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

[2] The Swedish Institute for Disability Research, Linköping University, Sweden

[3] The Swedish Institute for Disability Research, Jönköping University, Sweden

LIU LINKÖPING UNIVERSITY

Introduction

- The understanding of what disability is, and means, has been the subject of much debate. In Sweden, this has caused new recommendations to be issued from time to time as regards appropriate terminology.

- The quantity of historical Swedish texts that are digitised is increasing and can begin to be studied by digital methods.

Purpose

- Study lexical changes quantitatively in the domain of disabilities
- Use textual resources provided by Swe-Clarin
- Start collaboration between computational linguists, historians and disability researchers

Work in progress, here we report some early results.

The concept of disability (in Sweden)

- Words used to refer to disabilities and people with disabilities are manifold
  - Traditionally labelled the *medical* or *individual* model
  - challenged by the environmental turn in the *social model*
  - and later in Scandinavia the *relative*, or *relation model*
- To what extent are these views of disability found in official reports and media?

Disability terms in Swedish

- *Handikapp* ('disability') umbrella term for the many different terms that denoted special types of disability

- Environment related view *funktionsnedsättning* ('functional impairment') and *funktionshinder* ('functional impediment')

- Enabling measures signalled by terms such as *delaktighet* ('participation')

Data processing and analysis

- Official Reports of the Swedish Government (SOU, Statens Offentliga Utredningar) from 1922 to 2016
- Newspaper and literary texts used for comparison
- Texts were lowercased and stop words removed
- Grouped into decades; reports related to disability unevenly distributed over years
- Early texts contain many errors due to failing OCR
    - Word-based methods are often robust to noise.
    - General trends captured
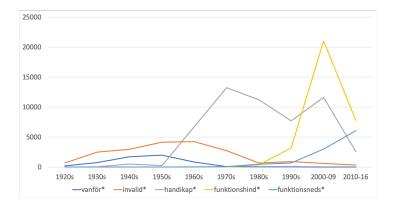   - Word embeddings showing plausible relations between terms.

Frequency changes

- Initial list of 60 words referring to disabilities and/or disabled persons over the last 100 years was manually produced by the disability researchers
- Data for the general terms: nouns or adjectives
- Comparing relative frequencies within a cohort of terms assumed to cover roughly the same semantic space over a decade.

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

Frequency results

Clear changes of dominant terms since the 1920:ies:

invaliditity/lame $\rightarrow$ disability $\rightarrow$ functional impairment/impediment

- Expected from the adoption of a relational model
- Happen quite slowly; with full force much later than one could expect
- The change is not abrupt; several disability models are at work simultaneously.

Terms in context

Adjectival attributes of the noun
handikapp, 'disability', from 1
billion tokens in the Swedish
Language Bank: the SOU-files,
all newspaper data and a corpus
of novels.

1. svår	233
2. psykisk	147
3. fysisk	148
4. grav	63
5. social	125
6. livslång	58
7. motorisk	37
8. olik	174
9. neurologisk	27
10. allvarlig	43
11. mental	23
12. medfödd	15
13. medicinsk	31
14. synlig	16
15. neuropsykiatrisk	10
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

Results from *handikapp* in context

- 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'

- Many overlaps with the corresponding lists for the words *sjukdom*, 'disease' and *funktionshinder*, 'functional impairment'.

- The medical model is well represented in the data.

Concordances

The Korp concordancer; parsed version of the SOU-texts

De som har allvarliga	**medicinska handikapp**	t ex.
STATENS OFFENTLIGA UTREDNINGAR		(stödjer ej
arbetsföra av 2:a graden innehöll individer med mera avancerade	**medicinska handikapp**	.
t enkel och arbetsförmågan kan bedömas med kännedom om det	**medicinska handikappet**	, som det
a en tillfredsställande arbetsinsats endast om hänsyn tas till deras	**medicinska handikapp**	vid val a
mma betingelser kan dessa som ovan framhållits trots avsevärda	**medicinska handikapp**	ofta gör
arbetsförhållanden för arbetstagare, som på grund av avsevärda	**medicinska handikapp**	icke kan
Det	**medicinska handikappets**	tidigare
Det	**medicinska handikappets**	som kan
hänsyn handikapp I första hand kan det	**medicinska handikappet**	påverka
Om en person trots	**medicinskt handikapp**	bedömt
individuell prövning och först sedan det blivit trots ett	**medicinskt handikapp**	men son
som tidigare nämnts är det en allmän princip att den som har ett	**medicinskt handikapp**	är skyldi
varande vederbörande tillräcklig principer för bedömning har ett	**medicinskt handikapp**	av rätt s
etsplacering dande syfte när det gäller äldre personer med något	**medicinskt handikapp**	.

LINKÖPING
UNIVERSITY

Word embeddings

- Trained word embeddings for the full corpus of
  SOU-reports, and for each decade

- Word2Vec CBOW model as implemented in
  GenSim

- 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.

Nearest neighbors for selected disability terms

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

- For *functional impairment* and *functional impediment*, *disability* is one of the three closest neighbours only once in the period 1970-79.

- In the following period 1980-89, *disability* is the closest neighbour for both terms

Global projection of vector spaces

- 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.

- Cosine similarity was checked again in each model to see if there were any new words showing up.

- The picture we got for *funktionshinder* by considering the neighbours in each decade, is confirmed.

Forward temporal analogies for *funktionshinder*

funktions-hinder	1970-79	1980-89	1990-99	2000-09	2010-16
1970-79	=	sjukdoms-tillstånd	sjukdomar	sjukdomar	smärt-tillstånd
1980-89		=	handikapp	funktions-nedsättning	funktions-nedsättning
1990-99			=	funktions-hinder	funktions-nedsättning
2000-09				=	funktions-nedsättning

- In the 1970:ies analogously to words such as *sjukdomar*, 'diseases' in later years.
- From 1980 onwards more affinity with the terms *disability* and *functional X*.

Conclusions

- Analysis of the use of Swedish disability terms in resources made available by Swe-Clarin partners
  - While recommendations have effects, they are delayed, and several frames of thinking about disability live along side by side

- Quantitative word-based investigations, of even fairly noisy textual data, corroborate and enrich qualitative approaches and reveal patterns and trends in language use in new powerful ways

www.liu.se