Analysing Changes in Official Use of the Design Concept Using SweCLARIN Resources

Lars Ahrenberg, Daniel Holmer, Stefan Holmlid, Arne Jönsson
Department of Computer and Information Science
Linköping University, Linköping, Sweden
`firstname.lastname@liu.se`

Abstract

We show how the tools and language resources developed within the SweClarin infrastructure can be used to investigate changes in the use and understanding of the Swedish related words *arkitektur*, *design*, *form*, and *formgivning*. Specifically, we compare their use in two governmental public reports on design, one from 1999 and the other from 2015. We test the hypothesis that their meaning has developed in a way that blurs distinctions that may be important to stakeholders in the respective fields.

1   Introduction

What is the relation between architecture and design? As concepts in the minds of speakers or as professions where stakeholders sometimes compete and sometimes join forces to achieve their goals? In this paper we try to answer such questions using the resources developed for the analysis of Swedish by Språkbanken Text and distributed through the SweClarin portal. More specifically we want to study whether there is a change in the denotations and connotations of four related words: *arkitektur*, 'architecture', *form*, 'form', *formgivning* (cf. German *Formgebung*) and *design*. In particular, are there changes in their use and, perhaps, signs of a convergence? Its rationale is a hypotheses from colleagues working in design that there has been an increased effort to place architecture and design under the same umbrella, not least from the side of the Swedish government, and that this development has been detrimental for the design field.

Dictionary definitions of the four words vary. According to one of them[1], the word *design* was first observed in Swedish in 1948. It is defined there as *konstnärlig formgivning*, 'artistic form giving' using the older term *formgivning*. Over the years *design* has established itself as a synonym of *formgivning*, and also, as will be shown, become the more frequently used of the two. The word *arkitektur*, 'architecture', on the other hand is defined as a science with a related concrete meaning as 'artistic and technical design of buildings'. Thus, *arkitektur* can be defined in terms of *design* but also in other terms. Nowadays, also *design* can be studied at universities as a separate subject. The word *form* has many meanings, one of them being 'artistic form'. It is used in this sense by the private organisation *Svensk Form*, 'Swedish Form', established in 1845, and its journal, simply named *Form*.

The studies presented in this paper make use of the Sparv text analysis pipeline[2] (Borin et al., 2016), the senSALDO[3] sentiment lexicon (Rouces et al., 2019), and the Swedish Culturomics Gigaword Corpus (Rødven Eide et al., 2016).

2   Historical background

In 1997 the Swedish government proposed an action program for an area identified as *arkitektur och formgivning*. Two years later an official governmental report (SOU), entitled *Mötesplats för form och*

---

[1]Nationalencyklopedins Ordbok, 1995 edition.
[2]https://spraakbanken.gu.se/sparv/#/sparv-pipeline
[3]https://spraakbanken.gu.se/resurser/sensaldo

design, 'A meeting-place for form and design'[4] proposed a new initiative for design. The report argued, however, that it was reasonable that its proposals had a clear connection to architecture, as its proposals related to buildings and building sites and an upcoming 'Year of Architecture', referring to the year 2001.

A few years later, in 2009, the Swedish Museum of Architecture was given a new responsibility to cover also "other fields of design" and its name was changed to ArkDes: The Swedish Centre for Architecture and Design. Its mission is "to increase knowledge of and cultivate debate about how architecture and design affect our lives as citizens."[5] More recently a new SOU-report was requested which was ready in 2015. With the title *Gestaltad livsmiljö: en ny politik för arkitektur, form och design*, 'Shaped habitat: a new policy for architecture, form and design', it brought the three concepts architecture, form, and design closer together and proposed the establishment for a new public body dealing with them jointly.

3   Data

In addition to the two SOU reports mentioned above we have used the news sections of the Swedish Culturomics Gigaword Corpus (Rødven Eide et al., 2016) from relevant time periods for comparisons. Frequencies for the terms of interest in the different datasets are shown in Table 1 and Figure 1.

	arkitektur	design	formgivning	form	arkitektur, form och design	all tokens
Gw 1990-99 (News)	793	1,008	204	10,421	0	60,037,845
Gw 2010-2015 (News)	1,595	4,042	303	28,102	0	168,998,305
SOU 1999:123	43	334	139	180	0	37,880
SOU 2015:88	318	317	26	301	0	47,345
- retokenized*	148	163	26	131	170	46,665

*The triad *arkitektur, form och design* considered as one unit, see Section 4

Table 1: Frequencies of the investigated words in different corpora.

