

On the Semantics of Concept Drift: Towards Formal Definitions of Semantic Change

Antske Fokkens¹, Serge ter Braake², Isa Maks¹, and Davide Ceolin³

¹ Computational Linguistics, VU University Amsterdam, Netherlands
antske.fokkens@vu.nl, isa.maks@vu.nl

² Media Studies, University of Amsterdam, Netherlands sergeterbraake@gmail.com

³ Computer Science, Web and Media department, VU University Amsterdam,
Netherlands d.ceolin@vu.nl

Abstract. Semantic change and concept drift are studied in many different academic fields. Different domains have different understandings of what a concept and, thus, concept drift is making it harder for researchers to build upon work in other disciplines. In this paper, we aim to address this challenge and propose definitions for these phenomena which apply across fields. We provide formal definitions and illustrate how concept drift and related phenomena can be modeled in RDF through the use of context. We explain and support the definitions through an example from historical research and argue that a formal modeling of semantic change in RDF can help to better interpret data.

Keywords: Concept Drift · Semantic Change · Digital Humanities · RDF

1 Introduction

Semantic change and **concept drift** are important fields of study for many academic fields, including history, linguistics, philosophy, political science and computer science. However, not every discipline uses these terms in the same way. This is not a problem in itself, but does create confusion when these fields communicate or even work together, like in digital humanities projects. In this paper, we aim to tackle this challenge with particular focus on how the relation between concept drift in Semantic Web research and other disciplines.

We define **concept drift** as a change in the intension (the definitions and associations) of a concept, while the (rigid) core of the concept remains stable [11, 16]. If the core of the concept changes as well, and we have a new concept, we speak of **conceptual replacement**, following Kuukkanen [11, p. 367-370]. We argue that these definitions are applicable across domains. We outline how concept drift relates other forms of semantic change (which are sometimes considered concept drift as well). The main contributions of this paper are the following:

1. We propose definitions of concept drift and related phenomena that are applicable across domains.

2. We show how contexts can be used to model concept drift and related phenomena in RDF.
3. We illustrate the complex interaction of various forms of semantic change and argue formal modeling of these processes can improve research across domains.

The rest of this paper is structured as follows. Section 2 provides an overview of work on concept drift in various disciplines. We introduce our definitions and formal models in Section 3. In Section 4, we illustrate the complex interaction between concept drift and related phenomena through an example from history. We conclude with a discussion in Section 5.

2 Related Work

Historians have been studying how concepts and their associations change over time for decades, to understand historical events [13]. Concept drift also receives plenty of attention from digital humanities researchers.⁴ Computer scientists and historians from the Translantis project, for example, have used vector coordinates to see how vocabularies shift over time, which in turn can help to detect concept change. Rather than taking a word as a starting point, they use the meaning, the concept, as an anchor, and monitor the evolving set of words that are used to denote it [10].

In sociology, some make a distinction between normal concepts and contested and contestable concepts, like ‘democracy’ and ‘freedom’, which leave room for lots of debate [5]. Contested concepts are considered to be of great interest to study in order to see how the political climate changes in a certain period of time. Historical linguists study causes of shifts in word meanings in general [8]. They tend to focus on linguistic mechanisms of semantic change such as metonymy and metaphor rather than on meaning shifts caused by socio-cultural reasons. Historical linguists generally look at a much broader range of concepts than sociologists. Within the Semantic Web, researchers seem to focus on a wider range of changes for concepts, where some consider any change related to a concept a form of concept change, regardless of whether it is a change in its intension, its labeling or its extension [16].

Within the humanities and social sciences a clear definition of ‘concept’ is often lacking [11, p. 363]. Recently the philosophers Betti and Van den Berg have argued that to study the history of thoughts (and concepts), researchers should make use of more explicit models, frameworks, to avoid the scent of arbitrariness in what they are doing [1].

