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 Optical Character Recognition, 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 disability. 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).

There are some qualitative studies in the area of disability research in Sweden, (for example (Gustavsson Holmström, 2005; Gardeström, 2006; Larsson-Severinsson et al., 2009; Holme, 2000; Lindberg and Grönvik, 2011; Oliver, 2013; SOU 2019:23, 2019; Söder, 1982; Söder, 2013), that describe and analyse the changes in disability concepts. A problem with these studies is that they often are based on each other and on a selection of texts that both present and argue for renewing the concepts of disability in politics and society. Some of these studies therefore show similar results focusing the very same narrative.

In this article we argue that quantitative studies can contribute to enhance and enrich knowledge about changes of disability concepts. A more varied picture of the concepts of disability may emerge when new possibilities arise for analysing larger amounts of data. In this paper we give some examples to support this view, using Swe-Clarin resources and word space models. 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 many 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 explain 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 therefore 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 compensational measures but to change how society works. According to the social model, disability is about combating 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 *relational model* has gained ground. According to this approach, the question of what becomes a disability is not given but is shaped 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 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 discrimination 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 (Centralkommitté, 1972) 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.

Given this history we set out to study the usage of disability related terms over time with the following questions in mind:

- How fast are new understandings of disability, and new recommendations as to term usage, in particular, followed in official reports and media?

- What can quantitative studies based on language technology contribute to our understanding of disability related terms and concepts?

3 Methods and results

We decided to use the Official Reports of the Swedish Government (henceforth: SOU[1]) as our primary source for the study. The SOU's are ordered by the government and are primarily used by them, the parliament and organisations that may have an interest in the issues presented in an SOU. They include a number of reports where disability is the main focus, as well as reports with other main themes where disability may be a side theme or just touched upon in some sections.

---

[1]An acronym for Statens Offentliga Utredningar.

Four types of studies have been performed:

- Frequency studies, where the main methodological issue has been handling inflection in noisy text (Section 3.3)

- Studies of co-occurring terms, using the Swe-Clarin Korp web service (Section 3.4)

- Word space modelling of relevant terms (Sections 3.1 and 3.5)

- A study of temporal analogies based on transformations of word spaces (Section 3.6)

3.1 Data and preprocessing

The SOU reports covering the years from 1922 to 2016 were downloaded from the Swedish Language Bank[2], the resource hub of Swe-Clarin[3]. For the study on co-occurring terms, we have included newspaper texts, also from the Swedish Language Bank.

A main issue for the methods we wanted to use was the quality of the older texts that have been scanned using optical character recognition (OCR). There are frequent misreadings of certain letter combinations, but more seriously, hyphenated words are recognised as two separate units and the OCR does not at all handle the common two-column format, thus, often completely misrepresenting the original text. It was our hope, however, that the methods would be robust enough to allow general trends in the data to be captured even in the presence of noise. To check this we looked at the ten nearest neighbours in vector spaces created as described below in Section 3.5. This was done both for disability related words and for common words, and for each decade covered by the corpus. It turned out that the absolute majority of neighbours were either morphological variants, misrepresented variants, semantically related words such as near synonyms or co-hyponyms, or syntactic attributes. As shown in Figure 1, the word *handikappad* for the period 1970-79 had neighbours such as *handikappade*, a morphological variant, *handi-*, a hyphenated part, *gravt*, 'seriously', a syntactic attribute, and a number of semantically related words. For the word *potatis*, 'potato(es)', all neighbours are related to food.

```
>>> wd = 'handikappad'
[('handikappade', 0.7128802537918091)
'familjeförsörjare', 0.6845875978469849)
'sjuk', 0.6721851229667664)
'arbetshandikappad', 0.6708623170852661)
'gravt', 0.6570059061050415)
'arbetsanställning', 0.651633620262146)
'rörelsehindrad', 0.6509082317352295)
'invalid', 0.6493630409240723)
'handi-', 0.646192193031311)
'invalidiserade', 0.6459981203079224)]
```

```
>>> wd = 'potatis'
[('sockerbeta', 0.910276472568512)
'spannmål', 0.9080594778060913)
'oljeväxter', 0.9044820070266724)
'rotfrukter', 0.8834930658340454)
'grönsaker', 0.8735833764076233)
'foder', 0.8698905110359192)
'djupfryst', 0.8694902658462524)
'köksväxter', 0.8692591190338135)
'socker', 0.867942214012146)
'mjöl', 0.8586941957473755)]
```

Figure 1: Ten nearest neighbours for the words *handikappad* and *potatis* for the sub-corpus of SOU:s covering the period 1970-79.

