Studying Disability Related Terms with Swe-Clarin Resources

Lars Ahrenberg[1], **Henrik Danielsson**[2], **Staffan Bengtsson**[3], **Hampus Arvå**[1], **Lotta Holme**[2], **Arne Jönsson**[1]

[1] Department of Computer and Information Science, Linköping University, Sweden
lars.ahrenberg|hampus.arva|arne.jonsson@liu.se

[2] The Swedish Institute for Disability Research, Linköping University, Sweden
henrik.danielsson|lotta.holme@liu.se

[3] The Swedish Institute for Disability Research, Jönköping University, Sweden
staffan.bengtsson@ju.se

Abstract

In Swedish, as in other languages, the words used to refer to disabilities and people with disabilities are manifold. Recommendations as to which terms to use have been changed several times over the last hundred years. In this exploratory paper we have used textual resources provided by Swe-Clarin to study such changes quantitatively. We demonstrate that old and new recommendations co-exist for long periods of time, and that usage sometimes converges.

1 Introduction

Digitisation (with OCR) of textual material previously available only in print has enabled large-scale quantitative studies of the recorded past. Coupled with methodological developments in natural language processing (NLP) and other fields, researchers in the humanities and social sciences can study the past in new and powerful ways. Well known examples can be found in the study of literature (Moretti, 2005; Jockers, 2013), in cultural history (Michel et al., 2011) and language change (Tahmasebi et al., 2018).

It is noteworthy that much of the research so far has been conducted on English data. As the quantity of historical Swedish texts that are digitised is increasing, linguistic change in Swedish, whether ”natural” or prompted by technological innovations or by the recommendations of public authorities, can begin to be studied by digital methods. In this study we are concerned with lexical changes in the domain of disabilities. This domain is of special interest in a Swedish setting as the understanding of what disability is, and what it means, has been the subject of much debate, causing new recommendations to be issued from time to time as regards appropriate terminology (see Section 2). To the best of our knowledge this is the first quantitative study of Swedish disability terms.

The study is a collaboration between computational linguists on the one hand and historians and disability researchers on the other. It is ongoing; in this paper we report some early results.

2 The concept of disability (in Sweden)

Several models and perspectives have been discussed and proposed in relation to disability. The traditional way to approach this field has been labelled the *medical* or *individual* model. This is foremost a term that has been introduced by its opponents, as a contrast, and can hardly be said to have its own advocates. The medical model tends to reduce the phenomenon to body functions and bodily deficits. Thus, disability occurs on an individual level, since it is the restrictions caused by physical or mental deviations or flaws that in the end explains why someone experiences problems in everyday life. The inherent logic of the medical model is to a large extent guided by ideas of bodily normality and so much of the attention is directed to compensation.

This way of approaching and understanding disability has been challenged by the environmental turn first materialised in the so-called *social model*, a model that emerged within British disability activism in the 1970s. As opposed to the medical model, disability is rather viewed as the outcome of social, structural and institutional barriers. What turns an impairment into a disability depends on how the society is

constructed. If society creates barriers in forms of both physical inaccessibility and degrading attitudes leading to various actions of discrimination the answer is not about normalising measures but to change how society works. According to the social model, disability is about combating all these social barriers.

One objection to the social model has been its alleged neglect of impairments and the body as well as the experience of the individual. As a part of this criticism competing models have been developed. In Scandinavia the *relative*, or *relation model* has gained ground. According to this approach the question of what becomes a disability is not given but comes as a result of the interaction between the individual and the surrounding environment. A person with a certain impairment can be disabled in one specific context or situation but not in another. It all depends on how the environment is constructed and what type of support is available. While the social model's claim of universal barriers, injustice and oppression is difficult to maintain, the relational model is close to it by emphasising that disability must be understood in relation to the environment.

A great breakthrough for the disability movement in Sweden came in the 1970s when the Disability Federation Central Committee introduced a joint disability programme, called A Society for All. As early as the 1960s, the concept of disability in official documents and legal texts included some social model elements, and in the programme A Society for All it was claimed that society and the environment should be designed according to the needs of all citizens. It was not enough to bring the individual to society; society must also be made accessible. An important question for us is to what extent this view of disability is found in official reports and media.

The term *handikapp* ('disability') was introduced as an umbrella term for the many different terms that denoted special types of disability. *Handikappad* ('disabled') was something a person was, but with the introduction of an environment related view, other words such as *funktionsnedsättning* ('functional impairment') and *funktionshinder* ('functional impediment') were recommended. More recently these words, too, have been put into question, and a shift of attention to enabling measures has been proposed signalled by terms such as *delaktighet* ('participation'). These changes are not only replacements of forms but of (desirable) denotations and connotations.

