Analyzing Segregation Discourse in Sweden: Technological Methods and Empirical Data

Dimitrios Kokkinakis[1,2], Daniel Wojahn[3], Johan Järlehed[1]
[1]Department of Swedish, Multilingualism, Language Technology, [2]Språkbanken Text,
[3]Södertörn University
[1,2]University of Gothenburg
[1]first.last@svenska.gu.se; [3]first.last@sh.se

Abstract

This paper outlines some of the empirical resources and language technology tools to be used in the project "Language(s) of segregation: Interdisciplinary perspectives on spatial, social, and symbolic division in cities." The aim of this project is to examine the construction of segregation discourse in Sweden, its implementation as urban policy, and its impact and experience in everyday life. By integrating perspectives from linguistics, public administration, and urban ethnography, this study analyzes various forms of segregation — such as educational and residential — using large corpora to identify patterns and address related disparities. Key resources include political discourse, social media, and press coverage, while language technology tools like word vectors, network and sentiment analysis, and topic modeling will be employed. Preliminary findings provide early insights into the complex dynamics of segregation across different contexts.

1 Introduction

This paper provides an overview of the various empirical resources and language technology tools that are aimed to support the on-going "Language(s) of segregation: Interdisciplinary perspectives on spatial, social, and symbolic division in cities"[1] project. The project takes a linguistic, public administration and urban ethnographic perspective with the purpose to *analyse and demonstrate how segregation has been discursively created as a problem in Sweden, is implemented as an urban government policy and administration practice, materializes in the linguistic landscape, and is negotiated in everyday life by speakers of different languages and residents in different socio-economic areas.* Segregation comes in various forms and dimensions, such as school, labour market, age, gender, race, residential; each of these forms is influenced by distinct social, economic, and institutional factors. Analyzing these forms of segregation often involves examining extensive empirical data, which enables researchers to identify patterns and develop effective strategies to address the resulting disparities across various contexts, thereby deepening our understanding of this pressing challenge.

The textual resources described (cf. Section 4.1) are drawn from three major genres: national *parliamentary* discourse on 'segregation'; local, regional and national *Swedish press*; and the online social media forum *Flashback*. The language technology tools to employ encompass a range of techniques, including lexical and phraseological analysis using word vectors (e.g., embeddings; cf. Section 4.2) and other statistical methods, network and sentiment analysis (cf. Section 4.3) to assess political attitudes, and topic modelling to uncover underlying themes (cf. Section 4.4).

2 Background and Related Studies

Segregation is a complex phenomenon which has been studied across various disciplines, including sociology, economics, and urban studies. Depending on the type of segregation, e.g., residential, educational, occupational, or age, the data, consequences, and policy interventions may differ significantly. The sources of empirical data used in such studies often include longitudinal surveys (Leoncini, et al., 2024), social networks

[1] The project is part of the "National research programme in segregation", coordinate by the Swedish Research Council, which in turn implements the Swedish Government's strategy and action plan to counteract and reduce segregation in Sweden, and covers research in all scientific fields, as well as cross-disciplinary and multi-disciplinary approaches.

(Echenique & Fryer, 2007; Reme, et al., 2022), administrative data, interviews and ethnographic observations (Lau, 2023) or corpora, primarily used for qualitative, critical discourse analysis (Backvall, 2019). The diverse data sources examined, enable researchers to analyze patterns of segregation, assess its impact on different populations, and develop e.g., targeted policy interventions aimed at mitigating its effects and promoting social integration.

In the Swedish context, studies alongside other societal research, include Brodén's et al. (2023) investigation of the analytical benefits of tracing parliamentary discourse through neologisms as an explorative approach to identify significant patterns of 'terrorism'. Ohlsson et al. (2022) present an initial exploration of the Swedish bicameral parliament's (*Tvåkammarriksdagen*) data using a mixed methods approach, showing how the concept of 'market' was represented linguistically over time in lexical dimensions such as the term's definite forms and compounds. In an exploratory study, Pettersson Ängsal et al. (2022) extract and analyze relevant texts on 'political terror' and 'terrorism', then visualize and mapp the discourse by developing key terms and identifying associated individuals, places, groups, and states. Finally, Fridlund et al., (2023), explore how Swedish nonfiction and fiction during the Cold War represented 'political terror', using both distant and close reading methods. Fridlund et al. demonstrate how text mining and historical expertise reveal patterns and significant aspects of the cultural discourse on terrorism in Sweden.