From these initial trials we concluded that the corpus, in spite of its noisiness, had sufficient quality to yield interesting results. However, we can not rely on absolute numbers, but the relative changes and differences that can be observed are sizeable enough to be trusted.

Another issue is that we discovered that there were some missing SOU-reports in the available files. This should be rectified in future releases of them.

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

---

[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.

> **General terms:** *handikap*, *handicap*, funktionsneds*, funktionshind*, funktionsvari*, funktionsnivå*, funktionsrätt*, *funkis*, *funkofobi*, funktionbegräns*, *delaktighet*
> **Contested / Group-specific terms:** rörelsehind*, rullstolsbu*, vanför*, lam, lyt*, krympling*, invalid*, fallandesjuk*, svagsint*, andesvag*, vansinnig*, ofullkom*, sinneströg*, vanskapt*, imbecill*, kretin*, debil*, fåne, fånig, utvecklingsstör*, sinnesslö, obildbar*, idiot, efterbliven, döv, döf*, hörselskadad*, dövstum*, dumbe , blind, synskad*, dövblind, fallandesot, epilepsi
> **Vague terms:** *abnorm*, halt, ofärdig, missbildad*, onormal*, avvikande, galen*, särskol* partiellt arbetsför*, personer med, för alla, särskilda behov, särskilt stöd

Table 1: A list of Swedish terms and general words referring to disabilities. The asterisk indicates positions where the term may be extended.

3.2 Disability-related terms in Swedish

The term *handikapp* ('handicap') was introduced in the 1950s as an umbrella term for the many different terms that denoted special types of disability. *Handikappad* ('handicapped') was something a person was, but with the introduction of an environment related view, other words such as *funktionsnedsättning* ('functional impairment'), *funktionshinder* ('disability'), and *funktionsvariation* ('functional variation') 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.

3.3 Frequency changes

An initial list of some 60 words and phrases referring to disabilities and/or disabled persons over the last 100 years was manually produced by the disability researchers. See Table 1.

The list included scientific terms from disability studies, the recommended and contested terms of disability politics, and terms and general words for disabilities affecting specific capabilities, such as sight, hearing, and movement. From the initial experiments of finding nearest neighbours reported above, we found a few additional terms including adjectives such as *psykisk*, 'mental', which are common in multi-word disability terms. It also turned out that some of the words were too infrequent to be included in the study.

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.

The multi-word terms were re-tokenized as single tokens before processing.

The words were grouped into three categories for the frequency studies:

- *Fysiskt handikapp* ('Physical handicap') using the Swedish terms *funktions**[5], *funktionsneds*, handikap*, invalid*, lyte*/lytt*, rullstolsb*, rörelsehind*, vanför**

- *Psykisk sjukdom* ('Mental disability or illness') using the Swedish terms *mentalsjuk*, sinnessjuk*, dår[ae]*, gale?n*, psykisk rubbning*, psykisk sjukdom*, psykisk störning**

- *Förståndshandikapp* ('Mental retardation') using the Swedish terms *efterbliv*, förståndshandi*, idiot*, imbecill*, utvecklingsstör**

Thus, we are comparing relative frequencies within a cohort of terms assumed to cover roughly the same semantic space over a decade.

---

[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').

