

Studying Concept Shift in Political Ontologies

Shenghui Wang and Stefan Schlobach and Janet Takens and Wouter van Atteveldt¹

Abstract.

For some years now ontologies have been used in Social Science, e.g., in the annotation of newspaper articles for disambiguating concepts within Media Analysis. The static view of those ontologies becomes problematic when they are associated with instances over distant years. These ontologies and annotations have now become objects of study in their own right, as they implicitly represent the shift of meaning of political concepts over time. In this paper, we address the problem of concept shifting from different aspects: extensional correlations between concepts from the same and different time periods, as well as intensional links, provided by hierarchical information and manually built links between concepts. We carry out an empirical study of concept shift in a case-study from Communication Science on a corpus with ontologies describing the Dutch election campaigns since 1994. Different types of concept shift are identified and the preliminary results lead to more open questions which needs to be investigated further.

1 Introduction

Large datasets of coded textual content are often required to answer research questions in Communication Science. Traditionally, textual documents are annotated with controlled vocabularies (of increasing expressiveness). Gradually, communication scientists adopt the Semantic Web techniques in their coding routine [9]. The idea is to code the meaning of sentences and articles in a formalised graph representation (called NET) similar to RDF triples. In these triples actors and issues are taken from an ontology, and the predicate usually represents opinions and moods. Take as an example a sentence in a newspaper article during the Dutch election period in 2006.

Het Openbaar Ministerie (OM) wil de komende vier jaar mensen- handel uitroeien. (*The Justice Department (OM) wants to eliminate human trafficking within the next four years.*)

The sentence is coded as `<om, -1, human trafficking>`, where `om` and `human trafficking` are two concepts in the ontology, while `-1` indicates the Justice Department is negative about human trafficking. After the whole corpus is triplified, communication scientists then do the analysis over such RDF graphs.

However, these datasets are often comprised of textual content from different time periods or genres. Similar to the dataset created from newspapers during five recent Dutch election campaigns from 1994 to 2006, the Manifesto Research Group is creating a dataset of 2162 topic-coded party manifestos from 25 countries during the period from 1945 to 2006 [1, 4]. Another example is the Comparative Policy Agendas project, which is combining coded media content with manifestos, legislative texts, and government press statements

of the last decades [3]. Since large manual coding projects require a lot of time, energy and money, it is often only possible to create such data sets by combining the efforts of coders from different time periods, research groups or even countries.

In order to reach interoperability between the datasets coded with different ontologies, or during different periods of time or even from different countries, there are two issues which need to be addresses: *ontology merging* and *concept shift*:

Ontology merging. If different datasets are combined to create a larger dataset, the ontologies or coding schemas used in both datasets need to be the same. If they are different, they need to be merged by identifying which differently named fields correspond to each other, or, more likely, creating more general concepts which subsume concepts from both ontologies. For example, if the ontology of one dataset contains a concept `illegal immigration`, and the other contains `labour migrants`, we can merge the ontologies by creating a more general concept `immigration` that contains both concepts. However, if good mappings do exist, it can be difficult and time-consuming to create these mappings for large ontologies.

Concept shift. Even if data from different datasets is coded using the same ontology, or if a good mapping has been found between the different ontologies, there is the danger that the concepts have a different semantics in different datasets. Interpretation of these concepts can vary depending on the time period, the location or the cultural background of the coders, which leads to the problems of reliability [5, 6]. In this paper, we do not consider this reliability issue and assume the interpretation of concepts are more or less stable within a certain period of time. We focus on whether the semantics of concepts shift over time. For example, we study whether the concept `feminism` shift in its meaning from centuries ago to today. We will investigate methods which can detect and represent such shift.

2 Related work

The problem addressed here is related to a broader problem of *ontology change*, which refers to the “problem of deciding the modifications to perform upon an ontology in response to a certain need for change as well as the implementation of these modifications and the management of their effects in depending data, services, applications, agents or other elements” [2]. Flouris *et al.* classified the problem of ontology change into 11 subfields, including *ontology evolution* and *ontology versioning*. The existing research on ontology evolution and versioning mainly addresses this problem at the macro ontological level, that is, the effect of certain change operations over the ontology elements, including concepts, relations and

¹ Vrije Universiteit Amsterdam, email: {swang, schlobac}@few.vu.nl, {JH.Taken, WH.van.Atteveldt}@fsw.vu.nl

instances, as well as the interoperability issue between different variants (versions) over time.