3 Research Questions and Objectives

Some of the research questions we just began[2] to explore in the project relate to various distant reading dimensions of how 'segregation' is conceptualized, conveyed and represented within the Swedish Parliament and among members in the social media forum Flashback. The use of language technology tools can enhance the analysis, specifically by identifying subtle patterns, implicit references, and emerging trends that might otherwise be overlooked. Thus, part of the corpus processing objectives includes methods to:

- analyze the lexical manifestations of 'segregation' and closely related concepts within the text samples. This includes identifying dominant and significant collocation and representation patterns, major compounds, and any similarities or differences in the ideological perspectives of political parties as reflected in the parliamentary text. Additionally, examine the connotation polarity profile of 'segregation' to understand its semantic nuances.
- examine how the framing of 'segregation' has evolved over time in Swedish parliamentary discourse and investigate any societal and historical events that have intersected with and influenced the changing perceptions of segregation in Sweden.
- identify texts that discuss segregation indirectly, without explicitly using the term itself. This can be accomplished using word and document embeddings to find terms and phrases, and even documents related to segregation.
- track the evolution and sentiment of the term 'segregation' using temporal analysis and sentiment analysis techniques, including trend analysis over time and sentiment scoring methods to assess shifts in public perception and discourse.
- apply topic modeling to uncover underlying themes and trends.

4 Framework for the Study

This section provides an outline and status of the textual resources and computational linguistics' framework for the project. One major challenge we face concerns the selection and representativeness of the data, which can be crucial for drawing more generalizable conclusions from the sample. The segregation-related textual data so far consists of:

- *'Riksdagens öppna data' – parliamentary data*: these texts cover a period of fifty-four years (1970-2024) which includes parliamentary motions, interpellations, minutes from the chamber, reports, statements and decisions ('*Betänkande*') and other related document.
- *'Riksdagens öppna data' – policy documents* and *official reports ('Statens offentliga utredningar')*: these are reports by government commissions of inquiry.

[2] Any figures or descriptions related to the resources in this paper are based on the data collected by mid-September 2024 (the "current status"). Collection and processing are ongoing processes during Fall 2024/Spring 2025.

- *press/news/media articles*: these (will) include both local press, national newspapers, magazines and trade media sources primarily read for professional purposes (*cf*. footnote 2).
- *social media ('Flashback forum')*: these data encompass various relevant to segregation discussion threads, with pertinent metadata such as publication date and user information included for comprehensive analysis.

We currently use *Språkbanken Text (SB*[3]) as a repository for the political document collections and some of the social media and press corpus as well as other tools such as *Retriever Research*[4], to compile the rest of the *press/news/media* data. Most of the texts are publicly accessible and cover a relatively long period of time.

The whole data will be imported into SB's infrastructural tools, such as *Korp*, *Mink* and *Strix* (Borin et al, 2012), while some of the resources and exploratory tools will be available as *Colab notebooks*[5]. The SB tools allow researchers to search for word instances in the data set over the complete timeline and get immediate access to the query words in their contexts as well as getting a statistical overview of word use; information includes also a variety of metadata about these texts, such as publication date. The study of the 'segregation' concept over time will also be further examined qualitatively using the linguistic representations in context and discourse analysis.

4.1 Methodology

Regarding the empirical data, ongoing work is on place with the aim to add more textual content to these subsets, particularly for social media and press sources (*cf.* footnote 2). The Swedish parliamentary texts, accessible through Språkbanken Text, consist of a well-defined collection of documents that are diverse in structure, style, and perspective. We are currently narrowing them down using selected keywords (cf. Section 4.2). The keywords enable us to extract a subset of the whole parliamentary data which is exported, and further divided into small number of subcollections, such as *motions* and *interpellations*. Regarding the *press/news/media articles*, the original goal of the project is to collect at least 80,000 of such documents in which the term 'segregation' or one of its inflections or compounds (either as prefix or suffix element) is used. Flashback threads are extracted in which various forms of 'segregation' are the primary focus of discussion. The status of these relevant threads, some already collected from *Språkbanken Text* resources, are provided in Appendix A.

The finally collected dataset will be used for the statistical analysis of how often 'segregation' and its forms are used across different corpora, genres, and time periods. Moreover, this study examines the grammatical structures in which 'segregation' and its variants appear, with a focus on common phrases, clause structures, and sentence roles. The textual data will also be the necessary input to the technologies overviewed; sections 4.2-4.4, below.

