

# Semantic tagging and the Nordic tradition of everyman's rights

Kimmo Kettunen 

School of Humanities, Finnish Language and Cultural Research,  
University of Eastern Finland, Finland

Matti La Mela

Department of Digital Humanities, University of Helsinki, Finland  
and Department of Business Studies, Uppsala University, Sweden

## Abstract

This article uses semantic tagging to analyse the Nordic concept of everyman's rights (a right of public access to nature) in protocols of the Finnish parliament. In the analysis, we use a novel tool, a lexical semantic tagger for Finnish (Finnish-language Semantic Tagger), which is used to tag key discussions about everyman's rights in the Finnish parliament. The article has two contributions as follows: first, it presents a method that combines semantic tagging and similarity analysis of corpora (keyness) for studying the formation of political concepts in large textual data. Secondly, it sheds light on the Nordic access rights and the underlying customary everyman's rights. Despite its central role in public debate, the history of the concept has not been well researched. Our analysis shows that the legislative context could be clearly detected with our approach, and that the method allowed us to describe shifts in the meaning of everyman's rights in the legislative discussion.

### Correspondence:

Kimmo Kettunen, School of Humanities, Finnish Language and Cultural Research, University of Eastern Finland, Finland.

### E-mail:

Kimmo.kettunen@uef.fi

## 1 Introduction

The Nordic countries stand out internationally for their broad public access rights to nature (Elands and Wirth, 2010). In Sweden and Finland, the rights of access are based on the customary principle of *allmansrätten*, everyman's rights, which allows everyone to roam on public and private land (La Mela, 2014; Tuunanen *et al.*, 2015; Øian *et al.*, 2018). In both countries, the right is widely used for recreational purposes, employed in tourism and country branding, allows commercial foraging of non-timber resources, and is often portrayed as a tradition or cultural heritage in the Nordic countries (Sandell and Svenning, 2011; Sténs and Sandström, 2014; Øian *et al.*, 2018). It is not surprising that the role and the legal status of everyman's rights are debated and contested in public by the different stakeholders

(e.g. Sténs and Sandström, 2013; Tuulentie and Rantala, 2013; Sténs and Mårald, 2020). Despite being a key concept in how landownership and public access rights are framed, we have little knowledge about the development of this term and its uses in public discussion. This article sets out to study everyman's rights in the Finnish parliamentary debates between 1940 and 2000. It demonstrates how the method of semantic tagging can be employed for studying the formation and the use of concepts in large textual masses. More precisely, the article asks in which legislative discussions everyman's rights have appeared, and how the concept has taken shape in the long-term—to become a building block of national identity in the Nordic countries.

In the article, the debates are studied with methods of semantic analysis and by using a novel tool FiST, a

Finnish-language Semantic Tagger (Kettunen, 2019). FiST implements the multi-lingual University Centre for Computer Corpus Research on Language Semantic Analysis System (USAS) framework for Finnish and its Finnish lexicon uses the lexicon of the English Semantic Tagger as a descriptive model (Löfberg, 2017). In practice, the tagger processes text and annotates words by their meaning class that are defined in the lexicon. In this study, we use the Semantic Tagger for annotating a subset of the parliamentary protocols, where everyman's rights were debated. Moreover, we use keyness analysis to highlight which meaning classes are over-represented and thus giving meaning to the everyman's rights. The aim is to study the changes in this syntagmatic semantic neighbourhood (Sahlgren, 2008) of the notion everyman's rights.

The article has two key contributions. First, it demonstrates how semantic tagging, available for several languages, can be used for analysing the formation of political concepts in large textual data, in our case parliamentary debates. Internationally, parliaments have published their data in digital format and digital versions of parliamentary texts have been analysed at least since the publication of the first multi-lingual EuroParl version (Koehn, 2005). These data have been used in research on political language and culture from different viewpoints, e.g. topic detection, sentiment analysis, gender studies, textometrics, political stance, and political group differences (e.g. Quinn *et al.*, 2010; Abercrombie and Batista-Navarro, 2020). Content wise annotated parliamentary corpora are still rare, and analysis of the data is many times based on current statistically oriented Natural Language Processing (NLP) approaches (e.g. Fišer *et al.*, 2018). The present article introduces the method of lexical semantic tagging for parliamentary documents. In contrast to probabilistic models, which generate meaning as a bag(s) of words (e.g. topics) commonly occurring together (e.g. Guldi, 2019), semantic labelling offers us a possibility to generalize the findings easily to meaningful categories. Moreover, semantic tagging has been employed successfully to give meaning to political discourse (Prentice, 2010) or to trace and analyse political talk (Demmen *et al.*, 2018). This article shows how the semantic tagging and the keyness analysis allow us to describe changes in the use of the term everyman's rights as debated in the Parliament.

Secondly, by studying change with the help of a semantic lexicon, we can shed light on the discursive history of the everyman's rights. Even though often narrated as an age-old institution, we know that the legal concept itself developed only from the 1930s onwards—with the ongoing urbanization, growth of free time, and new practices of access to nature (Sténs and Sandström, 2014; La Mela, 2016). Emblematically, we find the Finnish notion of everyman's rights among the neologisms presented in the Modern Finnish Dictionary of 1979 (Uudissanat, 1980). In contrast to previous scholarship which has been mainly interested in the legal concept of everyman's rights from today's perspective (e.g. Åslund, 2008), this article emphasizes the historicity of the concept and studies its uses in actual political public talk. We build on the work by Sténs and Sandström (2013, 2014), who have studied the political views on the regulation of the public access rights in Sweden and focused particularly on the non-timber forest resources. In contrast to their use of selected Swedish parliamentary motions, protocols, official inquiries, and bills, we can process digitally the whole set of parliamentary data in Finland.

The article proceeds as follows: Section 2 presents our approach of semantic tagging with the FiST tool and the digitized parliamentary data used in the article. The study builds on the idea of semantic (lexical) fields and the differences (keyness analysis) in which tagged categories are part of the semantic fields. In the section, the use and appearance of everyman's rights in our material are described. The study corpus is built around the key legislative debates where the term everyman was used. Section 3 provides the semantic analysis of everyman's rights and close reading of findings. In our study period, we see first the stabilization of a core for the everyman's rights, and later in the period, the broadening of the range of use of the term by the MPs about identities, Nordicness, and nationhood. Section 4 concludes and gives suggestions for further research.