Cognitive science has shown that people link many associations with concepts, some being more important than others [11, p. 354]. For example, the concept of ‘pig’ has the associations of ‘mammal’, ‘four-legged’ and ‘farm animal’. While a three legged pig living in a castle still is a pig, few would say that a creature that looks like a pig but lays eggs actually is a pig. Any definition of a

⁴ http://www.helsinki.fi/collegium/events/conceptual_change/abstracts.pdf

concept has a sense of arbitrariness and is based on choices of what associations to include [14, p. 35-37]. It therefore is desirable to study concepts with as much flexibility as possible.

Kuukkanen makes a distinction between the core of a concept, determined by a historian, and its margins. By making this distinction, it is easier to develop a vocabulary to talk about semantic changes. When the core of a concept changes he speaks of *conceptual replacement* [11, p.367-370]. Wang et al. speak of the ‘rigid core’ of a concept in this respect, and when that rigid core changes, the concept is not the same concept anymore. What belongs to the rigid core should ideally be determined by human domain experts (‘oracles’) [16, p. 8]. We adapt Kuukkanen’s and Wang’s view that it is useful to distinguish between core and margin associations and that domain experts are the ones to decide which properties are part of the concept’s core. We define concept drift as changes in the concept’s associations while the core stays in tact. This means we limit our definition of concept drift to what Wang et al. call *changes in intension*. In the next section we will elaborate on this definition and other forms of semantic change that have been treated as concept drift in the Semantic Web literature.

3 Modeling Semantic Change in RDF

Wang et al.’s research forms a central piece in work on concept drift within the Semantic Web community. They provide definitions of three forms of concept drift as well as formulas to calculate changes. We first address their definitions and explain how they relate to the ones we propose. We then introduce formal ways of modeling concept drift, which is not covered by Wang et al. [16].

3.1 Semantic change and concept drift

Wang et al., state that concept drift can involve changes in the concept’s intension (its associations), in its extension (its reference) or in its label (words used to refer to it). We follow Frege’s [6] view that the sense of a concept lies in its intension and thus only intensional changes are concept drift. We see changes in extension and labels as phenomena of semantic change related to concept drift. Changes in extension ([1]) and changes in label can be relevant to study when investigating concept drift, but this is not necessarily the case. We will illustrate the relevance of extensional changes for three types of concepts.

The concept of ‘pig’ has an extension that changes with every pig that is born or slaughtered. Changes in this concept’s extension are not relevant for investigating semantic change or ontological representation.⁵ Other concepts have extensions that inherently change (without the sense of the concept changing), where changes are relevant for formal modeling. The concept of ‘government’ is an example: a change in government does not mean the concept of ‘government’

⁵ The approximate number of living pigs may be relevant for certain use cases, but it is difficult to find a study where the exact extension is important to know.

changed, but representing what government is in place is relevant ontological information. Finally, there are concepts where the extension is closely connected to the intension of the concept. This is the case for Wang et al.’s example of the European Union, which is partially defined by its member states.

Semantic changes that concern the relation between labels (words) and the concepts they refer to are also often related to concept drift. When a new or different label is used for a concept, this often reflects an intensional change. For instance, when someone chooses to use the word *pig* to refer to a ‘police officer’, this invokes connotations that reflect the way they look at the police. We also observe changes where the same label ends up referring to a different concept. A well-known example of such change is the English word *cute* (based on Frermann and Lapata (2016) [7], following the Oxford English Dictionary [15]). In the early 18th century, the word appeared meaning ‘clever’. In the late nineteenth century the meaning had shifted to ‘cunning’. It then shifted to its modern day meaning of ‘sweet/attractive’. We assume that, in this case, the concepts of ‘clever’, ‘cunning’ and ‘sweet/attractive’ have not changed. Such changes are forms of what Kuukkanen calls **conceptual replacement**. In Section 4 we will see that conceptual replacement often interacts with concept drift.