[5]* means that we use all linguistic forms

Figures 2, 3, and 4 show the frequencies of the various groups of terms. An immediate observation for all three groups is that there have been big changes. Dominant terms from the earliest periods have (almost) disappeared from current official discourse, and the terms that replaced them have given way to even newer terms. An exception to this general pattern is the term *handikapp* ('handicap') which is still current.
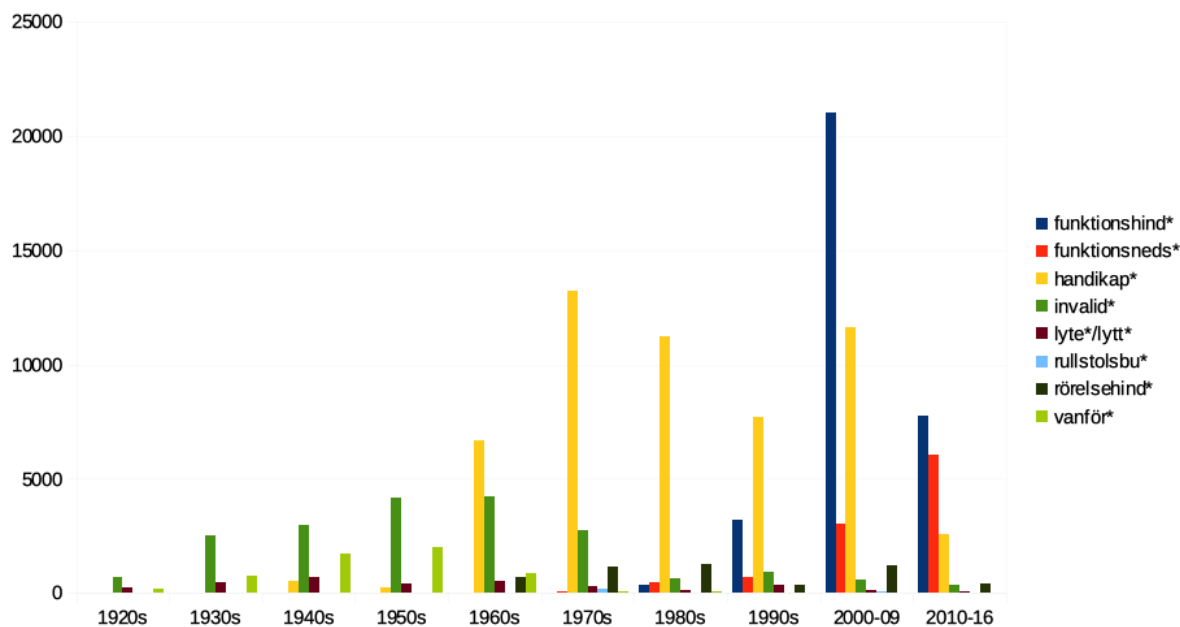

Figure 2: Usage of some Swedish Physical disability terms 1922-2016 by decades. Legend: *funktionshinder* ('disability'), *funktionsnedsättning* ('functional impairment'), *handikapp* ('handicap'), *invalid*, ('invalidity'), *lyte/lytt*, ('deformity', 'crippled'), *rullstolsburen*, ('carried by wheelchair'), *rörelsehindrad*, ('disabled wrt movement'), *vanför*, ('crippled, lame')

For instance, in the domain of Physical disability, Figure 2, we can see clear changes of dominant terms since the 1920:ies. Initially, during the period from 1920-1950 terms like *invaliditet/vanförhet* ('invalidity/lameness') dominate. Then from 1960 the term *handikapp* ('handicap') establishes itself and dominates until the term *funktionshinder/nedsättning* ('disability/functional impairment') is introduced around 2000. *Handikapp* does not disappear, however. Similar changes of dominant terms can be seen in relation to various forms of mental illness/retardation, Figures 3, and 4. For mental illness the total frequency of all terms has decreased.

The changes in dominant terms are largely in agreement with our expectations. However, while the change from *handikapp* ('handicap') to *funktionshinder/nedsättning* ('disability/functional impairment') 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, several disability models seem to be at work simultaneously.

3.4 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 co-occurrences with words in the context. The Korp concordancer[6], 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 5.

Figure 5 shows that *handikapp* ('handicap') is often accompanied by attributes referring to extent: *svår*, ('hard, difficult'), *grav*, ('grave'), *allvarlig*, ('serious'), *lätt*, ('light'), or kind: *psykisk*, ('mental'),

