-Tomaszczyk (2007), Romesburg (2004)). Collocation graphs can be seen as a cluster of words for the phrase generating it, ordered by PMI.

As an example, the first three words in the cluster (or collocation graph) for “håret hennes (her hair)”, computed from approximately 12 000 concordance samples, go like this:

Word/translation	frequency	PMI
duften/smell	350	36.42
børstet/combed	198	32.59
fingerne/the fingers	565	30.73

There are 29 Norwegian body words in all going into this study, some duplicated in singular and plural, resulting in 21 unique body parts from head to toe.

Each cluster is cut down to its top 200 words which are then compared using two standard similarity measures, the cosine and the Jaccard-similarity, where the former accentuates similarity of the clusters as weighted distributions, the latter highlights the set equality of the clusters.

Results

Our research question is how references to body parts differ when referenced using a pronoun, or with no pronoun specified. Note that even though no pronoun is specified, it is not required that the reference is done without a possessor, so there will be a certain overlap in the samples.

Our main result is that female and male possessive constructions generate clusters that are closer than standalone body words. Looking at the clusters themselves we see that some is due to the expressiveness of the body, for example “øynene/the eyes”, which ranks on top between female and male parts, has words like “glimt/sparkle” and “lyste/lightened” which is used to express emotion emanating from the beholder. These words are absent for the cluster for the word “øyne“eyes”. Also, words like “hendene/the hands” in the possessive construction yield words of action like “grep/gripped”, “slapp/released”, while outside the possessor construction generic actions like “klappe/clap” is found.

The next step is to study these constructions with respect to a distinction between classes of works, using the metadata of national bibliography, and also the

difference across media types such as newspapers, books and journals.

Bibliography

- Brezina, V., McEnery, T., & Wattam, S. (2015).** [Collocations in context: A new perspective on collocation networks](#). *International Journal of Corpus Linguistics*, 20(2), 139-173.
- Delsing, L. O.** (1998) "Possession in Germanic", in *Possessors, Predicates and Movement in the Determiner Phrase*, ed. Artemis Alexiadou & Chris Wilder, John Benjamins, Amsterdam
- Forth, C E, and Crozier, I., eds.** (2005) *Body parts : critical explorations in corporeality* Lanham : Lexington Books.
- Lewandowska-Tomaszczyk, B (2007) *Corpus linguistics, computer tools, and applications : state of the art* P.Lang Frankfurt am Main, New York
- Lødrup, H** (2011), "Norwegian Possessive Pronouns: Phrases, Words or Suffixes?" in *Proceedings of the LFG11 Conference*, Butt, Miriam and King Tracy eds., CSLI Publications, <http://csli-publications.stanford.edu/>
- Mahlberg, M** (2013). *Corpus Stylistics and Dicken's Fiction*. Routledge.
- Romesburg, H. C.** (2004) *Cluster analysis for researchers*, Lulu Pr.