## 2 Semantic Tagging of Parliamentary Protocols: Method and Data

The meaning of the concept of everyman's rights, in Finnish *jokamiehen/oikeus* (singular form) *jokamiehen/*

*oikeudet* (plural form), is studied in a syntagmatic way in our study (Sahlgren, 2008). This means that the semantically interpreted word surroundings (contexts) of the token *everyman*, in Finnish *jokamies*, are analysed. This way we get a distributional view of the co-occurrences of semantically categorized words in the contexts of *everyman* (Erk, 2012; Boleda, 2020). We study the change in the concept of *everyman's* rights through semantic lexical tagging and the analysis of the relative frequencies of the semantic tags, which allow us to see which semantic categories are overrepresented in the discussions about the *everyman's* rights. Our tools and methods in this study owe much to traditional corpus analytics. We produce both a study corpus (the 'everyman corpus') about the occurrences in the parliamentary proceedings where the *everyman's* rights is discussed, which is contrasted with a comparative corpus containing the complete proceedings. For studying the relative frequencies, we use keyness analysis, which highlights statistically significant differences or similarities in the relative occurrences of key item (or important) words in two corpora (Gabrielatos, 2018). Moreover, we study the significant differences also qualitatively, i.e. both distant and close reading is used in a digital humanities manner.

## 2.1 Semantic annotation schema of the data

This study uses a novel prototype lexical semantic tagger, FiST (Kettunen, 2019), for the analysis of the parliamentary protocols. We use it for production of annotated data out of the parliamentary proceedings of the Finnish parliament. The tagger has been developed using freely available components: FinnPos morphological tagger (Silfverberg *et al.*, 2016) and a 46,000-lexeme description Finnish semantic lexicon published in 2016 (Löfberg, 2017; Multilingual USAS, 2016). The semantic lexicon of the USAS framework is based on the modified and enriched categories of the Longman Lexicon of Contemporary English (McArthur, 1981).

Semantic tagging of FiST is based on the idea of semantic (lexical) fields. Wilson and Thomas (1997, p. 54) define a semantic field as 'a theoretical construct which groups together words that are related by virtue of their being connected—at some level of generality—with the same mental concept'. According to

**Table 1.** Semantic field of Money and Commerce

I Money and Commerce
I1 Money generally
I1.1 Money: Affluence
I1.2 Money: Debts
I1.3 Money: Price
I2 Business
I2.1 Business: Generally
I2.2 Business: Selling
I3 Work and employment
I3.1 Work and employment: Generally
I3.2 Work and employment: Professionalism
I4 Industry

Dullieva (2017) 'a semantic field is a group of words, which are united according to a common basic semantic component' (cf. also Lutzeier, 2006; Geeraerts, 2010). The descriptive approach taken in the USAS framework is quite generic: the USAS lexical meaning classes cover phenomena of the world quite extensively, but the inner structure of the semantic classes may vary in specificity.

Semantic lexicon of USAS is divided into 232 meaning classes or categories, which belong to 21 upper-level fields. Table 1 exemplifies one upper-level semantic field, Money and Commerce, and its meaning classes.<sup>1</sup>

FiST is a first version of the semantic tagger and it lacks some features, especially word sense disambiguation (Navigli, 2009) and proper handling of compounds. However, it achieves a lexical coverage of 82–91% in several types of modern standard Finnish texts. Worth noting is that the proceedings of the EU parliament<sup>2</sup> with its 28.6 million words got coverage of 90.9%. This example data implies clearly that parliamentary notes should be analysable with FiST. The tagger is described and evaluated in more detail in Kettunen (2019). FiST produces semantically categorized words in base form, and the semantic categorizations give a better way to inspect themes of discussion than plain words without any markings (Klebanov *et al.*, 2008; Rayson, 2008<sup>3</sup>).

We assume that the semantic annotation of our tagger provides a better way for analysis of the data than, e.g. general machine learning tools, such as Mallet,<sup>4</sup> which introduce topics as un-interpreted and ungrouped keywords taken from the actual texts.<sup>5</sup> Semantic labelling offers us the possibility to

generalize the findings easily to meaningful categories, even if the semantic categorization used is a general linguistic model that has not been tailored particularly to any specific use. From experience of the use of the semantic categorization in English, however, we know that the categorization has been useful in many kinds of studies.<sup>6</sup> Moreover, in the case of the multi-lingual USAS lexicon, semantic labelling allows us to compare the political debates in several national languages in future work, which is not directly possible with probabilistic topic models.

Out of the publications related to the UCREL English tagger the most relevant study for our setting is Klebanov *et al.* (2008). They utilized UCREL's semantic tagging for analysis of Margaret Thatcher's and Tony Blair's speeches in comparison to two other methods, unsupervised topic detection Latent Dirichlet Analysis (LDA) and a method based on supervised experimental data. As a strength of the dictionary-based method, they consider good interpretability of the categories of the lexicons. The main weakness of this type of approach is too general or abstract categories that may not reflect the actual texts and language use closely enough. Klebanov *et al.* suggest that a method based on pre-defined lexical categories is best suited for analysis in a comparative setting, where the basic setting of the texts or speech situations are similar enough. Our task fulfils this demand very well, as all the analyzable data derive from parliamentary proceedings. Another study, Murakami *et al.* (2017), found that varying granularity of the semantic classes may lead to problems in analyses.

## 2.2 Parliamentary proceedings and their semantic annotation

The parliamentary protocols of the Finnish parliament 1907–2000 have been digitized and published as open data by the Parliament of Finland in June 2018.<sup>7</sup> The documents record the work of the unicameral Parliament of Finland established in 1906 and offer a unique view on the key national legislative reforms and public issues of the 20th century. The documents contain the law proposals and petitions, preparatory and committee work, and the transcribed plenary debates. The Web pages of the Parliament contain only pdf versions of the digitized data. Text

versions of the data in our study have been produced using *pdf2otext* utility.<sup>8</sup>

With the background of FiST's lexical coverage in general and earlier morphological analysis of the parliamentary protocols (La Mela, 2020), we started to annotate them with FiST. As a result of the tagging, words of the texts are either given semantic tag annotation or tag Z99 as a mark of an unknown word for the semantic lexicon. FiST reaches a lexical coverage of 55–78% in the complete material, which, although slightly low, we consider feasible for the purposes of our study. Complete year-to-year coverage figures are given in Kettunen and La Mela (2020).

A short example of the tagger's output for a text snippet *Suomen eduskunnan vastaus. Hallituksen esitys* ('The answer of the Finnish government. The proposal of the government') would be:

- (1) *suomi* Noun Z2/Q3;
- (2) *eduskunta* Noun G1.1;
- (3) *vastaus* Noun Q2.1;
- (4) PUNCT;
- (5) *hallitus* Noun G1.1/S5+ S7.1+/S5+; and
- (6) *esitys* Noun Q2.2 X7+ K4 X9.2.