4.2 Pre-Trained Word Embeddings

Embeddings, as vector representations of words or documents[6], offer advantages in identifying word or document similarities. For instance, by mapping words into a continuous vector space, embeddings capture semantic relationships based on context, enabling models to discern subtle nuances in meaning (Mikolov, et al., 2013). This vectorized representation allows for the measurement of word similarity through simple mathematical operations like *cosine similarity*, which can effectively identify synonymous or closely related words even when they do not share explicit lexical features.

To identify words analogous to 'segregation', we currently examine several available Swedish word embedding models, cf. Appendix B for the results of two av them *FastText* embeddings (Grave et al., 2018), and a BERT model ("Bidirectional Encoder Representations from Transformers"; Devlin et al., 2019) trained by the National Library of Sweden (KBLab). The word embeddings, which capture the collocative patterns of 'segregation' in context and analyze the surrounding words and 'segregation' lexical variations, may provide a deeper understanding of semantic relationships. This would allow for more accurate identification of word similarities and nuanced interpretations of meaning within complex linguistic structures.

[3] https://spraakbanken.gu.se/en.
[4] https://www.retrievergroup.com/product-research.
[5] A Colab notebook with the current Flashback forum part of the corpus (ca 2 400 structured instances) can be found here: https://tinyurl.com/3ek68vbx.
[6] We also plan to apply sentence embeddings using available Swedish models from KBLab, such as KB-SBERT: https://huggingface.co/KBLab/sentence-bert-swedish-cased.

4.3 Polarity Identification and Analysis

Sentiment analysis can help identify polarization in the discourse on 'segregation'. If the analysis reveals a wide range of sentiments from highly positive to highly negative, this might indicate a polarized topic. Understanding the degree of polarization can inform further analysis of the underlying reasons for divergent views on segregation. We plan to apply the Swedish VADER[7] resource for this analysis.

Thus, sentiment analysis can determine and quantify the emotional tone associated with 'segregation' in various texts. It also provides a quantitative measure of how the term is perceived in different contexts and time periods, which may reflect societal attitudes towards 'segregation'.

By doing so, we hope to track how the sentiment associated with 'segregation' changes over time, providing insights into shifts in public opinion, societal attitudes, or the impact of significant events related to the term.

4.4 Topic Modeling

Prevalent themes or topic across the data can be explored using the topic modeling, e.g., BERTopic (Grootendorst, 2022). BERTopic is a modular topic modeling framework, which utilizes pre-trained language models and applies clustering techniques to identify prevailing topics. This enables a nuanced exploration of overarching themes, offering insights into the nature of these discussions in the forum, using a state-of-the-art pre-trained language model for Swedish[8]. Topics can be enhanced by incorporating sentiment analysis (Section 4.3), adding another layer of depth to the overall analysis. Examining the sentiment within each topic where 'segregation' is discussed, we might gain a more nuanced understanding of how different thematic areas (e.g., education, social) frame and emotionally color discussions of 'segregation'. In the context of analyzing 'segregation' and its properties topic modeling can achieve several key objectives by *identifying thematic patterns*; *exploring contextual variability*; *tracking semantic evolution*; and *revealing the term's contextual and thematic associations*, supporting a more detailed analysis of its usage, meaning, and evolution.

4.5 Qualitative Analysis

The analysis will combine quantitative tools from corpus studies with qualitative approaches from discourse studies to examine the discursive construction of 'segregation' within the data. This approach will apply discourse analysis to explore how 'segregation' contributes to arguments, narratives, and rhetorical strategies, revealing its role in shaping public discourse across different contexts. The study will conduct an exhaustive corpus-assisted diachronic and comparative discourse analysis of Swedish media and political texts from 1970 to the present. This method will allow us to trace the evolution of 'segregation' over time; arguments associated with the term and the connotations of 'segregation', enhancing semantic analysis by uncovering nuanced meanings and distinguishing between its legal and social applications, thereby offering deeper insights into the broader implications of its usage.

5 Discussion and Current Status

The purpose of this paper is to provide a concise overview of the orientation and preliminary concepts regarding the objectives, framework, and project concept, while also emphasizing some of the available data and modeling resources intended for use and exploration. The project's empirical data has primarily been guided by the degree of accessibility, availability and variability in terms of both scope and representativeness of the texts. While corpus linguistics methods offer several advantages, it is crucial to remain mindful of the persistent challenge of *sample bias*, which can significantly influence the findings. This issue is not unique to quantitative approaches; but affects qualitative methods, underscoring the importance of carefully considering sample selection and representativeness in any linguistic analysis.