In this paper, we look at the change problem at a micro level, that is, the underlying semantics of concepts (which can also be extended to other elements of ontologies, such as relations). Under the same labels, the semantics of concepts may change over time. We expect this to be more likely the case for the extensional semantics, reflected by the contextual information associated with the concepts, which are sensitive to the time periods, locations, and even the cultural background of human coders. We propose to measure the changes of concept semantics to provide users a finer-grained access to the background knowledge represented in the form of a formal ontology.

3 What is concept shift?

In philosophy of language, the concepts *extension* and *intension* signify two different ways of defining the semantics of a concept. *Intensional* semantics define a concept by listing the defining properties of a concept (*i.e.* its intension), such as stating that a bird is an egg-laying vertebrate that can fly. For each object it can be determined whether it matches this definition by checking whether it has these properties. *Extensional* semantics, on the other hand, define a concept by listing the objects that match its definition (*i.e.* its extension) [7]. For example, the concept Dutch cities can be defined by listing Amsterdam, Rotterdam, The Hague, etc.

In the communication science study, the intension of a concept in the ontology stands for the core part of the concept that the communication scientists want to measure, while the extension is formed by all the instance contextual information assigned to the concept by the human coders. We state that concept shift occurs when *the extensional and intensional semantics of a concept differ, and the extensional semantics changes over time*.

4 How can we detect concept shift?

If the intensional semantics of a concept is relatively stable, then we can measure the extensional semantics using its extensions, that is, the instances associated with it. We propose to study concept shift by using extensional semantics to derive relations between concepts from the same years and different versions of the concepts used in different years. Combining with intensional information of the concepts, such as hierarchical information, we observe the subtle changes in meaning of the related concepts over time.

Extension mappings between different years Following our previous work [8] we use information retrieval techniques to calculate document similarity between annotated articles, which we use subsequently to identify similarity between concepts. As reported in [10], chains between the most similar of these concepts thus produce graph-like structures (lines, trees, or DAGs), which “tell their own” story of Dutch politics over the past 15 years.

Co-occurrence links within same years Since the extensional semantics of one concept are reflected by the instances which are associated with it, we can study the co-occurrence of this concept and other concepts. By comparing the overlap between instances, we can measure the strength of the association between concepts within the same year. In this way, if one concept is correlated with different concepts in different years, it is likely the extensional semantics of this concept may change between these years.

Hierarchical links within same years The intensional semantics of a concept can be defined in the ontology by its property and its structural association within the ontology. Unfortunately, the ontologies at hand have no intensional properties defining the concepts involved. We can only rely on little hierarchical information to obtain some indication of the intensional semantics of the concepts involved.

Once these three kinds of links are put together, we can generate a whole picture of the change of concept semantics over time.

5 Experiments

For our experiments we used 5 different ontologies from the Dutch election campaigns in 1994, 1998, 2002, 2003 and 2006. Our experiment were conducted by mapping each of those ontologies with each other. Each ontology was used manually to annotate (around 5000) newspaper articles of the respective campaign. There exists a manually created (intensional) alignment between these 5 ontologies, which gives us an opportunity to compare the difference between extensional and intensional semantics.

As reported in [10], the consistency between manual (intensional) and extensional mappings is not big, as the precision and recall in terms of the manual mappings are not satisfactory. However, it is not the goal to reproduce the manual mappings by using the extensional method, instead, we would like to study the differences between extensional and intensional mappings, which will say something about the concept shift in general.

Figure 1 gives two examples of such chains of concept shift. The three types of links introduced in the previous section together with manual mappings provided by the domain experts are combined and visualised in terms of different colours:

- blue links: extensional mappings between different years
- red links: manual (intensional) mappings between different years
- black links: co-occurrence links within the same years
- green links: hierarchical links within the same years, with the starting concept being the child concept.

