Using *lemon* to Model Lexical Semantic Shift in Diachronic Lexical Resources

Fahad Khan, Federico Boschetti, Francesca Frontini
CNR-ILC
Pisa, Italy
{name.surname}@ilc.cnr.it

Abstract

In this paper we propose a model, called *lemonDIA*, for representing lexical semantic change using the *lemon* framework and based on the ontological notion of the perdurant. Namely we extend the notion of sense in *lemon* by adding a temporal dimension and then define a class of perdurant entities that represents a shift in meaning of a word and which contains different related senses. We start by discussing the general problem of semantic shift and the utility of being able to easily access and represent such information in diachronic lexical resources. We then describe our model and illustrate it with examples.

Keywords: lemon, linked data, OWL, ontologies, perdurants, semantic shift

1. Introduction

In this paper we propose a model for representing lexical semantic change using the *lemon* framework. More precisely, we want to be able to track the shifts in meaning undergone by the lexical entries contained in some lexical resource (or resources), and to represent and access information about these meaning shifts in an intuitive way. We will limit our focus in this paper to changes in the meanings of lexemes (although in our examples we will focus on single words only) rather than trying to deal with so called grammatical semantic change – although this is a very closely related topic, see (Hollmann, 2009) for an overview.

The lexical resources that we particularly have in mind in this paper are those that contain etymological and/or other diachronically relevant information, as well as lexica for historical languages like Latin or ancient Greek in which the different stages of the language's development have to be taken into consideration. On the other hand the ideas we discuss are also applicable to any kind of general purpose lexicon and especially for wordnets.

We will work with the *lemon* model for lexical resources using the "semantics by reference" principles defined in (Cimiano et al., 2013). We will assume, given a lexicon L, that we have access to an ontology O which provides the semantics for L. Each lexical entry l in L (or at least the lexical entries we are interested in) will be associated with one or more ontology vocabulary items c in O that serve as extensions for l. In addition in this work we will assume that there is a time interval t associated with each correspondence between an entry and a vocabulary item.

We will employ the notion of *perdurant* commonly used in ontology modelling for representing change over time, to represent the shift in meaning of a lexical entry l from an original meaning c_0. For us this process of meaning shift becomes a perdurant entity to which we can explicitly refer. A perdurant here can be thought of as an event or a process that may be composed of different temporal as well as spatial parts.

We have called our new model *lemonDIA*.

In the next section, Section 2 we discuss the general problem of semantic shift with a particular emphasis on historical languages. Then in Section 4 we present our proposed model and give examples to illustrate its use, before finally discussing ideas for further work in the conclusion.

2. Lexical Semantic Change

The meaning of words in a language can often shift quite drastically over time, sometimes even over relatively short time scales. For example until a few decades ago the adjective *gay* was used to refer to someone as 'happy' or 'carefree', whereas this meaning is no longer in common currency and the word is now primarily used to refer to someone as a homosexual. To take another example, the word *fantastic* was once used to mean that something was a product of the imagination, but now refers to something as being very positive.

Theoretical linguistic research in this area has identified a number of different typologies of semantic change, e.g., there are semantic shifts where the meaning of a word becomes more general, and others where the meaning becomes more specific. The thesis that there exists a general pattern whereby semantic changes tend to lead to words becoming less *objective* and more *subjective* (with these words being used in a special technical sense closely related to their everyday meaning), the so called process of subjectification, has also been proposed and found widespread acceptance, again see (Hollmann, 2009).

Moreover in the case of modern lexica for ancient Greek or Latin there is a clear need for tools to assist philologists and historical linguists in the study and representation of lexical semantic shift.

For example it was quite common in the ancient world, after a major change in a predominant epistemic paradigm (e.g., from pre-socratic to post-socratic) or in a governing religious or socio-cultural framework (e.g. from pagan to Christian), that terms in numerous important domains would be affected by semantic change – even if in many cases a prior general purpose meaning and a new more domain specific meaning were able to coexist for a certain period of time.