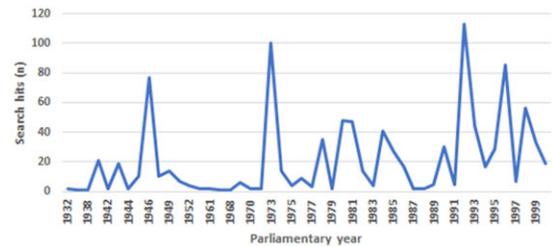
We can see different types of tags in the result. The second- and third-word tokens have unambiguous single tags; others have multiple tags, as semantic ambiguity is not resolved in the tagger. In most of the cases, the first tag is probably the right one, as the most frequent tag for each word is the first one in the semantic lexicon (Löfberg, 2017, p. 74). In the literature of word sense disambiguation, this is known as the most frequent meaning baseline, which is many times hard to outperform with disambiguation methods (Navigli, 2009).<sup>9</sup> Many of the disambiguation methods also have a bias towards the most frequent sense of the word (Preiss, 2006; Postma *et al.*, 2016). A slash tag (/) means that the word can belong to two or more categories. Pluses after tags indicate a positive position on a semantic scale.

## 2.3 Keyness analysis of FiST's output and the study corpus

As a research method, we used keyness analysis to study FiST's output for tracing changes in the semantic neighbourhood of everyman (syntagmatic relations; Sahlgren, 2008); or distributions, in the vector

space parlance (Erk, 2012; Boleda, 2020). We are interested in describing the content of everyman's rights, detecting in which legislative debates the term appears, and portraying changes in the usage of the term as used in the parliament. For obtaining the distributions, we defined a meaningful context of study. As our context window, we have chosen twenty words on the left and right side of occurrences of the token *jokamies* (everyman). From a linguistic point of view, we can justify the choice of our word window size in the everyman corpus as follows:

- quality of the data is not good, and sentence boundaries cannot be found automatically many times, thus we need to rely on mere word counting for the context (further details of the quality of the data are described in Section 2.3.1 and [Supplementary Appendix SI](#));
- most interesting things related to the notion of *jokamies* are probably said near to the word itself, either immediately before or after it. It is possible that a larger context window, say fifty words on both sides of *jokamies*, could be informative, but we leave this for a later stage;
- the window size is a 40-gram, which is eight to twenty times larger than word *n*-grams used many times in NLP (the size of *n* varying usually from 2 to 5; cf. [Manning and Schütze, 1999](#); [Lison and Kutuzov, 2017](#))—a large enough context is needed for meaningful syntagmatic relations;
- according to the sentence length statistics of [Niemikorpi \(1991, pp. 176, 177\)](#) Finnish of different genres (forty-six text types from 1960s) have a mean average sentence length of 12.83 words. The shortest mean length of sentences in Niemikorpi's data is 5.8 words, longest 24.3. Thus we can assume that our window length covers about one sentence before and after token *jokamies* at least. This can be considered a meaningful contextual window<sup>10</sup>;
- a very large context window, e.g. the whole document, would be futile, because in a large context nearly every word can co-occur with every other; on the other hand, a very narrow context, like few words, would cause a serious sparse data problem, while words co-occur very rarely with each other in a small textual window ([Sahlgren, 2008](#)).



**Fig. 1** Search hits per parliamentary year for everyman: ‘\*okamie\*’ (1932–2000)

For the selection of our study corpora, we searched for all the occurrences of everyman in the minutes of the plenary debates. The keyword used was ‘\*okamie\*’, which is the truncated keyword for different inflections of Finnish everyman. There were five wrong search hits that did not relate to everyman, which we removed. As a result, the keyword appeared 998 times in the minutes since 1932. We preprocessed the word tokens in the documents, and then tagged the corpus with FiST (presented in [Kettunen and La Mela, 2020](#)). The lexical coverage in the everyman corpus follows the trend but is somewhat higher than in the minutes of the parliament in total, being 80.6% for the complete everyman corpus. This is due to the preprocessing of the tokens and the capturing of tokens in Finnish language debates (with less technical or procedural vocabulary present). Search hits of \*okamie\* for different years in the minutes are depicted in [Fig. 1](#).

The timeline confirms the previous observations that the term everyman's rights itself became commonly used only during the second half of the 20th century ([La Mela, 2016, 2020](#)). The peaks in the [Fig. 1](#) depict the nature of parliamentary work and show how the term has been more topical during certain years and when specific legislative projects were debated. We find the first appearance of ‘everyman’ in the minutes in 1932. Obviously, the term ‘*jokamies*’, everyman, has been used in Finnish language already in the 19th century when referring to the common man, e.g. in technical or law manuals. The results show three clear peaks in discussing the everyman's rights: 1946, 1973, and 1992, which took place in different legislative and societal contexts. Based on this, we decided to focus our semantic analysis on these 3

years. Finally, the comparative corpora required in the keyness analysis were built, and they consist of minutes from 3 years preceding and succeeding the years we analyse (presented in [Kettunen and La Mela, 2020](#)). We wanted the comparative corpus to surpass the context of the legislative process where everyman's rights were discussed, but also to limit our comparison to the specific public language use of the time. After the year 1930, the Parliaments in Finland have usually functioned for 3 or 4 years (the complete term is 4 years).

### 2.3.1 Quality of the parliamentary data

[La Mela \(2020\)](#) has earlier analysed the Finnish parliamentary proceedings using LAS, a Linguistic Analysis Command-Line Tool that wraps up several existing analysis tools in a single package. La Mela's analysis implied that the word-level recognition of the data is relatively high most of the time. Between the years 1918 and 1927, however, there is a clear drop in recognition, but otherwise the recognition rate is between 70 and 90% most of the time and for the last few years clearly over 90%. The morphological recognition rate can be considered a rough estimation of the quality of the digitization ([Kettunen and Pääkkönen, 2016](#)). For clean modern standard Finnish, the recognition rate would be around 95% ([Pirinen, 2015](#)).

Tagging results of our semantic tagger are, however, lower than results of morphological recognition. Lexical coverage of the semantic tagger varies from ca. 57 to 79% (cf. Fig. 2A in [Supplementary Appendix S1](#)). Minutes of the parliament have overall a slightly better recognition rate than the whole data, but the differences are most of the time small. Average recognition rate for the minutes is 68.9%, and for the whole data 67.7% with our semantic tagger.

With detailed analysis of the data, we found some reoccurring error types in the data (cf. [Supplementary Appendix S1](#)). These errors were not easily correctable and based on the analysis we estimated that the simplest and most effective way to improve lexical coverage of the semantic tagger somewhat was to correct line-ending hyphenations in the text files. We performed this for all files of the comparative corpus 1946, 1973, and 1992. Each protocol file could contain approximately 35,000–50,000 line-ending hyphens that broke words.