---
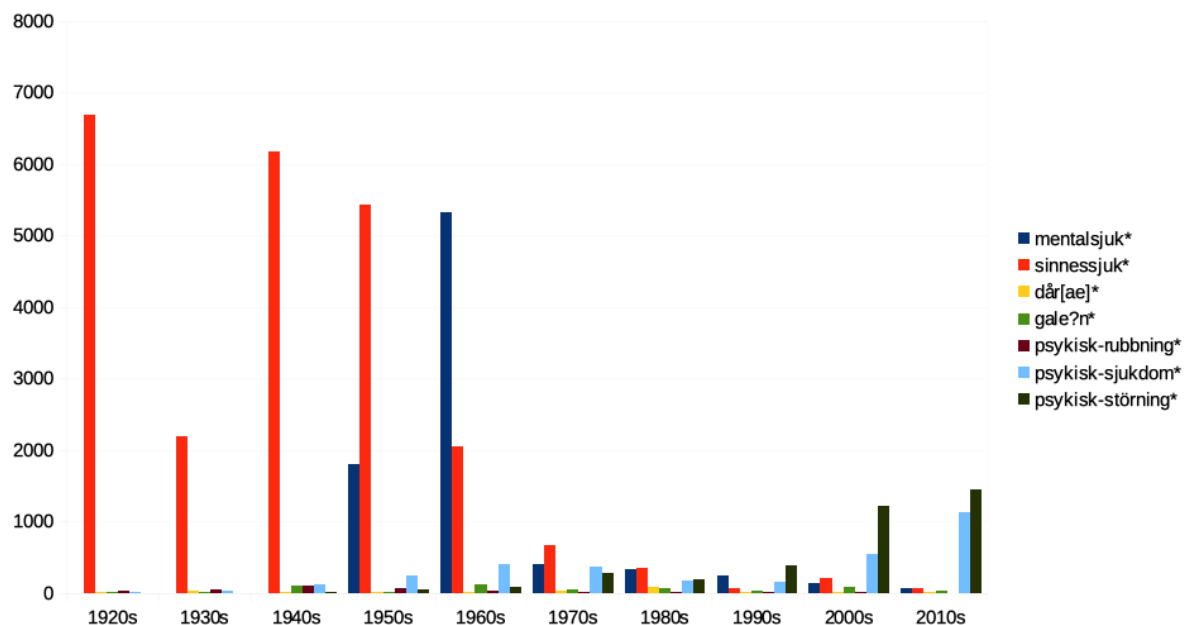
[6]https://spraakbanken.gu.se/korp

Figure 3: Usage of some Swedish *mental illness* terms 1922-2016 by decades. Legend: *mentalsjuk*, ('mentally ill'), *sinnessjuk*, ('insane'), *dåre*, ('lunatic'), *galen*, ('mad(ness)'), *psykisk-rubbning*, ('mental disorder'), *psykisk-sjukdom*, ('mental illness'), *psykisk-störning*, ('mental disorder')
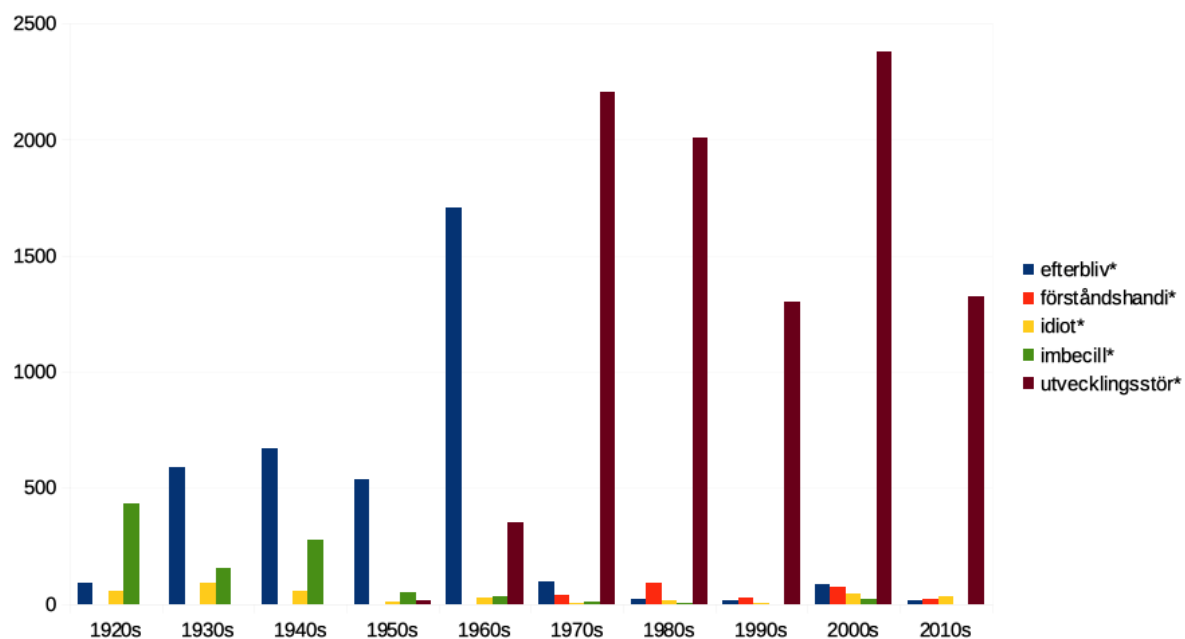
Figure 4: Usage of some Swedish *mental retardation* terms 1922-2016 by decades. Legend: *efterbliven*, ('retarded'), *förståndshandikappad*, ('intellectually handicapped'), *idiot*, ('idiot'), *imbecill*, ('imbecile'), *utvecklingsstörd*, ('mentally retarded')