3.2 Modeling semantic change and concept drift

We propose to model the phenomena outlined above through the use of contexts and Lemon (lexicon model for ontologies) [12].⁶ Lemon is a W3C-standard that can describe complex lexical information about words and link them to ontologies in the Semantic Web. Lexical knowledge is linked to ontological knowledge through so-called lexicalSense nodes. Lexical knowledge may refer to spelling variants (e.g. American vs. British spelling) or morpho-syntactic information such as part-of-speech or plural forms. The lexicalSense that links this knowledge to the ontology can be used to define contextual and usage conditions, such as the connotations of ‘pig’ as a reference for police man or the time span in which *cute* referred to ‘clever’.⁷ As such, Lemon directly provides the means to model label changes for concepts as well as conceptual replacement.

Changes in the intension or extension of the concept, on the other hand, are not captured in the lexicalSense (since the label of the concept may remain stable). We follow Bouquet et al. [2] and place information dependent on a context in the same graph. Figure 1 illustrates how the concept ‘democracy’ has drifted. We define the core meaning as ‘voting power for people’. In addition, the concept can be associated with various properties that can be context dependent. In Figure 1, the first context (c1) associates democracy with secret ballots and a multiple party system and assumes there is a minimum age. This could represent how a (well-functioning) democracy is perceived in Europe in the 21st century. Context 2 (c2) not only assumes an age limitation, but also restricts voting rights

⁶ <http://lemon-model.net>

⁷ <https://github.com/ctrl/GRaSP/blob/master/examples/concept-drift/cuteLemon.png> illustrates *cute*’s conceptual replacement.

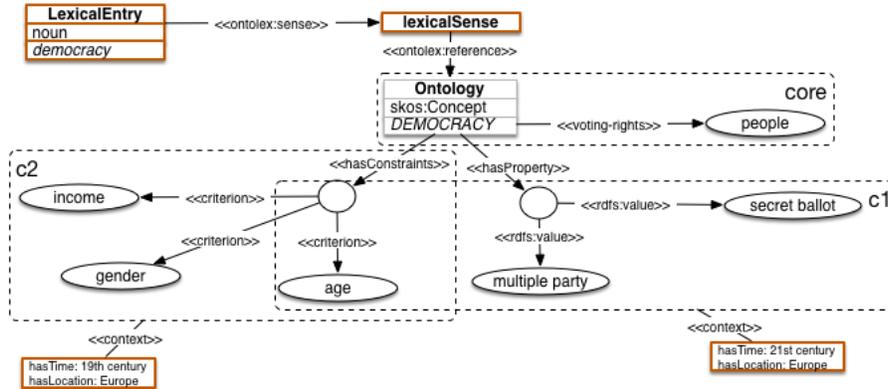


Fig. 1. Democracy in different contexts

to a gender and income, fitting the context of many European countries in the late 19th century. We can imagine other variations that capture how democracy was seen in (e.g.) the Soviet Union, Ancient Rome or Athens. Note that we can also use these contexts to model extensional changes (e.g. to represent which countries were part of the European Union in a specific time).

3.3 Using formal representations

Conceptual changes can be detected by comparing changing in an ontology [16]. As Horrocks points out, “in computer science an ontology is an engineering artefact, usually a model of (some aspect of) the world; it introduces vocabulary describing various aspects of the domain being modelled, and provides an explicit specification of the intended meaning of the vocabulary” [9]. Ontologies can be changed to improve their ability to model the world. However, we must distinguish between the case when this improvement takes place because previous versions of an ontology were inadequate, imprecise, erroneous (ontology versioning), and when it happens because a change that occurred in the world needs to be tracked. By making the context in which a certain definition applies explicit, we can distinguish changes in the world from changes that are extensions or corrections of the ontology. This furthermore allows us to use the ontology for interpreting data from or about the past. Concept changes in history start long before the Semantic Web existed and interpretations from historical contexts will therefore need to be added directly as representations of the past. We will describe what this might look like in DBpedia and GeoNames.