## 3 Semantic Analysis of Everyman's Rights: From Access Rights to National Identities

We focused on the formation of the Finnish concept of everyman's rights in the 20th century based on the changes in the semantic neighbourhood of the term 'everyman'. In our case, we were interested in the categories that were more frequent in the everyman corpus than in the comparative corpus. [Table 2](#) shows the over-represented categories in the everyman corpus based on %DIFF metric,<sup>11</sup> and signals statistical significance for the values. The table lists categories with a value of at least 1.00, thus, categories at least two times more represented in the everyman corpus. The non-significant categories are also listed, while they may open new questions for the analysis. These categories have low frequencies in the everyman corpus.

We noted that certain FiST categories are common for the everyman corpus in all 3 years. The general fields concerning Emotional discourse (E), Government and public domain (G), Entertainment (K), Movement (M), Substances (O), and the World and environment (W) are semantic categories common to the discussions about the everyman. These categories are, however, not equally represented. As we will see, the differences reflect the themes of the specific debates in the parliament, but also highlight changes in the ways the concept of everyman's rights becomes used by the MPs.

### 3.1 Everyman's fishing rights

In the early 1940s, we find the term 'everyman's fishing' in our search results. This appears as the decade, when the term 'everyman' became used more commonly in relation to nature and its resources ([La Mela, 2020](#)). At the time, the MPs discussed the legislation about 'everyman's right to fish' (*jokamiehen kalastus/oikeus*). These were temporary fishing rights created in the war years to alleviate food shortages, and they allowed households and war evacuees (the 'everyman fishers') to fish without rights to the local fishing waters. In 1946, the debate concerned whether the rights should be continued or even made fixed. The debate on everyman's fishing is a Finnish particularity

**Table 2.** The overrepresented subfields in the everyman corpus in relation to the complete minutes ( $\pm 3$  years), for years 1946, 1973, and 1992

Discourse field	Fist subdivision	%DIFF/100		
		1946	1973	1992
<b>General</b>				
A General and Abstract	A8 Seem		<b>6.62*</b>	1.61
	A9 Getting and giving; possession	<b>1.24*</b>		
	A12 Easy/difficult			
	A14 Exclusivizers/particularizers		<b>1.50</b>	
	A15 Safety/Danger	1.35	1.74	<b>2.00</b>
B The body and the individual	B2 Health and disease	1.53	<b>2.95</b>	
E Emotional states, actions, processes	E1 General (Emotional Actions, states)		2.65	
	E2 Liking	<b>6.25</b>		
	E3 Calm/Violent/Angry		1.33	
	E4 Happy/sad		<b>2.00</b>	
	E5 Fear/bravery/shock			1.23
	E6 Worry, concern, confident	6.22		1.80
F Food and farming	F4 Farming and Horticulture	<b>4.34*</b>		1.03
G Govt and Public domain	G2 Crime, law and order	<b>1.17*</b>	<b>6.63*</b>	<b>2.59*</b>
H Architecture, buildings, houses, and the home	H2 Parts of buildings	<b>4.52</b>		
	H3 Areas around or near houses	1.91	<b>5.46</b>	
	H4 Residence	2.15	<b>2.63</b>	1.03
I Money and commerce	I4 Industry			1.10
K Entertainment, sports, and games	K1 Entertainment general	2.52	<b>32.56*</b>	2.12
	K5 Sports and games generally	1.27		<b>2.58</b>
L Life and living things	L2 Living creatures generally	<b>4.54*</b>		<b>1.36</b>
M Movement, location, travel, and transport	M1 Moving, coming and going		<b>1.20*</b>	
	M4 Movement/transportation: water	<b>41.50*</b>		<b>4.04</b>
	M5 Movement/transportation: air			<b>30.94*</b>
	M7 Places		<b>1.08*</b>	
O Substances, materials, objects, and equipment	O1 Substances and materials generally	<b>2.75</b>		
	O2 Objects generally	<b>1.51</b>		
	O4 Physical attributes		<b>1.39</b>	<b>1.15</b>
Q Linguistic actions, states, processes	Q4 The Media			1.67
W The world and our environment	W2 Light	3.94		
	W3 Geographical terms	1.32		<b>1.54</b>
	W4 Weather			1.16
	W5 Green issues		<b>48.33*</b>	<b>6.53*</b>
X Psychological actions, states, processes	X1 General (Psychological actions, states)		1.20	
	X2 Mental actions and processes		<b>1.07</b>	
	X3 Sensory		<b>1.72*</b>	
	X5 Attention		1.46	
	X7 Wanting; planning; choosing		<b>1.06</b>	

Notes: The table lists categories with a %DIFF/100 value of at least 1.00. Asterisk marks values with Bayes Factor of at least 2, and bolded values have  $P < 0.01$ .

and is not found in the Swedish discussion about access rights (cf. [Sténs and Sandström, 2014](#)).

The discussion on ‘everyman’s fishing’ in 1946 is very visible in the categories. The categories M4 (Movement: water) and F4 (Farming and Horticulture) are much over-represented, and by searching them in the word windows, we find that they label terms ‘to fish’ and ‘fishing’. These categories also tag words, such as ‘boat’ and ‘(fishing) trap’. Moreover, we see that M4 and W3 (Geographical terms) mark the terms ‘water’, ‘bodies of water’, ‘river’, and ‘sea’. In relation to the previous, the category L2 (Living creatures) labels the different fish—‘salmon’ and ‘whitefish’—and the terms ‘fish’ and ‘fish stock’. Interestingly, in contrast to the other years, it is only in 1946 that the category about possession A9 (Getting and giving; possession) is over-represented, even though one major perspective to everyman’s rights is the conflict between landowners and the nature recreationists. The category A9 has labelled common verbs, such as ‘to give’ and ‘to receive’, but we also find nouns and verbs describing property relations such as ‘owner’, ‘to own’, and ‘landowner’.