fysisk, ('physical'), *neurologisk*, ('neurological'), *medfödd*, ('congenital'), *livslång*, ('life-long'). There are many overlaps with the corresponding lists for the words *sjukdom*, ('illness') and *funktionshinder* ('disability'). This clearly gives the impression that the medical model is well represented in the data.

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 det
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:
12.medfödd	15	individuell prövning och först sedan det blivit trots ett	**medicinskt handikapp** men son
13.medicinsk	31	som tidigare nämnts är det en allmän princip att den som har ett	**medicinskt handikapp** är skyldi
14.synlig	16	varande vederbörande tillräcklig principer för bedömning har ett	**medicinskt handikapp** av rätt s
15.neuropsykiatrisk	10	etsplacering dande syfte när det gäller äldre personer med något	**medicinskt handikapp** .
16.socialmedicinska	14		
17.socialmedicinskt	10		
18.lätt	23		
19.språklig	16		
20.somatisk	14		
21.osynlig	12		
22.psykosocial	9		
23.speciell	25		
24.intellektuell	10		

Figure 5: Adjectival attributes of the noun *handikapp* ('handicap'), 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.

3.5 Word embeddings

To obtain richer models we have trained word embeddings for the full corpus of SOU-reports, and for each decade. Given a sufficiently large corpus, learned vectors for frequent words give a good representation of their similarity. In the following, we focus on four Swedish words that have been used to try out the techniques. They are: *handikapp* ('handicap'), *funktionshinder* ('disability'), *funktionsnedsättning* ('functional impairment'), and *delaktighet* ('participation'). We used the SOU data from the 1970s and forward, split into decades as our corpus.

We used word2vec (Mikolov et al., 2013) with standard parameter settings as implemented in the GenSim framework[7]. Word2vec models are models created by a shallow, two-layer neural network that is trained on relevant data. As the training algorithm initializes the vectors with random numbers, ten models were trained for each decade, giving us 50 models altogether. This was done to check the stability of the models. The top three neighbouring word vectors for each term of interest was then checked in every model.

Some neighbours appeared in all ten models, but because of the random starting number there were also some variation. The found words were sorted on the basis of similarity and descending frequency.

Results are shown in Table 2. We could see a change in moving from the 1970:ies to the 1980:ies. For the terms *funktionshinder* ('disability'), and *funktionsnedsättning* ('functional impairment'), the term *handikapp* ('handicap') is one of the three closest neighbours only once in the period 1970-79. In the following decade 1980-89, *handikapp* ('handicap') is the closest neighbour for both terms. From this decade on these three terms are close neighbours in the word space. We could see from Figure 2 that the three terms are all used in the 1990:ies and later in the 2000:s, albeit with *handikapp* ('handicap') becoming the least frequent. Table 2 tells us, however, that there is close similarity in usage.

---

[7]https://radimrehurek.com/gensim/index.html

Term	1970-79	1980-89	1990-99	2000-16
handikapp	handikappade	handikappade	funktionshinder	funktionshinder
funktionshinder	funktions-rubbningar	handikapp	funktions-nedsättningar	funktions-nedsättning
funktionsnedsättning	psykiska	handikapp	funktionshinder	funktionshinder
delaktighet	gemenskap	gemenskap	jämlikhet	jämlikhet

Table 2: Nearest neighbours for selected disability terms in different decades

3.6 Temporal analogies and cross-decade projections

Vector spaces for the different decades were compared 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 which words appeared. Most analogies were to the same word, but for some words mappings shifted to new words, as listed in Table 3.