DBpedia contains definitions of several forms of democracy, including Soviet Democracy, Direct Democracy, Representative Democracy and Pseudo-Democracy. They are defined as `dct:subject` of `dbpedia:Democracy`. They can be linked to specific societies through the `dct:subject` predicate as well (e.g. Soviet Democracy is linked to the Soviet Union and Direct Democracy to Ancient Greek society). DBpedia does however not make explicit that the concept of ‘Direct

Democracy’ is a form of democracy that was practiced in Athens during a specific time and that thus ‘Democracy’ in the context of this society refers to a direct democracy. In our model these relations could be made explicit by creating a `skos:sameAs` link between ‘Democracy’ and ‘Direct Democracy’ for the context of Athens in the Ancient Greek society. A scholar interpreting Plato may use this information to deduct that the negative view of democracy expressed in Plato’s view is not necessarily aimed at modern day democracy.

GeoNames provides information about locations and geo-political entities. One of GeoNames’ challenges is that such entities constantly change: countries and cities appear and vanish and some areas have a controversial political status. GeoNames indicates that countries that no longer exist such as Czechoslovakia are ‘historical political entities’, but it currently does not offer any information about the country’s context. Though this may be easy to add for a country from the past that was relatively stable throughout its existence, this becomes more challenging for countries such as Poland, which are current countries, but have changed drastically throughout history (from being one of the largest empires in Europe to completely disappearing as a political entity twice to moving West after World War II). If GeoNames wants to accurately represent this information, time-dependent contexts are essential.

The examples above are still relatively straight-forward cases of concept drift and semantic change. In the next section, we will illustrate that the interaction between various phenomena can become quite complex and that for accurate historical interpretation, a wide range of concepts needs to be taken into account.

4 Semantic changes in history

In 1541, Catharina de Chasseur ([3, for her biography]) was arrested on the suspicion of forgery and distribution of false coins. Many historians have described and interpreted this episode in different ways over the past two centuries. One contemporary had testified that Catharina, the *huysfrou* of Lord Van Assendelft, was *quaet ende boos*. To modern Dutch these two expressions would literally translate to ‘housekeeper’ and ‘angry and angry’. In the sixteenth century, however, these words translated to ‘wife’, ‘evil’ and ‘bad’. Trained historians are aware of such semantic changes and will use the necessary dictionaries to interpret their source material well.

Both *boos* and *quaet* refer to a different concept in modern Dutch (though they may occur with their old meaning in some fixed expressions). The changes related to *huysfrou* are slightly more complex. Conceptual replacement has taken place, because the modern word *huisvrouw* no longer refers to spouse. However, it still typically refers to a woman taking care of the household for her own family. At the same time, its original meaning ‘wife’ has had a stable core meaning (a person of the female gender who is legally married to another person), but outside its core the intension heavily depends on context (time, place and culture). Even though a wife still is a female person married to another person, this other person could only be a person of the male gender in the sixteenth century.

The sixteenth century ‘wife’ would furthermore be more associated with ‘family ties’, ‘status’ and ‘bearer of children’ than with ‘romantic love’ and ‘equal’ or ‘other/better half’. This concept drift, the concept of ‘wife’ with a different intension, is imperative to identify for the historian, because they can lead to different interpretations of the past.

Catharina de Chasseur was the wife of the Lord van Assendelft. The Van Assendelfts opposed this marriage because Catharina was not of nobility. To them the concept of ‘wife’ did not fit her for their son, because she was socially mismatched. While modern observers could judge the Van Assendelft family’s behavior as cruel and uncaring, they behaved mostly in line with the customs of their time. As such, the correct interpretation of sources is even more dependent on concept drift than on words changing meaning (conceptual replacement).

The example of *huysfrou* and ‘wife’ illustrates the sometimes complex interaction between a shift in label and concept drift. The current word *huisvrouw* may have lost the core meaning of ‘spouse’, but it still shares many associations with its original lexicalSense.⁸ It may be an interesting topic of investigation how the concept drift of ‘wife’ relates to the lexical semantic change of *huysfrou*: did the drifting concept induce the lexical semantic change, or the other way around? The RDF representation proposed in this paper allows researchers to make both forms of change explicit supporting this kind of investigation.

5 Conclusion and Discussion

The main goal of this paper was to show how various disciplines look at concept drift in order to increase understanding and hence facilitate collaboration between disciplines. We argued that out of the three forms of concept drift addressed by Wang et al. [16], only intensional change is (likely to be) understood as concept drift across domains. Extensional change and label change are related phenomena that can be indicative of concept drift, but this is not always the case. The relevance of extensional change in particular depends on the kind of concept under consideration.