We plan to continue the collection and structuring of the data, with a focus on the analysis through a critical examination of how the concept of 'segregation' is addressed across different text types and genres. Comparing results from different types of texts and discourses will likely provide deeper insights into how the concept of 'segregation' is understood, framed, and communicated across various contexts.

[7] VADER stands for "Valence Aware Dictionary and sEntiment Reasoner" (Hutto & Gilbert, 2014); details about the svVADER cf Kokkinakis et al. (2023).

[8] The Swedish Sentence Transformer model "`sentence-bert-swedish-cased`, trained by KBLab is used.

Acknowledgments

This work is part of the project: *Segregationens språk: tvärvetenskapliga perspektiv på rumslig, social och symbolisk uppdelning i städer* ("Language(s) of segregation: Interdisciplinary perspectives on spatial, social, and symbolic division in cities"), financed by The Swedish Research Council, with Dnr: 2023-00667. The project both utilizes and enhances Språkbanken Text's research infrastructure, with the work being conducted in close collaboration with Språkbanken Text, which actively contributes to the organization's e-infrastructure initiatives. The *Nationella språkbanken* is jointly funded by its 10 partner institutions and the Swedish Research Council (2018-2024; Dnr 2017-00626).

References

Karin Backvall. 2019. Constructing the Suburb. Swedish Discourses of Spatial Stigmatisation. *Geographica* 21. 93 pp. Uppsala: Dep of Social and Economic Geography. ISBN 978-91-506-2742-8.

Lars Borin, Markus Forsberg and Johan Roxendal. 2012. Korp – the corpus infrastructure of Språkbanken. *In Proceedings of the 8th Conference on Language Resources and Evaluation (LREC)*. Pages 474–478. Istanbul, Turkey.

Daniel Brodén, Leif-Jöran Olsson, Mats Fridlund, Magnus P Ängsal, Patrik Öhberg. 2023. *The Diachrony of the New Political Terrorism. Neologisms As Discursive Framing in Swedish Parliamentary Data 1971–2018*. Digital Humanities in the Nordic and Baltic Countries Publications 5 (1). Oslo, Norway: 79-89. https://doi.org/10.5617/dhnbpub.10651

Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*. Pages 4171–4186. Minneapolis, Minnesota. c2019 Association for Computational Linguistics

Federico Echenique, Roland G. Fryer, Jr. 2007. A measure of segregation based on social interactions. *Quarterly J of Economics* 122, no. 2: 441-485.

Mats Fridlund, Michael Azar, Daniel Brodén, Michael McGuire. 2023. The Cultural Imaginary of Terrorism: Close and Distant Readings of Political Terror in Swedish News and Fiction During the Cold War. *The 7th Digital Humanities in the Nordic and Baltic Countries Conference (DHNB)*. Norway

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning Word Vectors for 157 Languages. *In Proceedings of the 11th Conference on Language Resources and Evaluation (LREC)*. Miyazaki, Japan. European Language Resources Association (ELRA). https://aclanthology.org/L18-1550

Mutanaarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv: https://doi.org/10.48550/arXiv.2203.05794.

Clayton J. Hutto and Eric Gilbert. 2014. A Parsimonious rule-based model for sentiment analysis of social media text. *The 18th International. Conference on Weblogs and Social Media*. (ICWSM-14). Stanford, USA.

Dimitrios Kokkinakis, Ricardo Muñoz Sánchez, Mia-Marie Hammarlin. 2023. Scaling-up the Resources for a Freely Available Swedish VADER (svVADER). *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*. Pp. 667–672. Tórshavn, Faroe Islands. https://aclanthology.org/2023.nodalida-1.66.

Mandy HM Lau. 2023. Residential Age Segregation: Evidence from a Rapidly Ageing Asian City. *J Popul Ageing*. Mar 13:1-21. https://doi.org/10.1007/s12062-023-09416-7.