The Latin word *otium* (leisure, rest time) offers an excellent example of such a semantic shift, which in this case

occurred over several different, clearly defined, stages. The original meaning as attested in archaic texts can be understood as "the state prevailing during the absence of war", and referred to the state of leisure enjoyed by soldiers in an army camp, especially in winter.
During the classical age, the word assumed a very positive connotation and related to the absence of political duties: "time that is free from duties or responsibilities", "time available for hobbies and other activities that you enjoy" (which especially meant in this case the study of philosophy). Later in the Middle Ages, *otium* gained a double meaning. The first was positive: in the case when this "freedom from activity (work or strain or responsibility)" was devoted to God. The second was negative: when it meant "leisure time away from work devoted to rest or pleasure", and thus corresponded to the deadly sin of sloth. This latter meaning was to prevail during the medieval ages. Finally, Renaissance era Latin restored the classical meaning of *otium* according to which it meant freedom from social and political duties with the extra time instead being devoted to philosophical studies.
All of the meanings of *otium* quoted above are represented in Latin WordNet but as, we hope, the above demonstrates, there is a real need for a tool that can assist in the discovery and representation of this kind of semantic-conceptual evolution over different time periods.
Along with Latin there are currently wordnets in development for ancient Greek (Bizzoni et al., 2014) as well as for several other languages - such as Sanskrit (Kulkarni et al., 2010) - with long and well documented histories of use and for which the representation of semantic shift would be particularly useful for different groups of researchers such as historians and philologists.

3. The *lemon* model

As more and more lexical resources are added to the linguistic linked open data cloud, it becomes increasingly important to develop tools and methodologies to exploit the data contained within them. *lemon* (Mccrae et al., 2012) is currently one of the most popular models for publishing lexical resources as linked open data in RDF/OWL and so we decided to work with it as a foundation. Along with its popularity *lemon* also has a clearly defined semantics which made its use in this work even more attractive.
In the *lemon* model we can represent the relation between a lexicon L and a ontology $O = (\Lambda_0, V_0)$ whose vocabulary items provide the semantics for the lexical entries in L using sense objects as follows. Each lexical entry l is related to a vocabulary item $c \in V_0$ via a (reified) sense object $s = \sigma^{(l,c)}$ if there exists evidence for a use of l in which it can be taken to mean c. We represent this state of affairs using the `sense` and `reference` relations as defined in *lemon*: `sense`(l, s), `reference`(s, c).
lemon does make provision for adding temporal information to lexica by defining a property `usedSince` of *lemon* sense objects. `usedSince` is a subproperty of the *lemon* context property and allows the addition of information specifying when a term was first used in a particular sense[1].

[1]See the lemon cookbook for further details at `http://lemon-model.net/lemon-cookbook/`.

The work in the rest of this paper however explores a more extensive modelling of word sense shifts using *lemon*.

4. Using Perdurants to Model Word Senses

Let us assume that the word *punk* is an entry in our lexicon, L. Here it is taken as both a noun that from the 1970s onwards came to refer to a follower of a certain youth culture, and also as a noun that from around the 1590s until the 1700s meant a prostitute[2]. We want to be able to represent both of these meanings along with their relevant periods of use.
We will take these two meanings to correspond to two different concepts c, c', respectively, in an ontology O. Under the *lemon* semantics by reference framework we define a sense $s = \sigma^{(punk,c)}$ that represents the meaning of *punk* as c, and another sense $s' = \sigma^{(punk,c')}$ representing the meaning of *punk* as c'. In addition let t represent a time interval [1976 - Current] and t' represent the 'fuzzy' interval [1590 - ?1750] (we will mention issues relating to the representation of intervals whether qualitatively or quantitatively, i.e., without fixed endpoints, in the next subsection).
The time intervals we are working with here could represent any number of things such as the first and last attested uses of that particular meaning of a word or they could represent an educated guess as to the relevant time period in which that meaning was current. So then we would like to be able to state something like the following: $\texttt{sense}(punk, s, t)$, $\texttt{reference}(s, c)$ and $\texttt{sense}(punk, s', t'), \texttt{reference}(s', c')$[3].
In other words we want to make the sense relation a fluent. The question then arises, how can we model this and keep within what is broadly speaking still the *lemon* framework? An obvious solution and the one which we will pursue in the rest of this paper is to model each sense s as a perdurant, namely as an object with a time interval associated with it[4]. Then the correspondence between a lexical entry l and vocabulary entry c represented by a given *lemon* sense object has a certain time dimension representing, say, the validity of this correspondence or the time during which it was commonly in use.
Clearly adding this temporal dimension is helpful because it enables us to plot the different meanings of a word over a given time period and also to see if and when these meanings overlap. It would also be very helpful to be able to track how a specific meaning changes or evolves over a certain time period and in this case it makes sense to talk about the sense of a word, when viewed as an event or a process, as something that has different temporal parts, some of which may have different lexical references (although to