We can see here that the ratio of *design* to *formgivning* is changing rather rapidly. For the 1990s the ratio is about 5:1, rising to to 13:1 for the period 2010-2015. Also in the SOUs the word *formgivning* is used less. Not only is it losing ground to *design* but also to *form*.

The word *arkitektur* is infrequent in the SOU from 1999 as the main topic of that report is design. However, there are many indirect references to architecture which we can see if we take into account compounds and derivations. In particular there are plenty of references to *Arkitekturmuséet*, 'The Museum of Architecture', and *Arkitekturåret*, 'The Year of Architecture'.

Figure 1: Bar chart showing relative frequencies in three corpora.

4   Analyses

Word distribution. To understand how the terms are used in general language, we looked at their distribution in the news sections of the Swedish Gigaword Corpus (Rødven Eide et al., 2016), for the periods 1990-99 and 2010-2015. Using word embeddings derived with the Gensim Word2Vec framework, we can observe the following:

- *design* and *formgivning* are close (synonyms) for both periods. The word *grafisk*, 'graphical', a common attribute to both terms, is about equally close.

- In the 1990:ies *formgivning* is a close neighbour to *arkitektur*, while *design* is further away. In the period 2010-15 the situation is reversed. In this period, *design* and *konst*, 'art' are competing for the place as closest neighbour to *arkitektur*.
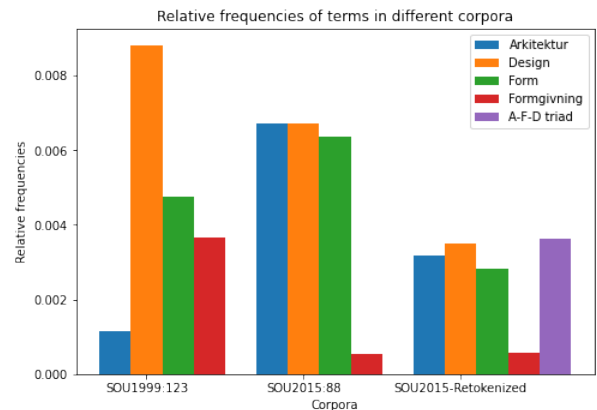
---

[4]SOU 1999:123

[5]https://arkdes.se/en/about-us/

- The word *form* does not turn up in the close vicinity of any of the other words. This is due to its many other, more common meanings such as 'type', 'sort', 'shape', 'state', 'mould'.

In order to compare the semantic space of the studied terms in the two reports, we used the temporal word analogies method as suggested by (Szymanski, 2017). This method works by transforming two vector space models to a common vector space, which acts like a link between the models, enabling the comparison of word vectors between two otherwise independent models. Thus, we can investigate shifts between the models, in the form of *"which word X in model M1 correspond to word Y in model M2?"*.

We trained a Word2Vec-model for each of the reports and applied the temporal word analogies technique to search for differences in usage of the studied terms, *architecture, design*, and *form*. Although the models themselves showed some differences when extracting and manually inspecting their most similar words, this method did not reveal any semantic shift of any of the studied terms between the two reports. It is possible, however, that this is due to the relative small size of the data and vocabularies used.

Part of the reason why the words turn up as close neighbours in vector space is that they are often coordinated, in pairs, triads or even longer ones that include words such as *konst*, 'art', and *hantverk*, 'crafts'. We also see trends of concept building via these coordinations. First in the name of the ArkDes Museum, and later, in the SOU from 2015, where the triad *arkitektur, form- och design* is very frequent and treated as such in proposals as well as in reactions to the proposals. In fact, out of the instances found in SOU 2015:88 as many as 170, or more than 50% for all three words, appear as part of this triad. For this reason we created a version of the report where this triad was treated as a single token (see Table 1).

Topic modelling of the SOUs. We have applied topic modelling to the reports to see whether they differ in their distribution of topics, using the Gensim package (Řehůřek and Sojka, 2010) on parsed versions of the reports. The number of topics per model were chosen to maximise the coherence score C_v (Röder et al., 2015), which resulted in the model for the 1999 SOU having 16 topics, and the 2015 SOU having 14 topics.

We could see for the 2015 SOU, that for the topics where *design* is among the 10 most relevant terms, so is *arkitektur*, and vice versa. For the majority of topics where this happens, *form* is also among the 10 most relevant terms.