3 Data processing and analysis

Our primary resource for this study are the Official Reports of the Swedish Government (henceforth: SOU[1]) from 1922 to 2016 as found at the Swedish Language Bank[2], the resource hub of Swe-Clarin[3].

For the studies on frequencies and word embeddings the texts were lowercased and stop words were removed and grouped into decades. It was necessary to use this coarse granularity as reports covering topics related to disability are unevenly distributed over years.

The SOU-files, especially for the earliest periods, contain many errors due to failing OCR. However, word-based methods are often robust allowing general trends in the data to be captured even in the presence of noise. As it turned out, also word embeddings could be produced, showing plausible relations between terms.

3.1 Frequency changes

An initial list of 60 words referring to disabilities and/or disabled persons over the last 100 years was manually produced by the disability researchers. For lack of space we only report data for the general terms introduced above.

The words are either nouns or adjectives, which means that they occur in Swedish text in a variety of inflectional forms, up to eight for nouns, up to ten for adjectives. They also form derivatives and compounds. We have assumed that sharing of a common stem implies sharing a meaning.[4] This is a simplification, but does not prevent the discovery of general trends. Thus, we are comparing relative

[1] An acronym for Statens Offentliga Utredningar.

[2] `https://spraakbanken.gu.se/swe/resurs/rd-sou#tabs=information` and `https://spraakbanken.gu.se/swe/resurs/sou#tabs=information`.

[3] The most recent version at the time of writing being published in July, 2017.

[4] For most words, the stem is identical to the look-up form in a Swedish dictionary. For some words, two forms are required due to stem variation, as in *galen* ('mad' singular) vs. *galna* ('mad' plural), *galning* ('mad person').

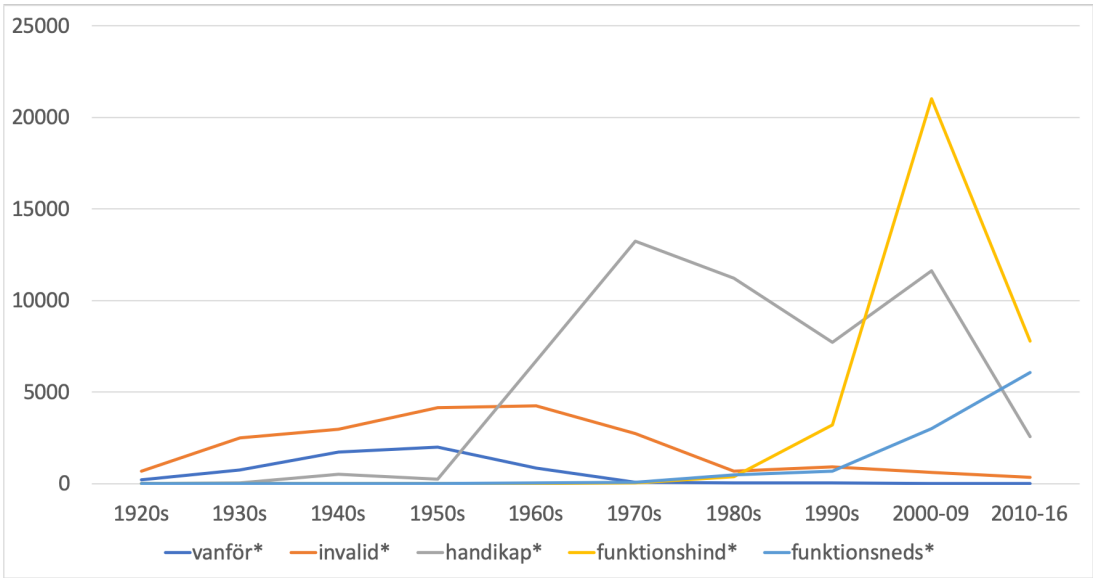

Figure 1: Usage of some Swedish disability terms 1922-2016 by decades.

frequencies within a cohort of terms assumed to cover roughly the same semantic space over a decade. See Figure 1. We can see clear changes of dominant terms since the 1920:ies:

$$\text{invaliditet/vanförhet} \rightarrow \text{handikapp} \rightarrow \text{funktionsnedsättning/funktionshinder}$$

In agreement with our expectations, these data indicate that there has been a change in the language used to refer to disabilities and persons with disabilities. While the last change is to be expected from the adoption of a relational model, it seems to happen quite slowly and with full force much later than one could expect. Also, the change is not abrupt. Thus, it seems that several disability models are at work simultaneously.

3.2 Terms in context

By looking at the contexts in which a word is used we can gain an understanding of how people use it. We may use a concordancer or simply look at cooccurrences with words in the context. The Korp concordancer [5], which has a parsed version of the SOU-texts, can display neighbours with different grammatical relations to a word with their frequencies. Korp also enables the generation of concordances for the pair of context word and key word. See Figure 2.