[2]For the purposes of the example we do not assume that these two meanings are related by a process of semantic shift, although this may well be the case.

[3]We could also of course add an extra argument for the reference relation instead, and model the relation between a sense and reference as varying with time; this way of modelling change over time could be handled in a similar way to the methodology we propose below.

[4]As we briefly discuss later, redefining the sense relation as a 3-place relation in OWL brings a host of problems with it.

avoid confusion with *lemon* sense objects we will refer to a sense viewed diachronically as a diachronic shift object).
For example the word *girl* originally referred to a young person of either male or female gender before shifting meaning so that it ended up referring only to young females, see (Hollmann, 2009).
So imagine that in our lexicon we have a entry $girl$ which during the interval $t1$ means "young human being" and that this class of entities is represented in our ontology by the concept c_1, and that during another, later, time interval t_2 it means "young female human being", and that this class is represented by the concept c_2 in our ontology. We want to be able to relate the senses $s_1 = \sigma^{(girl,c_1)}$ and $s_2 = \sigma^{(girl,c_2)}$ together as parts of another event 'object' representing the historical shift in meaning that took place from the word *girl* having the meaning of c_1 to its having the meaning c_2 along with other further shifts that might have also have taken place over time. In the next section we discuss how to do this using perdurants.

4.1. Perdurants, Endurants and Time Slices

When modelling fluents, i.e., relations and functions that change over time, common use is made in ontology design of the notion of a perdurant. By perdurants here we mean entities that have a temporal dimension and which we can indeed view as being composed of different temporal parts. Perdurants may have different properties which hold at some times during the time span associated with them but which do not hold at others; significantly enough, though, there will also be other, essential, properties which hold throughout the whole life span of the perdurant and by which they can be identified.
The notion of perdurant is often contrasted with the notion of an endurant by which we mean an entity that also has an associated life span but which is in some sense wholly present at each time point of its life span; so that unlike with perdurants we do not view endurants as being actually composed of different temporal segments. Another core idea which is related to that of perdurant is that of the *time slice*, which is a snapshot of a perdurant representing all (or perhaps a relevant subset) of the properties which hold of a perdurant at a given point in time.
Perdurants are a particularly popular method for representing fluents in OWL since they avoid the main pitfalls associated with the strategy of representing fluents using n-ary relations in OWL. In an influential paper (Welty and Fikes, 2006) Welty and Fikes describe these pitfalls in detail as well as laying out an approach in which all entities are represented as perdurants – although we do not pursue this approach in the current paper.

4.2. Description of the *lemonDIA* Model

Now we will give a description of *lemonDIA* our new model based on *lemon*. We define a new subclass of sense objects, so called lexical p-sense objects, which are defined similarly to normal *lemon* sense objects except that we define them as perdurants with a temporal dimension. These p-sense objects can be understood as follows.
Given a lexical entry l, an ontology vocabulary item c, and a time interval t, we propose the following criteria for determining the existence of a lexical p-sense object $s = \sigma^{(l,c,t)}$:

- We have evidence that the lexical item l was interpreted as c during the time period t.
- There exists evidence of a number of uses of the entry l being used to mean c during the time period t; and the set of these uses is represented by s.
- We are able to hypothesize the existence of a concept s giving the full lexical meaning of l when it was used to mean c during t.