For studying the actual positions taken by the MPs, we examined the key sentences which include the categories highlighted by FiST. The category G2 (Crime, law, and order), which incorporates also ethical terms, hints at ethical positions taken by the speakers. The category G2 marks the noun ‘a right’, but also the terms ‘to steal’, ‘robbery’, and ‘confiscation’. The debate about the ‘Temporary Fishing Rights Act’ (importantly, the law text itself did not use the term everyman’s fishing) concerned the effects of the previous temporary acts, which had allowed every Finnish citizen to fish in the country, and permitted the war evacuees, who were professional fishers, to practice their profession outside their home waters. Many speakers highlighted the negative effects of the ‘everyman fishing’ to the fish stock, but also to other property which had been harmed under the right. The positive aspects—the ways in which the situation of the poor had been alleviated—were emphasized; however, the right was interpreted as temporary and related to the specific social group and the distressful situation. These aspects were present in the floor of the socialist MP Rinne, who also criticized the current private water property rights:

[.] We hear also accusations against the everyman fishers, that they bring disorder to the fishing waters, they tear and damage the traps of professional fishers, and have there even been actual thefts of fish from the traps. This all is very reprehensible, and there must be order in fisheries. We have to find the means to preserve the fish stock and even increase considerably its current level. The so-called everyman fisher wants to be part of this but let us secure them their right to fish for their own household in the waters, where they want to exercise their fishing right.

[.] The current fishing legislation is outdated and needs reform, so that the people’s alimentation does not become weaker and that the fishing conditions in our country are made more democratic and satisfactory for the everyman fishers. [.]<sup>12</sup>

## 3.2 Discussion of the early 1970s

In the early 1970s, the legislative debates where everyman appeared concerned the enactment of the Outdoor Recreation Act. At the time, recreational use of nature was framed increasingly by (foreign) tourism, motorization, and by attention towards the limits set by industrial and private holiday homes ([Sténs and Sandström, 2014](#)). At this time, the term had become a known concept in the Nordic countries. Everyman’s rights were portrayed as something unique to the countries or the Nordic area more broadly (e.g. [Rosén, 1979](#)). The term itself was rather novel, the newspapers used the term to depict modern everyman’s rights since the 1950s ([La Mela, 2020](#)). Moreover, the legal status of the term remained under scrutiny, and in the Finnish case, the term was not included in the Outdoor Recreation Act.

In our results from 1973, the idea of roaming in nature is very visible and reflects the nature of the legislative debates on outdoor recreation. The category W5 (Green issues) includes the terms ‘nature’ and ‘nature conservation’, and in the category K1 (Entertainment) we find descriptions of the recreational activities such as ‘camping (spot)’ and ‘outdoor recreation’. The categories which marked the

fishing vocabulary are not visible anymore. Also, the category G2 (Crime, law, and order) is over-represented, but in contrast to 1946, it is not related to the vocabulary of stealing: the concept of everyman's rights is written in the minutes separately as 'jokamiehien oikeus', the right of the everyman, which makes the tagger find the token 'a right' (G2).

Our findings, however, not only reflect the topics of the legislative discussion, but also confirm that the concept of everyman's rights was generally known and used in the early 1970s. Moreover, we find that the term 'everyman' has become a fixed part of the concept. As shown in Fig. 1, the term everyman was used in the parliament at the turn of the decade, and the term referred to fishing rights, to common man and to access rights to nature. The term was used in the latter meaning in 1969 by social democrat MP Mäkinen, when he talked about the 'so-called everyman's right, traditionally recognized in Nordic law' during the debate about the building and planning law (*Minutes 1969*, I, 893). According to Mäkinen, this right had allowed the access to shore areas. It was, however, in the debate about the outdoor legislation in 1973 that the term 'everyman's right' (jokamiehien oikeus) became very visible. Almost all the search hits are about the term 'everyman's right' and carry the meaning of access rights to nature. Interestingly, what might be a sign of novelty of the concept, is that in the Swedish translation of the report the term 'varmansrätten' not 'allemansrätten'<sup>13</sup> was used in the committee reports (cf. MP Söderman in *Minutes 1973*, II, 1458).

Despite having a core meaning shared by the speakers, the concept remained contested in the debate, as was expressed by MP Kaarna of the Centre (Agrarian) party:

Whatever is meant by this everyman's right, it seems that everyone aims to interpret it in their own way, some of us very broadly. We have already definitions about the everyman's right to roam in other laws; for example, it is seen that one such ordinary right is that one can pick berries and mushrooms on other's land, as long as one does not pick from there, where they are cultivated.<sup>14</sup>

The main issue debated was whether the concept of everyman's rights should be mentioned and clarified in the Outdoor Recreation Act. The government

proposal concerned only public recreational routes, areas, and regulations about camping, but in addition, the social democrat MP Mäkinen *et al.* had presented a legislative motion which expanded the Act to cover roaming on private land. The motion noted how it was necessary to enforce the everyman's rights in the outdoor recreation legislation, this right had been effective as customary law since time immemorial and was shared by the Nordic countries (motion by MP Mäkinen *et al.*, Annexes 1972, 1703). This became the dividing line in the discussion; left-wing MPs emphasized how the decision to leaving out regulations about roaming on private land would endanger the whole right, whereas the centre-right highlighted how the public recreational routes improved possibilities of recreation and still kept the customary everyman's rights in force. The idea of the law motion was crystallized by the socialist MP Rytönen in his floor:

The motion is not about stepping towards the abolition of the holy land property rights, not a single step towards that direction. It means in practice that the everyman's rights about a freer right to roam, to forage berries and mushrooms regardless of the land property relations would be written in this law. It is about time to do so and do justice for those millions of citizens, who do not own land, but love the Finnish nature given to the Finns and want to roam in nature for recreation and healthiness.<sup>15</sup>

The motion was rejected also in the parliament, but MP Mäkinen proposed the parliament to include a statement about everyman's rights in the law description. This was approved and the statement that 'common use rights over land and water, everyman's rights, that had belonged to the country's citizens for time immemorial would remain intact' found its place in the parliament's response in approving the Act (*Minutes 1973*, II, 1776, 1785).

### 3.3 1990s and the European Community

Moving towards the 1990s, the public access rights became viewed in the European context. The debates about the EC membership in both Finland and Sweden in the early 1990s brought worries about the potential restrictions the EC regulations would impose on these traditional, even age-old access rights (*Sténs and Sandström, 2014; La Mela, 2016*, footnote

860). Emblematically, the Nordic Council of Ministers published the work titled ‘Allemansrätten i Norden’ (Zettersten *et al.*, 1997), which built a narrative of a common cultural heritage of access rights shared by all the five Nordic countries.

In our results from 1992, we find the same key categories as in the previous years. The categories W5 (Green issues), M4 (Movement: water), K5 (Sports and games), and L2 (Living creatures) mark terms concerning the nature and recreation such as ‘nature (conservation)’, ‘natural resources’, ‘to fish’, ‘to ride’, ‘to hunt’, and different animal names. However, the categories are in general less over-represented and less statistically significant than in 1973 or 1946. Also, the most over-represented category M5 (Movement: air) is not very frequent and has captured the terms ‘water plane’ and ‘helicopter’, which were raised by MP Karhunen of the centre-right National Coalition. Karhunen joked how the ‘everymen’ would need aircraft to be able to enjoy the lake areas, while the lakefronts were constructed full of private cottages. Importantly, her speech was not related to lakeside protection issues as such, but MP Karhunen took part in the debate about Finland’s EC membership and property rights of foreigners. MP Karhunen contrasted the Finns’ interests in buying wineries in Italy or France, with the aim to restrict foreigners to invest in lakeside properties (Minutes 1992, I, 546).