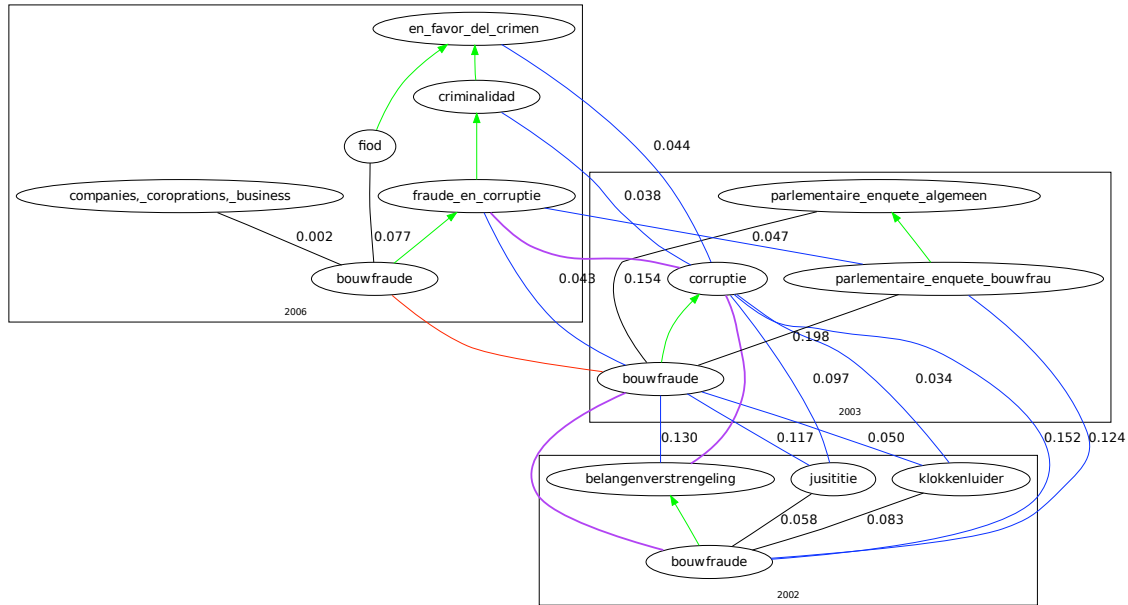
plus

- purple links: when extensional and intensional mappings agree with each other

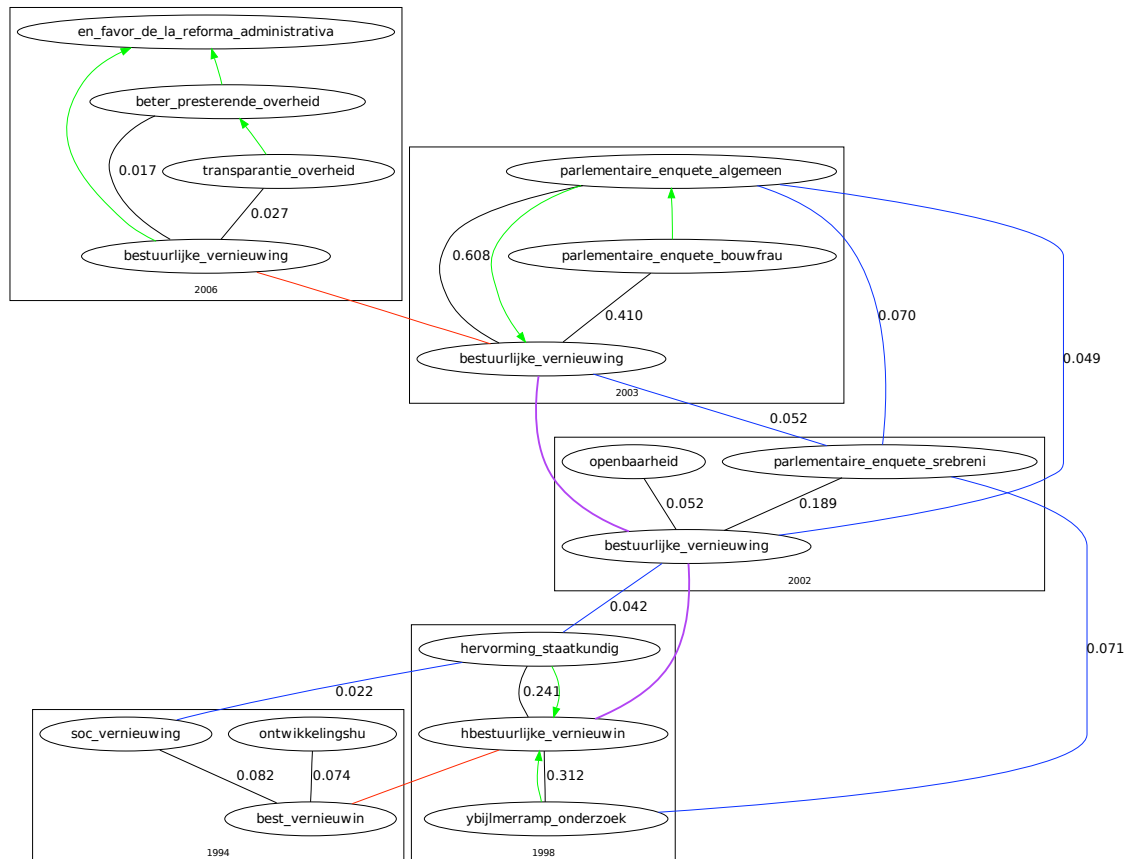
The number next to the links are the confidence value of such links. Manual mappings are considered to have a confidence value of 1, which is omitted in the figures.