Here the time intervals should be understood as being maximal in the sense that they represent the whole time interval in which l is interpreted as c.
So to return to the *punk* example we have the following $\texttt{pSense}(punk, s)$, $\texttt{reference}(s, c)$ and $\texttt{pSense}(punk, s')$, $\texttt{reference}(s', c')$ where the senses s, s' are now perdurants. We then add the following statements $\texttt{temporalExtent}(s, t)$ and $\texttt{temporalExtent}(s', t')$ where $\texttt{temporalExtent}$ is a relation between a perdurant and its associated time span given by a time interval.
How can we model the *girl* example given above, this time explicitly modelling the shift in meaning that took place between the two senses of *girl*? This will involve the definition of a class of diachronic shift objects.
Meanings or senses, like ideas, are not three- or even four-dimensional objects: that is although a meaning can be manifested in a physical format, the meaning itself has no spatial dimensions, similarly it can be argued that meanings are in some sense timeless. However as we mentioned above meanings can be associated with temporal intervals (and also with spatial dimensions if we think in terms of the geographical regions in which the language communities occur for which these meanings are common although we do not discuss this here). We will use these temporal intervals along with the 'dimension' of meaning provided by an ontology in order to view meanings as perdurants. We can motivate this solution in a (hopefully) intuitive way as follows.
Say there exists an initial correspondence between a lexical item l with a ontology concept c_0 at time t_0 and imagine we have a graph where the x-axis represents time (the time line is represented as a real line) and the y-axis is an enumeration of the concepts in our ontology O. We can visualize meaning shift by plotting the vocabulary items $c^i_1, ...c^i_k$ at each succeeding time point t_i on the basis of whether l means c^i_j at time t_i and where these meanings are related by the process of meaning shift to the original meaning at time t_0. Let $C_{(l,c_0,t_0)}$ be the set of all such c^i_j.
It is important to emphasise that the meanings in $C_{(l,c_0,t_0)}$ all derive from the original pairing of l and c_0 by a process of historical meaning shift, but that the lexical entry l may mean other unrelated things at each of these time points, if for example the word is homonymous and the other senses are not etymologically related to the meaning of l as c. (Also we will assume in this paper that if a lexical entry l means two different but related things, that the distinction is not made between these two things in our ontology

but that there is one 'closest' single vocabulary item c capturing the two meanings, we are allowed to assume these two senses are one – although this might be problematic in practice.)

In Fig 1 we give a chart representing the situation where a lexical entry has meaning c at time points $t0$ to time $t4$ and meaning c' at time $t4$ and $t5$. This chart and the succeeding one in Fig 2 are based on similar charts in (Bittner and Donnelly, 2004).

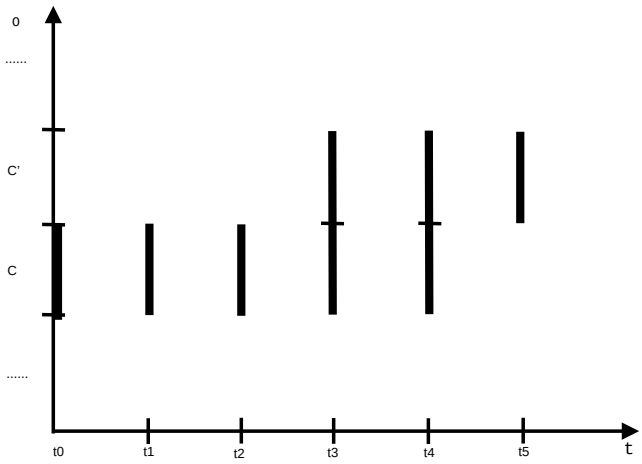

Figure 1: The meaning of a lexical entry l over time.

Taken together, at each time point t_i, the meanings c^i_j can be thought of as constituting a time slice of a perdurant object, d such that d represents the process of meaning shift of the lexical entry l with original meaning c_0.

In other words, given an ontology item $c \in C_{(l,c_0,t_0)}$ each p-sense $s = \sigma^{(l,c,t_c)}$, where t_c is the appropriate time interval, is related to d via a (perdurant) `partOf` relation (since a perdurant can have both time slices and other perdurants as parts, see (Bittner and Donnelly, 2004)); on the other hand we can think of d as the 'sum' of all the p-senses $s = \sigma^{(l,c,t_c)}$ where $c \in C_{(l,c_0,t_0)}$. We will refer to d as a diachronic shift object, for want of a better word, and use it to represent the meaning shift that words undergo over time. We will define a relation `diachronicShift` that holds between l and d.