Sentiment analysis of the SOUs. For sentiment analysis we used the Swedish SenSALDO 0.2 sentiment lexicon (Rouces et al., 2019) with sentiment scores -1, 0 and +1, and Vader (Hutto and Gilbert, 2014). What makes SenSALDO 0.2 unique in a Swedish context is that it assigns different sentiment values to different senses of a word, for instance the Swedish word *fara* can mean *danger* or *go (away)* where the former has a negative sentiment and the latter is neutral. SenSALDO comprises 12287 lexical entries where 8893 are unique words. Word sense disambiguation with the SenSALDO 0.2 lexicon is provided by the Sparv pipeline. Vader produces a compound score for each sentence, by summing the valence scores of the words according to their identified sense, and normalize this sum to be between -1 and +1.

The mean sentiment for the 1999 SOU is 0.106 and for 2015 it is 0.155. The difference is significant, $p < 0.001$ (all tests use Welch's t-test). Thus, the 2015 SOU uses overall a more positive tone.

	Form		Formgivning		Architecture		Design	
	Sentences	Sentiment	Sentences	Sentiment	Sentences	Sentiment	Sentences	Sentiment
1999	150	**0.1231**[*]	127	0.1098	42	0.173	323	**0.1256**[*]
2015	238	**0.2019**[*]	21	0.2218	246	0.2236	277	**0.2003**[*]

Table 2: Concept based sentiment. Number of sentences and mean concept sentence sentiment. [*]Significant, $p < 0.001$

We have also investigated the sentiment for each of the concepts in focus *architecture, design, form*, and *formgivning*. For each concept, sentiment is computed if the concept occurs in the sentence. The result is presented in Table 2 showing that the 2015 SOU has a more positive attitude towards the concepts *form* and *design*.

However, if we consider the triad *arkitektur, form- och design* and filter out all sentences containing it, none of the sentiments differ significantly, i.e. the significant difference for *form* and *design* in one way or another depends on the triad. The triad is not used much in 1999, only 3 occurrences, so we also compared the triad to the other concepts for 2015 and then it turns out that the difference in sentiment for *form*, 0.1490, is significantly different from the triad, 0.2745, $p < 0.001$, and for *design*, 0.1678, $p < 0.001$. That is, the triad is presented with a more positive sentiment in the 2015 SOU.

5 Conclusions

We have compared two public government reports concerning the design concept and its status as a domain for public support, financially and structurally. The first one was published in 1999 and the second one in 2015. In particular, we look at relations between the concepts of design and architecture, as they are presented in the reports and in comparable news corpora. We can see indications of a semantic convergence of the word *design* with the word *arkitektur*, especially in the latter report. Also, in the reports as in the news corpora it seems to overtake the role of the older word *formgivning*. Moreover we see that the compound term *arkitektur, form and design*, that is so frequent in the 2015 report, is significantly more positively described there, compared to *form* and *design* as individual terms, seemingly underlining a supporting policy that takes an integrative approach.

We also conclude that the tools and language resources available in the SweClarin infrastructure for analysis of Swedish texts enable comparisons of language use also over such short time spans as 20 years. In particular, we exploited the ability of the Sparv parser to identify word senses for sentiment analysis, and the Culturomics Gigaword Corpus for comparing official government reports with general language. Sentiment analysis seems to be the method that provides the most reliable results in our case, while the results from topic modelling and temporal word analogies are more uncertain due to the small dataset.

References

Lars Borin, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer, and Anne Schumacher. 2016. Sparv: Språkbanken's corpus annotation pipeline infrastructure. In *SLTC 2016. The Sixth Swedish Language Technology Conference, Umeå University, 17-18 November, 2016*.

C.J. Hutto and E.E. Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI*.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA. http://is.muni.cz/publication/884893/en.

Michael Röder, Andreas Both, and Alexander Hinneburg. 2015. Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, WSDM '15, page 399–408, New York, NY, USA. Association for Computing Machinery.

Jacobo Rouces, Nina Tahmasebi, Lars Borin, and Stian Rødven Eide. 2019. Sensaldo: Creating a sentiment lexicon for Swedish. *LREC 2018 - 11th International Conference on Language Resources and Evaluation*, pages 4192–4198.

Stian Rødven Eide, Nina Tahmasebi, and Lars Borin. 2016. The swedish culturomics gigaword corpus: A one billion word swedish reference dataset for nlp. In *Digital Humanities 2016. From Digitization to Knowledge 2016: Resources and Methods for Semantic Processing of Digital Works/Texts, Proceedings of the Workshop, Krakow, Poland*.

Terrence Szymanski. 2017. Temporal word analogies: Identifying lexical replacement with diachronic word embeddings. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 448–453, Vancouver, Canada, July. Association for Computational Linguistics.