1. svår	233	De som har allvarliga	**medicinska handikapp**	t ex.
2. psykisk	147	STATENS OFFENTLIGA UTREDNINGAR		(stödjer ej
3. fysisk	148	arbetsföra av 2:a graden innehöll individer med mera avancerade	**medicinska handikapp**	.
4. grav	63	t enkel och arbetsförmågan kan bedömas med kännedom om det	**medicinska handikappet**	, som de
5. social	125	a en tillfredsställande arbetsinsats endast om hänsyn tas till deras	**medicinska handikapp**	vid val a
6. livslång	58	mma betingelser kan dessa som ovan framhållits trots avsevärda	**medicinska handikapp**	ofta gör
7. motorisk	37	arbetsförhållanden för arbetstagare, som på grund av avsevärda	**medicinska handikapp**	icke kan
8. olik	174	Det	**medicinska handikappets**	tidigare
9. neurologisk	27	Det	**medicinska handikappets**	som kar
10.allvarlig	43	hänsyn handikapp I första hand kan det	**medicinska handikappet**	påverka
11.mental	23	Om en person trots	**medicinskt handikapp**	bedömt

Figure 2: Adjectival attributes of the noun *handikapp*, 'disability', from 1 billion tokens in the Swedish Language Bank. Apart from the SOU-files all newspaper data and a corpus of novels have been included. The icons to the right of an attribute provide links to a KWIC concordance, where the search word occurs with this particular attribute.

[5] https://spraakbanken.gu.se/korp

The figure shows that 'handikapp' is often accompanied by attributes referring to extent: *svår*, 'hard, difficult', *grav*, 'grave', *allvarlig*, 'serious', *lätt*, 'light', or kind: *psykisk*, 'mental', *fysisk*, 'physical', *neurologisk*, 'neurological', *medfödd*, 'congenital', *livslång*, 'life-long'. There are many overlaps with the corresponding lists for the words *sjukdom*, 'disease' and *funktionshinder*, 'functional impairment'. This clearly gives the impression that the medical model is well represented in the data.

funktionshinder	1970-79	1980-89	1990-99	2000-09	2010-16
1970-79	=	sjukdomstillstånd	sjukdomar	sjukdomar	smärttillstånd
1980-89		=	handikapp	funktionsnedsättning	funktionsnedsättning
1990-99			=	funktionshinder	funktionsnedsättning
2000-09				=	funktionsnedsättning

Table 1: Forward temporal analogies for the term *funktionshinder*.

To obtain richer models we have trained word embeddings for the full corpus of SOU-reports, and for each decade[6].

Because the training algorithm initializes with random weights, ten models were trained for each decade from the 1970 and forward, resulting in a total of fifty models. To control that the models were stable, the top three words by cosine similarity were checked for each one of the fifty models.

We could see a change in moving from the 1970:ies to the 1980:ies. For the terms *funktionshinder* and *funktionsnedsättning*, the term *handikapp* is one of the three closest neighbours only once in the period 1970-79. In the following period 1980-89, *handikapp* is the closest neighbour for both terms.

We have also compared the vector spaces for different decades using the technique of temporal analogies (Szymanski, 2017). The method enables comparisons of one vector space to another by a global transformation or projection. Each 'early' model was paired with each 'later' model, e.g. a model from 1970 was paired with all other 1970 models and all models from later decades. After the transformation, cosine similarity was checked again in each model to see if there were any new words showing up. Table 1 shows the closest analogy for the term *funktionshinder*. The picture we got for this word by considering the neighbours in each decade, is confirmed. In the 1970:ies this term was used differently, analogously to words such as *sjukdomar*, 'diseases' in later years. From 1980 onwards it shows more affinity with the terms *handikapp* and *funktionsnedsättning/funktionsnedsättningar*.

4 Conclusions

Quantitative word-based investigations of even fairly noisy textual data may corroborate and enrich qualitative approaches and reveal patterns and trends in language use that can be compared to advice and recommendations. Our analysis of the use of Swedish disability terms in resources made available by Swe-Clarin partners indicates that, while recommendations have effects, they are delayed, and that several frames of thinking about disability live along side by side.

References

Matthew L. Jockers. 2013. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press.

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, , Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

Franco Moretti. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.

Terrence Szymanski. 2017. Temporal word analogies: Identifying lexical replacement with diachronic word embeddings. In *Proceedings of the 55th Annual Meeting of the ACL (Short papers)*, pages 448–453.

Nina Tahmasebi, Lars Borin, and Adam Jatowt. 2018. Survey of computational approaches to diachronic conceptual change. *CoRR*, abs/1811.06278.

[6]We used the Word2Vec CBOW model as implemented in GenSim (`https://radimrehurek.com/gensim/index.html`)