In fact, in these legislative debates—which regarded hunting legislation, Finland’s EC membership, and foreign property and company ownership—the concept of everyman’s rights was not related directly to nature use, but it was used in debates regarding the country’s international status and citizens’ rights. Moreover, we find the expansion of the category A15 (Safety/danger) which has marked the terms ‘threat(en)’, ‘(en)danger’, ‘secure’. The category G2 (Law and order) includes discussion about ‘a right’, ‘legislation’, and ‘law’, but this is not related anymore to the written form, but the concept is written together, *jokamiehenoikeus*. In contrast to 1973 or 1946, the concept of everyman’s rights did not only refer to the tensions between the landowners and the outdoor recreationists, but the concept of everyman’s rights could be employed to point at national values and the national tradition of access rights. This was how MP Karhunen (above) employed the concept, and how the Green MP Paloheimo criticized the

property laws which he saw too unregulated concerning foreign land ownership:

Some have comforted the situation by noting how we have the everyman’s rights, and how it does not matter who owns the land in Finland, while everyone can roam freely, and everybody can forage mushrooms and berries etc. This is truly the practice now, [...], but this is no law, it is a mere practice.

[...] and as regards to the everyman’s right, of course after we have joined the EC, we have to start following the same European customs as they follow there. I can just imagine how the same circles [who have changed their minds on the liberal property laws] come and tell people—when we eventually start to dismantle the everyman’s right [...]—that we cannot be such hillbillies, that we have to have some euro habits. One does not roam in Europe on other’s land. As the people have come here from Europe and have bought land with high price, surely, they need to have the right to control it. We must change our practices. [...]<sup>16</sup>

## 4 Conclusion

The article has shown how semantic tagging can be used for studying political concepts in big textual data. We have used FiST—a lexical semantic tagger for Finnish—for analysing the concept of everyman’s rights (Nordic public access rights to nature) in the proceedings of the Parliament of Finland. We formed a sub corpus with  $\pm 20$ -word windows around the occurrences Finnish term *jokamies*, everyman, and which we tagged with FiST. We focused on specific years and their political context, where key debates took place in the parliament: 1946, 1973, and 1992. The tagging results were approached by looking at the relative occurrence of the semantic categories (keyness analysis), and by close reading the results to understand party political standpoints on the matter; the Finnish data are not currently annotated by speaker and party, which would enable to use this political metadata directly for refining the tagging results.<sup>17</sup>

We found that during our period ‘everyman’ became associated with the expression everyman’s

rights, which embodied the core meaning of ‘right of public access to nature’. Already in the 1940s, the term referred to access rights and concrete outdoor environments and was used to describe the wartime temporary ‘everyman’s fishing rights’. In the early 1970s, the categories regarding nature, concrete outdoor environments, and movement became central. Despite the common core acknowledged by the speakers, the MPs were divided about the scope of the right. In the early 1990s, the concept was used beyond the domestic access rights discourse, to appear in debates about national culture and institutions in the context of Finland’s EC membership. We can say that the modern meaning of everyman’s rights was shared in the public debate at least in the early 1970s, and in the 1990s, the concept carried values related to national identity. These results open interesting avenues for further research. What can we say about the chronology and the nordicness of the concept if we study the uses of the concept between the Nordic countries?

We used FiST for studying changes in the meaning and usage of the concept of everyman’s rights. We were able to capture features about the general meaning given to the concept, but also about the ongoing law projects, where the term everyman was used. In contrast to topic modelling methods, which require well founded model parameters and explanation of the word cluster interpretation, cf. Aletras *et al.* (2017), FiST provided an efficient description of the corpus through the semantic categories. The usefulness of semantic tagging for the analysis was in the description of the semantic content, which appeared particularly applicable when identifying differences and similarities. However, close reading of the relevant categories identified by FiST was a necessary step for validating the results and controlling the problems related to classification (Kettunen and La Mela, 2020). In accordance with Kilgarriff (2001), we also emphasize that corpus similarity (or difference) is complex. We have shown with generally used measures the differences and similarities in the everyman corpus and study corpora and believe that these show the different meanings of the concept of everyman in the ca. 50-year time span. We see, thus, that this approach combining semantic tagging, keyness analysis, and close reading of the results offers a fruitful way for studying changes in political concepts in large textual data.

## Acknowledgements

The analyses of the Parliamentary data were run in the Taito cluster of the CSC—It Center For Science Ltd.

## Supplementary Data

Supplementary data are available at *DSH* online.

## References

- Abercrombie, G. and Batista-Navarro, R. (2020). Sentiment and position-taking analysis of parliamentary debates: a systematic literature review. *Journal of Computational Social Science*, 3: 245–70.
- Aletras, N., Baldwin, T., Lau, J. H., and Stevenson, M. (2017). Evaluating topic representations for exploring document collections. *Journal of the Association for Information Science and Technology*, 68(1): 154–67.
- Åslund, Å. (2008). *Allemansrätten och Marknyttjande: Studier av ett Rättsinstitut*. Doctoral dissertation, Linköpings Universitet.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4): 77–84.
- Boleda, G. (2020). Distributional semantics and linguistic theory. *Annual Review of Linguistics*, 6:213–34.
- Demmen, J., Jeffries, L., and Walker, B. (2018). Charting the semantics of labour relations in House of Commons debates spanning two hundred years: a study of parliamentary language using corpus linguistic methods and automated semantic tagging. In Kranert, M. and Horan, G. (eds), *‘Doing Politics’: Discursivity, Performativity and Mediation in Political Discourse*. Amsterdam: John Benjamins, pp. 81–104.
- Dullieva, K. (2017). Semantic fields: formal modelling and interlanguage comparison. *Journal of Quantitative Linguistics*, 24(1): 1–15.
- Elands, B. and Wirth, V. (2010). Cross-European comparison. In Pröbstl, U., Wirth, V., Elands, B. and Bell, S. (eds), *Management of Recreation and Nature Based Tourism European Forests*. Berlin, Heidelberg: Springer, pp. 141–74.
- Erk, K. (2012). Vector space models of word meaning and phrase meaning: a survey. *Language and Linguistics Compass*, 6 (10): 635–53.
- Fišer, D., Eskevich M., and de Jong, F. (eds) (2018). *‘ParlaCLARIN: Creating and Using Parliamentary*