Riccardo Leoncini, Mariele Macaluso, Annalivia Polselli. 2024. Gender segregation: analysis across sectoral dominance in the UK labour market. *Empir Econ.* https://doi.org/10.1007/s00181-024-02611-1

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. 2013. Linguistic Regularities in Continuous Space Word Representations. *Proceedings of NAACL-HLT.* Pages 746–751. Atlanta, Georgia, 9–14 June 2013. c Association for Computational Linguistics

Claes Ohlsson, Victor Wåhlstrand Skärström and Henrik Björck. 2022. The Market as a Concept in Swedish Parliamentary Records from 1867 to 1970: A Mixed Methods Study. *Proceedings of the Digital Parliamentary Data in Action (DiPaDA) Workshop*. Co-located with 6th Digital Humanities in the Nordic & Baltic Countries (DHNB), Uppsala, Sweden, pages 22-34. https://www.diva-portal.org/smash/get/diva2:1655748/FULLTEXT01.pdf.

Bjørn-Atle Reme, Andreas Kotsadam, Johannes Bjelland, Pål Roe Sundsøy, Jo Thori Lind. 2022. Quantifying social segregation in large-scale networks. *Sci Rep* 12, 6474. https://doi.org/10.1038/s41598-022-10273-1

Magnus Pettersson Ängsal, Daniel Brodén, Mats Fridlund, Leif-Jöran Olsson and Patrik Öhberg. 2022. Linguistic framing of political terror: Distant and close readings of the discourse on terrorism in the Swedish parliament 1993–2018. *CLARIN Annual Conference*. 69-72

Appendix A

Status of the Flashback forum (24 threads, already sampled) related to 'segregation'.

Start-of-thread	URL	Title	#Posts
2005-03-15	www.flashback.org/t191958	Segregation- bra eller dåligt? (Segregation- good or bad?)	36
2009-02-23	www.flashback.org/p15283638	Segregation är positivt! (Segregation is positive)	88
2012-03-22	www.flashback.org/t1827271	Lögnen om att socioekonomisk status skulle förklara invandrares brottsbenägenhet (The lie that socio-economic status would explain immigrants' propensity for crime)	40
2014-05-06	www.flashback.org/t2370004	Segregation i skolan (Segregation in school)	54
2016-04-22	www.flashback.org/t2712618	Är inte segregation förbaskat trevligt? (Isn't segregation damn nice?)	24
2016-08-08	www.flashback.org/t2750231	Vilka orter kan man bo i om man vill slippa mångkulturen? (Which places can you live in if you want to avoid multiculturalism?)	185
2017-05-13	www.flashback.org/t2839409	Expressen avslöjar: Att bryta segregationen är bortom det möjligas horisont (Expressen reveals: Breaking segregation is beyond the horizon of the possible)	38
2017-09-02	www.flashback.org/t2869718	Kvotera in Svenskar i segregerade områden! (Quota for Swedes in segregated areas!)	71
2018-01-19	www.flashback.org/t2909298	Invandrare föreslår tvåstatslösning för Sverige (Immigrants propose a two-state solution for Sweden)	350
2019-07-21	www.flashback.org/t3060342	Starten på segregation redan i unga år? (The start of segregation already at a young age?)	18
2020-02-21	www.flashback.org/t3121960	Invandrare skapar själva segregation (Immigrants themselves create segregation)	13
2020-03-15	www.flashback.org/p70725538	Varför vill man ha segregation istället för integration? (Why do you want segregation instead of integration?)	22
2020-08-24	www.flashback.org/t3260714	Vill vänstern återinföra segregation? (Does the left [party] want to reintroduce segregation)	17
2020-11-08	www.flashback.org/t3279828	Svensken ska känna skuld pga invandrarnas segregering (Swedes must feel guilty because of the segregation of immigrants)	74
2021-05-17	www.flashback.org/t3328036	Invandrare som klagar på segregation och problem i samhället - Eskilstuna (Immigrants who complain about segregation and problems in society - Eskilstuna)	51
2021-09-04	www.flashback.org/t3353567	Hur segregerat är Sverige om 20 år? (How segregated will Sweden be in 20 years?)	64
2021-09-16	www.flashback.org/t3356567	Miljöpartiet vill bygga flerbostadshus mitt i villaområden för att minska segregation (The Green Party wants to build apartment buildings in the middle of residential areas to reduce segregation)	67
2021-11-05	www.flashback.org/t3367953	Hur bryter vi segregationen? (How do we break segregation?)	113
2022-08-07	www.flashback.org/t3435756	150 000 invandrare måste flytta från utsatta områden för att bryta segregationen (150,000 immigrants must move from vulnerable areas to break segregation)	24
2022-09-07	www.flashback.org/t3450849	Migrationsministern: "Vi har misslyckats med segregationen i Sverige" - ANDERS YGEMAN (S) SEGREGATION (Minister of Migration: "We have failed with segregation in Sweden" - ANDERS YGEMAN (S) SEGREGATION)	28
2023-10-08	www.flashback.org/t3561722	Taxichaufförens ord förklarar segregationen. (The taxi driver's words explain the segregation.)	29
2023-10-16	www.flashback.org/t3563405	Föräldrar i Sandviken rasar över tvångsflyttande av elever i integrationens namn (Parents in Sandviken rage over the forced relocation of students in the name of integration)	273
2024-04-12	www.flashback.org/t3600309	Svenskars empatistörning - orsak till segregationen? (Swedes' empathy disorder - cause of segregation?)	105
2024-06-29	www.flashback.org/t3615439	HUR i hela helvetet kan 60% av svenskarna inte se problemen med invandring? (HOW the hell can 60% of Swedes not see the problems with immigration?)	326
2024-07-13	www.flashback.org/t3617794	Det finns en ort utan invandrare (There is a place without immigrants)	177