We showed how various forms of semantic change can be modeled using lemon to link lexical entries to concepts in ontologies and using contexts to define how a specific concept was seen or what its extension was in a specific time and place. Making this information explicit can allow researchers to better interpret data [4] and distinguish ontological changes due to changes in the concepts from those that are corrections or extensions of the ontology. We described how the model might be used in DBpedia and GeoNames.

Obtaining this information, however, is challenging. Recent advances in distributional semantics have been shown to be highly effective in picking up semantic change from text [7], but they have mainly been evaluated on known cases of conceptual replacement. Capturing concept drift requires domain expertise and the examples above show that for a field such as history, relevant changes

⁸ An illustration of the concept and lexicalSense of *huysfrou* and ‘wife’ can be found at: <https://github.com/ctrl/GRaSP/blob/master/examples/concept-drift/spouce.gif>

may apply to concepts for everyday terms such as ‘wife’ to highly philosophical terms such as ‘freedom’. On the other hand, concept drift is intensively studied and there is a vast amount of knowledge about semantic change. If we can show that Semantic Web technology can help them, experts may be swayed to help the Semantic Web and represent their knowledge in RDF, leading to models showing how to interpret terms and concepts in various contexts over time.

6 Acknowledgments

This work was supported by the Amsterdam Academic Alliance Data Science (AAA-DS) Program Award to the UvA and VU Universities and NWO VENI grant 275-89-029. We would like to thank anonymous reviewers and the workshop’s organizers for their feedback. All remaining errors are our own.

References

1. Betti, A., van den Berg, H.: Modelling the History of Ideas. *British Journal for the History of Philosophy* 4(22), 812–835 (2014)
2. Bouquet, P., Serafini, L., Stoermer, H.: Introducing context into rdf knowledge bases. In: SWAP. vol. 5, pp. 14–16 (2005)
3. ter Braake, S.: Chasseur, Catharina de. In: *Digitaal Vrouwenlexicon van Nederland* (2006)
4. Ceolin, D., Noordegraaf, J., Aroyo, L.: Capturing the ineffable:collecting, analysing and automating web document quality assessments. In: *EKAW* (2016), to appear
5. Collier, D., Hidalgo, F.D., Maciuceanu, A.O.: Essentially Contested Concepts. *Journal of Political Ideologies* 11(3), 211–226 (2006)
6. Frege, G.: Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophischen Kritik* (100), 25–50 (1892)
7. Frermann, L., Lapata, M.: A bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics* 4, 31–45 (2016)
8. Geeraerts, D.: *Theories of lexical semantics*. Oxford University Press (2010)
9. Horrocks, I.: Ontologies and the semantic web. *Commun. ACM* 51(12), 58–67 (Dec 2008), <http://doi.acm.org/10.1145/1409360.1409377>
10. Kenter, T., Wevers, M., Huijnen, P., De Rijke, M.: Ad Hoc Monitoring of Vocabulary Shifts over Time. *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management* pp. 1191–2000 (2015)
11. Kuukkanen, J.M.: Making sense of conceptual change. *History and Theory* October(47), 351–372 (2008)
12. McCrae, J., Aguado-De-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., Wunner, T.: Interchanging lexical resources on the semantic web. *Lang. Resour. Eval.* 46(4), 701–719 (Dec 2012), <http://dx.doi.org/10.1007/s10579-012-9182-3>
13. Richter, M.: *The History of Political and Social Concepts. A critical introduction*. Oxford University Press, New York and Oxford (1995)
14. Saeed, J.I.: *Semantics*. Blackwell Publishers (1997)
15. Stevenson, A.: *Oxford dictionary of English*. Oxford University Press, USA (2010)
16. Wang, S., Schlobach, S., Klein, M.: Concept drift and how to identify it. *Web Semantics: Science, Services and Agents on the World Wide Web* 9(3), 247–265 (2011)