- Corpora', in *ParlaCLARIN 2018 Workshop Proceedings*. [http://lrec-conf.org/workshops/lrec2018/W2/pdf/book\\_of\\_proceedings.pdf](http://lrec-conf.org/workshops/lrec2018/W2/pdf/book_of_proceedings.pdf) (accessed 1 August 2020).
- Gabrielatos, C.** (2018). Keyness analysis: nature, metrics and techniques. In Taylor, C. and Marchi, A. (eds), *Corpus Approaches to Discourse: A Critical Review*. Oxford: Routledge, pp. 225–58.
- Geeraerts, D.** (2010). *Theories of Lexical Semantics*. Oxford: Oxford University Press.
- Guldi, J.** (2019). Parliament's debates about infrastructure: an exercise in using dynamic topic models to synthesize historical change. *Technology and Culture*, **60**(1): 1–33.
- Hakulinen, A., Karlsson, F., and Vilkuna, M.** (1980). *Suomen tekstilauseiden piirteitä: kvantitatiivinen tutkimus*. Publications of the Department of General Linguistics No. 6. Helsinki: University of Helsinki.
- Kettunen, K.** (2019). 'FiST—Towards a Free Semantic Tagger of Modern Standard Finnish', in *Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages 2019*. <http://aclweb.org/anthology/W19-0306> (accessed 1 August 2020).
- Kettunen, K. and La Mela, M.** (2020). 'Digging Deeper into the Finnish Parliamentary Protocols: Using a Lexical Semantic Tagger for Studying Meaning Change of Everyman's Rights (allemansträtten)', in *CEUR Workshop Proceedings, 5th Conference of the Association Digital Humanities in the Nordic Countries*, pp. 63–80. <http://ceur-ws.org/Vol-2612/> (accessed 1 August 2020).
- Kettunen, K. and Pääkkönen, T.** (2016). 'Measuring Lexical Quality of a Historical Finnish Newspaper Collection—Analysis of Garbled OCR Data with Basic Language Technology Tools and Means', in *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)* [http://www.lrec-conf.org/proceedings/lrec2016/pdf/17\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2016/pdf/17_Paper.pdf) (accessed 1 August 2020).
- Kilgariff, A.** (2001). Comparing Corpora. *International Journal of Corpus Linguistics*, **6**(1): 97–133.
- Klebanov, B. B., Diermeier D., and Beigman, E.** (2008). Automatic annotation of semantic fields for political science research. *Journal of Information Technology & Politics*, **5**(1): 95–120.
- Koehn, P.** (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *MT Summit 2005*. <http://home.pages.inf.ed.ac.uk/pkoehn/publications/europarl-mtsummit05.pdf> (accessed 1 August 2020).
- La Mela, M.** (2014). Property rights in conflict: wild berry--picking and the nordic tradition of Allemansträtt. *Scandinavian Economic History Review*, **62**(3): 266–89.
- La Mela, M.** (2016). *The Politics of Property in a European Periphery. The Ownership of Books, Berries, and Patents in the Grand Duchy of Finland 1850–1910*. Doctoral dissertation, European University Institute. <http://dx.doi.org/10.2870/604750> (accessed 1 August 2020).
- La Mela, M.** (2020). Tracing the emergence of Nordic allemansrätten through digitised parliamentary sources. In Fridlund, M., Oiva, M. and Paju, P. (eds), *Digital Histories: Emergent Approaches within the New Digital History*. Helsinki: Helsinki University Press, pp. 181–197.
- Lison, P. and Kutuzov, A.** (2017). 'Redefining Context Windows for Word Embedding Models: An Experimental Study', In *Proceedings of the 21st Nordic Conference of Computational Linguistics 2017*, pp. 284–288. <https://www.aclweb.org/anthology/W17-0239> (accessed 1 August 2020).
- Lutzeier, P. R.** (2006). Lexical fields. In Allan, L. (ed.), *Concise Encyclopedia of Semantics*. Oxford: Elsevier, pp. 470–73.
- Löfberg, L.** (2017). *Creating Large Semantic Lexical Resources for the Finnish Language*. Doctoral dissertation, Lancaster University. 10.17635/lancaster/thesis/3 (accessed 1 August 2020).
- Manning, C. D. and Schütze, H.** (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Magnusson, M., Öhrvall, R., Barrling, K., and Mimno, D.** (2018, April 4). Voices from the Far Right: A Text Analysis of Swedish Parliamentary Debates. 10.31235/osf.io/jdsqc (accessed 24 February 2021)
- McArthur, T.** (1981). *Longman Lexicon of Contemporary English*. London: Longman.
- Minutes.** (1969, 1973, 1992). Minutes of the Parliament of Finland. Suomen valtiopäiväasiakirjat. [Online] <https://avoindata.eduskunta.fi/#/fi/digitoidut/> (accessed 1 August 2020).
- Multilingual USAS** (2016). [Online] <https://github.com/UCREL/Multilingual-USAS> (accessed 1 August 2020).
- Murakami, A., Thompson, P., Hunston, S., and Vajn, D.** (2017). 'What is this corpus about?': Using topic modeling to explore a specialised corpus. *Corpora*, **12**(2): 243–77.
- Navigli, R.** (2009). Word sense disambiguation: a survey. *ACM Computing Surveys*, **41**: 10–69.
- Niemikorpi, A.** (1991). Suomen kielen sanaston dynamiikka. *Acta Wasaensia*. Vaasa: Universitas Wasaensis, **26**.
- Pirinen, T. A.** (2015). Development and use of computational morphology of Finnish in the open source