Appendix B

Generation of word vectors for 'segregation' (left) in two different Swedish models.

KBLab: 'bert-base-swedish-cased-new'		fastText: 'cc.sv.300.bin'	
segregationen	0.9526	segregering	0.8179
segregerar	0.9510	segregationen	0.8108
segregerade	0.9495	bostadssegregation	0.8107
segregeras	0.9406	boendesegregation	0.7694
diskrimineringssystem	0.9352	skolsegregation	0.7499
ojämlikhet	0.9344	bostadssegregationen	0.7424
stigmatisering	0.9340	segregationsproblem	0.7361
utanförskapsområden	0.9319	segregeringen	0.7250
ojämlikheter	0.9273	Boendesegregation	0.7236
instabilitet	0.9267	Segregation	0.7227
repression	0.9261	skolsegregationen	0.7130
otrygghet	0.9261	rassegregation	0.7092
vaccinklyftor	0.9259	boendesegregationen	0.7033
jämlikhetsperspektiv	0.9228	segregationspolitik	0.6991
diskriminera	0.9220	segregationens	0.6906
radikalisering	0.9206	segregerad	0.6876
diskrimineringen	0.9184	Segregationen	0.6869
urbanisering	0.9184	segregerande	0.6816
stresseffekter	0.9181	segregerade	0.6748
undanträngningseffekter	0.9173	Bostadssegregationen	0.6674
diskriminerar	0.9170	segregerar	0.6624
stressnivå	0.9169	könssegregation	0.6574
diskriminerande	0.9164	Boendesegregationen	0.6566
prevention	0.9157	segregerat	0.6541
selektion	0.9153	rassegregationen	0.6393
hederskultur	0.9152	Könssegregation	0.6317
åldersdiskriminering	0.9147	Segregering	0.6293
terrorkampanj	0.9145	segregera	0.6202
alkoholkonsumption	0.9132	segregans	0.6165
nyfattigdom	0.9126	ojämlikhet	0.6137
repressiv	0.9126	integration	0.6131
samhällspridning	0.9113	rassegregering	0.6122
extremism	0.9110	klassklyftor	0.6119
sterilitet	0.9105	rasism	0.6051
moralism	0.9100	utanförskap	0.6045
samhällseffekter	0.9095	desintegration	0.6011
samhällsförändringar	0.9092	bostadslöshet	0.6004
islamism	0.9082	marginalisering	0.5943
invandringi	0.9080	urbanitet	0.5900
stressfaktorer	0.9080	rasism.	0.5870

The two first columns to the left of the above table are the results generated by the pre-trained Swedish model ('`bert-base-swedish-cased-new`') implemented by KBLab that can be loaded from: https://huggingface.co/KBLab/bert-base-swedish-cased-new. To the right of the same table, are the results generated by the pre-trained Swedish model ('`cc.sv.300.bin`') implemented in "fastText" (right) that can be loaded from: https://dl.fbaipublicfiles.com/fasttext/vectors-crawl/cc.sv.300.bin.gz. The model's parameters were *Vector Size* = 300; *Corpus Size* = 1G and *Vocabulary Size* = 50052. The numbers shown are the cosine similarities between the queried word embedding and its nearest neighbor word embeddings, returned by the two models.