What kinds of concept shift can be detected?

- association shifting
 - In 1994, *best vernieuwin* (administrative reforms) is strongly related to *ontwikkelingshu* (development aid) and *soc vernieuwing* (social reforms); while in 1998, it is related to *hervorming staatkundig* (state reforms) and *ybijmerramp onderzoek* (Inquiry in Bijlmer disaster). It suggests the concept “reforms” was about social reforms and development in 1994, but in 1998, when political parties mentioned “reforms”, they particular meant the reforms due to the disaster that happened in 1998. While again, in 2002, it is related to the right of expressing opinions about reforms (*inspraak*).



(a) Bouwfraude (builder's fraud)



(b) Bestuurlijke Vernieuwing (administrative reforms)

- generalisation/specialisation of concepts
 - `bouwfraude` (builder's fraud) in 2003 is linked to `fraude en corruptie` (fraud and corruption) in 2006, which is the parent concept of `bouwfraude` — **this concept shifts to have a more general semantics**
 - The two equivalent concepts in 2002, `bestuurslijke vernieuwing` and `bestuurlijke vernieuwing` (administrative reforms), are both mapped to `parlementaire enquête algemeen` (parliamentary inquiry; general) in 2003, which is a child concept of `bestuurlijke vernieuwing` — **this concept shifts to have a more specific semantics**
- overlapping concepts
 - `vno_ncw` in 1994 is mapped to `rwerkgevers` in 1998, which co-occurs extensively with `rvno`, the intensional mapping of `vno_ncw` of 1994. — **this concept shifts to have a different but overlapping semantics**

- [10] Shenghui Wang, Stefan Schlobach, Janet Takens¹, and Wouter van Atteveldt, 'Mapping-chains for studying concept shift in political ontologies', in *Proceedings of the Ontology Matching workshop (OM 2009)*, Washington, USA, (October 2009).

6 Open questions

We have shown in [10] that studying a representative sample of mapping chains through our domain expert indicates clearly that mapping chains can be interesting given the reasonable quality of the extensional mappings our method can produce.

However, we have a few open questions which need to be investigated further:

- What types of concept shift can we identify and automatically detect?
- What is an appropriate (formal) representation for the detected shift?
- How can we evaluate the detected concept shift, both qualitatively and quantitatively?

REFERENCES

- [1] Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, Eric Tanenbaum with Richard C. Fording, Derek J. Hearl, Hee Min Kim, Michael McDonald, and Silvia Mendez, *Mapping Policy Preferences. Estimates for Parties, Electors, and Governments 1945-1998*, Oxford University Press, Oxford, 2001.
- [2] Giorgos Flouris, Dimitris Manakanatas, Haridimos Kondylakis, Dimitris Plexousakis, and Grigoris Antoniou, 'Ontology change: classification and survey', *Knowledge Eng. Review*, **23**(2), 117–152, (2008).
- [3] B. D. Jones and F. R. Baumgartner, *The politics of attention: How government prioritizes problems*, University of Chicago Press, Chicago, 2005.
- [4] Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, Ian Budge, and Michael Macdonald, *Mapping Policy Preference II: Estimates for Parties, Electors and Governments in Eastern Europe, the European Union and the OECD, 1990-2003*, Oxford University Press, Oxford, 2006.
- [5] K Krippendorff, 'Reliability in content analysis', *Human Communication Research*, **30**, 411–433, (2004).
- [6] Matthew Lombard, Jennifer Snyder-Duch, and Cheryl Campanella Bracken, 'A call for standardization in content analysis reliability', *Human Communication Research*, **30**(3), 434–437, (2004).
- [7] A. S. Maida and S. C. Shapiro, 'Intensional concepts in propositional semantic networks', *Cognitive Science*, **6**(4), 291–330, (1982).
- [8] Balthasar Schopman, Shenghui Wang, and Stefan Schlobach, 'Deriving concept mappings through instance mappings', in *Proceedings of the 3rd Asian Semantic Web Conference*, Bangkok, Thailand, (2008).
- [9] Wouter Van Atteveldt, *Semantic Network Analysis: Techniques for Extracting, Representing, and Querying Media Content (dissertation)*, BookSurge, Charleston, SC, 2008.