Table 4 shows the closest analogy for the term *funktionshinder* ('disability'). 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*, ('illnesses') in later years. From 1980 onwards it shows more affinity with the terms *handikapp* ('handicap') and *funktionsnedsättning* ('functional impairment').

Early	Later	Start term	New word	Nr of occurrences Max is 100
1970s	1980s	funktionshinder	sjukdomstillstånd	91
1970s	1980s	funktionsnedsättning	sjukdom	82
1970s	1990s	delaktighet	gemenskap	70
1970s	1990s	funktionshinder	sjukdomar	93
1970s	1990s	funktionsnedsättning	psykiska	95
1970s	2000s	funktionshinder	sjukdomar	79
1970s	2000s	funktionsnedsättning	sjukdom	98
1970s	2000s	handikapp	funktionsnedsättningar	96
1970s	2010s	funktionshinder	smärttillstånd	79
1970s	2010s	handikapp	funktionsnedsättningar	100
1980s	1990s	funktionshinder	handikapp	73
1980s	1990s	funktionsnedsättning	handikapp	100
1980s	2000s	funktionshinder	funktionsnedsättningar	98
1980s	2010s	funktionshinder	funktionsnedsättningar	100
1980s	2010s	handikapp	funktionsnedsättningar	93
2000s	2010s	funktionshinder	funktionsnedsättning	100

Table 3: Temporal analogies of selected terms. For pairs of periods not listed, the term was mapped onto itself.

4 Discussion

So, what can we learn from our analyses in relation to the two questions asked at the beginning?

As regards term usage, it is hard to find examples of overall terms being used in the early decades. Instead, various impairments are at work that affect a person in different ways. But as early as the 1940s we

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

Table 4: Forward temporal analogies for the term *funktionshinder* ('disability').

see various umbrella terms present in the SOU-reports. Both *funktionsnedsättning* ('functional impairment') and *handikapp* ('handicap') are now being seen in the official discourses surrounding disability. But *handikapp* ('handicap') is still used only to a limited extent. These findings indicate that disability in the early period was interpreted foremost according to a medical model understanding of disability with the problem mainly located within the individual. However, it is hard, from this analysis, to decide whether or not this meaning was attached to the *handikapp* ('handicap') term as early as the 1940s.

However, the results raise the question whether some kind of conceptual change had started to occur and whether the SOU investigators had reasons to choose a new terminology when talking about the body. This, in turn, raises questions concerning how and to what extent the social model approach was preceded by the previous conceptual landscape that somehow helped prepare the way for the social model's emphasis on the societal dimension of disability, for instance by introducing an umbrella term that puts more focus on the collective aspect of a phenomenon. What we clearly see is how the introduction of *handikapp* ('handicap') has been broadly established during the 1970s. It is interesting that this change happens in parallel with the emergence of the social model perspective gaining ground in both disability theory and disability activism. The fact that the more relative, and in that sense more progressive, disability concept is being used in official reports suggests that the official discourse to some extent is affected by wider conceptual terminology changes within the society. The results indicate that disability was now being interpreted in another way and that it, broadly, had turned into a more collective phenomenon.

At the same time, the data show that those changes did not happen overnight, but that the process of change was rather slow and that old and new concepts lived side by side for a considerable amount of time. Another thing that is evident in the data is how other concepts become increasingly important concerning disability. Not only do we witness an increase when it comes to summary concepts that underline more of a social model and relativistic approach to disability, but we can also see how concepts linking it to social policies intentions takes more place in the SOU-reports.

As to our second question, what quantitative studies can contribute, it must be understood that in disability theory the concepts that are being used are very much linked to the overall philosophy or principle guiding the way a phenomenon is supposed to be understood and interpreted. In that way, the quantitative analysis not only shows how several vital concepts in relation to disability and the body are represented in a certain text, but how these concepts change over time, which indicates that the phenomenon of disability is being surrounded by changing ideas and interpretations. It is also interesting that various concepts live side by side and that we do not see a uniform language used in the governmental reports and the text corpus under scrutiny. This, in turn, might suggest that some uncertainty how to understand and describe disability, and how a message was supposed to be presented in relation to the logic of a certain text, has prevailed.