- and open science era: notes on experiences with Omorf Development. *SKY Journal of Linguistics*, **28**: 381–393.
- Postma, M., Izquierdo W., Agirre, E., Rigau, G., and Vossen, P.** (2016). ‘Addressing the MFS Bias in WSD Systems’, In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, pp. 1695–1700. <https://www.aclweb.org/anthology/L16-1268.pdf> (accessed 1 August 2020).
- Preiss, J.** (2006). A detailed comparison of WSD systems: an analysis of the system answers for the SENSEVAL-2 English all words task. *Natural Language Engineering*, **12**(3): 209–28.
- Prentice, S.** (2010). Using automated semantic tagging in Critical Discourse Analysis: a case study on Scottish independence from a Scottish nationalist perspective. *Discourse & Society*, **21**(4): 405–37.
- Quinn, K. M., Monroe, B. L., Colaresi, M., Crespin, M. E., and Radev, D. R.** (2010). How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science*, **54**(1): 209–28.
- Rayson, P.** (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, **13**(4): 519–49.
- Robertson, F.** (2019). ‘A Contrastive Evaluation of Word Sense Disambiguation Systems for Finnish’, in *Proceedings of the Fifth International Workshop on Computational Linguistics for Uralic Languages 2019*. <http://aclweb.org/anthology/W19-0304> (accessed 1 August 2020).
- Rosén, B.** (1979). *Allemansrätt, allemanskyldighet*. Stockholm: Rabén & Sjögren.
- Sahlgren, M.** (2008). The distributional hypothesis. *Rivista di Linguistica*, **20**(1): 33–53.
- Sandell, K. and Svenning, M.** (2011). *Allemansrätten och dess Framtid: Utredning om Allemansrätten*. Stockholm: Naturvårdsverket.
- Silfverberg, M., Ruokolainen, T., Lindén K., and Kurimo, M.** (2016). FinnPos: an open-source morphological tagging and lemmatization toolkit for Finnish. *Language Resources and Evaluation*, **50**: 863–78.
- Sténs, A. and Sandström, C.** (2013). Divergent interests and ideas around property rights: the case of berry harvesting in Sweden. *Forest Policy and Economics*, **33**: 56–62.
- Sténs, A. and Sandström, C.** (2014). Allemansrätten in Sweden: a resistant custom. *Landscapes*, **15**(2): 106–18.
- Sténs, A. and Mårald, E.** (2020). ‘Forest property rights under attack’: actors, networks and claims about forest ownership in the Swedish press 2014–2017. *Forest Policy and Economics*, **111**: 1–13.
- Tuulentie, S. and Rantala, O.** (2013). Will ‘Free Entry into the Forest Remain?’ Argumentation Analysis of Finnish and Swedish Discussions on Everyman’s Rights. In Müller, D. K., Lundmark, L., and Lemelin, R. H. (eds), *New Issues in Polar Tourism. Communities, Environments, Politics*. Dordrecht: Springer, pp. 177–88.
- Tuunanen, P., Tarasti, M., and Rautiainen, A.** (eds). (2015). *Everyman’s Rights and the Code of Conduct on Private Land*. Finnish Ministry of the Environment. <https://julkaisut.valtioneuvosto.fi/handle/10138/159060> (accessed 1 August 2020).
- Uudissanat: Nykysuomen Sanakirja.** (1980) [1979]. *Kotimaisten Kielten tutkimuskeskuksen kielitoimisto, Suomalaisen Kirjallisuuden Seura*. Porvoo: WSOY.
- Wilson, A. and Thomas, J.** (1997). Semantic annotation. In Garside, R., Leech, G., McEnery, T. (eds), *Corpus Annotation: Linguistic Information from Computer Text Corpora*. New York: Longman, pp. 53–65.
- Zettersten, G., Scheibel Nielsen, H. H., Nordberg, L., Snaebjörnsdóttir, H., Thaulow, M., and Ahlström, I.** (1997). *Allemansrätten i Norden*. Köpenhamn: Nordisk ministerråd.
- Øian, H., Fredman, P., Sandell, K., Sæþórsdóttir, A. D., Tyrväinen, L., and Søndergaard Jensen, F.** (2018). *Tourism, Nature and Sustainability. A Review of Policy Instruments in the Nordic Countries*. [Online] Nordic Council of Ministers. [www.norden.org/nordpub](http://www.norden.org/nordpub) (accessed 1 August 2020).

## Notes

- 1 USAS Semantic Tag Set, <http://ucrel.lancs.ac.uk/usas/USASsemanticTagset.pdf>. Tagset translated in Finnish <https://github.com/UCREL/Multilingual-USAS>.
- 2 <http://www.statmt.org/europarl/archives.html#v6>.
- 3 Generalization to semantic fields can also help in data sparseness problem: single words may have very few occurrences in the texts, but semantic fields have more occurrences and thus differences or similarities may be easier to establish in the data (cf. Kilgariff’s discussion on data sampling, 2001, p. 246).
- 4 <http://mallet.cs.umass.edu/>.
- 5 The same applies also for one of the most used unsupervised machine learning topic detection methods, LDA (Blei, 2012). The bag(s) of words given as topics by the

- topic modelling software can be grouped under common themes or titles by a group of evaluators afterwards, but that is not without its problems. Results of LDA are also partly dependent on choice of parameters and their tuning. Automatic creation of more abstract topic descriptions out of topic word lists is possible, but needs ontology or some other resource (Aletras *et al.*, 2017). In contrast to our approach of semantic tagging, we find the strengths of the topic modelling method in how the granularity of the analysis can be adjusted (cf. Guldi 2019). The method of seeded topic modelling would appear to us as a promising way to study specific themes in large textual corpora such as parliamentary debates (cf. Magnusson *et al.*, 2018).
- 6 Cf. publication listing at <http://ucrel.lancs.ac.uk/wmatrix/#apps>.
  - 7 <https://avoindata.eduskunta.fi/#/fi/digitoidut/>.
  - 8 <http://www.xpdfreader.com/>.
  - 9 Robertson (2019) has implemented and evaluated different well known disambiguation methods for Finnish data. But as these tools use Wordnet's semantic descriptions as their meaning representation, these are not usable with our approach.
  - 10 In another study from the late 1970s (Hakulinen *et al.*, 1980) with neutral factual text from newspapers the mean average length of 5016 sentences is 13.3 words.
  - In a random one million sentence news data collection of 2011 from the Leipzig corpus (<http://wortschatz.uni-leipzig.de/en/download/>), the mean average length of sentences is 11.8 words.
  - 11 The %DIFF value signals the normalized frequency of a category in the everyman corpus in comparison to the normalized frequency in the comparative corpus (complete minutes of  $\pm 3$  years): value 1 is twice the frequency of a category in the everyman corpus, value 2 three times the frequency. (Gabrielatos, 2018, p. 236). The formula for counting is  $((NFC1 - NFC2) * 100) / NFC2$ , where NFCs are normalized frequencies of items in the two corpora.
  - 12 Minutes 1946, I, 210.
  - 13 'Var man' or today's 'envar' means everyman in the collective sense, whereas 'alleman' also refers to 'allmän' and 'allmänningar' (common, the commons). La Mela, 2014, pp. 271–72.
  - 14 Minutes 1973, II, 1454.
  - 15 Minutes 1973, II, 1530.
  - 16 Minutes 1992, VI, 5764–65.
  - 17 The Finnish parliamentary data are currently being transformed into linked open data in the research project 'Semantic Parliament', which will improve the usability of the data. <https://seco.cs.aalto.fi/projects/semparl/en/>.