In that respect, it is also crucial to understand the meaning of an SOU-investigation. These texts were used not only to depict something but also to perhaps suggest reforms and to reason in relation to substantiate change in some area. Thus, concepts and terminology constituted an important part of that process.

What the data tells us is that disability represents a field in which there were limited consensus on what concepts to be used during those periods, which in turn suggests that representations of disability are movable and open and that they are characterised by the negotiating nature in which they are introduced, maintained and finally abandoned and replaced. This also touches on the notion of language seen from a constructionist point of view, in which language not only reflects reality but also helps to construct and shape it. The analysis conducted here thus suggests how disability and the body, as a phenomenon, are not given in terms of their meaning. The fact that there are different terms used during the same period indicates that the official discourse concerning disability had no fixed borders.

The approach can also assist in problematizing changes over time. Figure 2 shows how the individual based concept of *funktionsnedsättning* ('functional impairment') is highly present in the SOU-reports in the last twenty years and that it even, approaches the more environmentally related concept of *funktions-hinder* ('disability') in frequency in the decade 2010-16. One question nourished by this outcome is whether this trend indicates some kind of reaction towards the social model thinking and relativistic understanding of disability, quite well established in the second half of the 1900s? This is just one example of how a quantitative content analysis can be a starting point for analysing discourses and understandings of disability in more depth.

5 Conclusions

Our analysis of the use of Swedish disability terms in resources made available by Swe-Clarin partners indicates that, while recommendations have effects, they seem to be delayed. Moreover, several frames of understanding disability live along side by side; especially the development and political use of different terms in a broader and official political context.

Most language technology methods used in this project utilise publicly available libraries, that we deliberately used without any parameter optimisation to illustrate what can be achieved "out of the box". Such optimisation, and more advanced preprocessing, would provide more reliable results, and will be carried out in future research projects on studying the *handikapp* ('handicap') concept. One obvious first improvement would be to lemmatize or stem the corpus to produce more stable word vectors. Further improvements include improved Optical Character Recognition of the SOU-reports used in the study. A comprehensive study would also require that access can be given to the full set of SOU-reports.

Using similarities in vector spaces, as a language technological tool, is a point of departure in analysing how a certain concept has been used in relation to its textual context. It underlines the fact that a certain word or concept never stands alone but derives, for instance, rhetorical meaning and content from its surroundings, which in turn can illustrate how a certain concept has changed its meaning content. The creation of the word2vec models presented above was based on central concepts linked not only to disability concepts as such but also to wider discourses guiding disability policy and the desired outcome of social policies.

From the results obtained, the conclusion can be drawn that quantitative analysis of even fairly noisy textual data can contribute with results that show a more complex picture of the history of disability concepts. This may in turn generate hypotheses to be investigated further by both quantitative and qualitative studies.

References

Handikappförbundens Centralkommitté. 1972. Ett samhälle för alla – princip- och handlingsprogram för handikappförbundens centralkommitté.

Elin Gardeström. 2006. *Handikapprörelsen och forskningen: en studie om relationerna utifrån handikapprörelsens perspektiv*. Sundbyberg: Handikappförbundens samarbetsorgan (HSO).

Marie Gustavsson Holmström. 2005. Den sociala modellen. In Bert Danermark, editor, *Sociologiska perspektiv på funktionshinder och handikapp*, pages 59–81. Lund: Studentlitteratur.

Lotta Holme. 2000. Begrepp om handikapp. en essä om det miljörelativa handikappbegreppet. In Magnus Tideman, editor, *Handikapp: synsätt, principer, perspektiv*, pages 67–78. Lund: Studentlitteratur.

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

Marie-Louise Larsson-Severinsson, Olov Andersson, and Magnus Tideman. 2009. Det relativa handikappbegreppets framväxt och etablering [elektronisk resurs].

Lars Lindberg and Lars Grönvik. 2011. *Funktionshinderspolitik: en introduktion*. Lund: Studentlitteratur.

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.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc.

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

Mike Oliver. 2013. The social model of disability: thirty years on. *Disability & Society*, 28(7):1024–1026.

SOU 2019:23. 2019. Styrkraft i funktionshinderspolitiken.

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.

Mårten Söder. 1982. Handikappbegreppet: en analys utifrån who:s terminologi och svensk debatt.

Mårten Söder. 2013. Swedish social disability research: a short version of a long story. *Scandinavian Journal of Disability Research*, 15(1):90–107.

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