In Fig 2 we represent the diachronic shift object for the example in the previous figure.

To return to the *girl* example, given our previous definitions we have that $\texttt{pSense}(girl, s_1)$ with $\texttt{reference}(s_1, c_1)$, $\texttt{temporalExtent}(s_1, t_1)$, and $\texttt{pSense}(girl, s_2)$ with $\texttt{reference}(s_2, c_2)$, $\texttt{temporalExtent}(s_2, t_2)$. Accordingly we can define a diachronic sense object d with $\texttt{diachronicShift}(girl, d)$ such that there exists a `partOf` relation between d and s_1 and d and s_2. In Figure 3 we represent the *girl* example with a diagram.

Note again that the diachronic shift object d captures the meaning shift of a lexical item from an initial meaning of c_0 by encompassing meanings that are related historically. The word *page*, for example, has two historically unrelated senses which we wouldn't want to include in a single diachronical sense object, at least not from an etymological point of view.

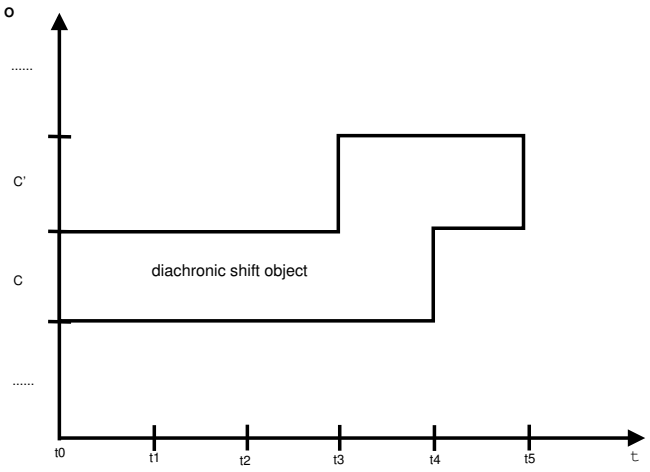

Figure 2: The meaning of a lexical entry l over time represented as a perdurant.

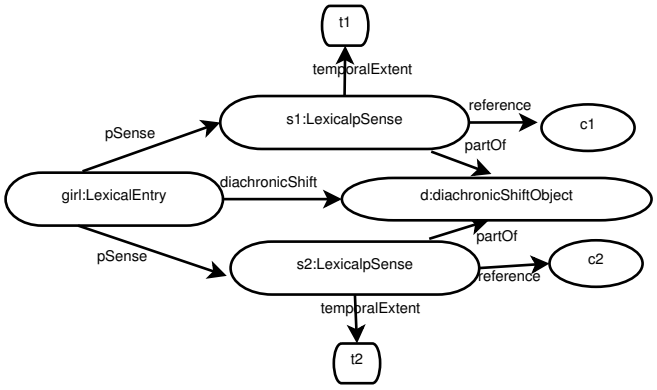

Figure 3: The *girl* example.

In Figure 4 we present a diagram of the *lemonDIA* model.

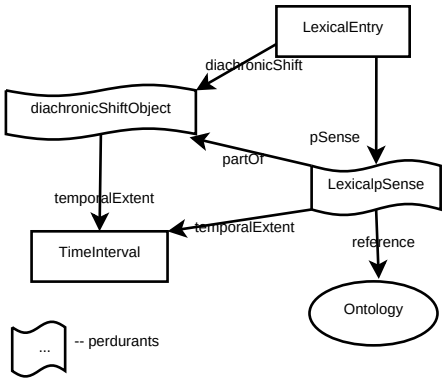

Figure 4: The *lemonDIA* model.

5. Conclusions

We have outlined a model, called *lemonDIA*, for representing lexical semantic change using the *lemon* framework and the ontological notion of the perdurant. The description of the model given above needs to be considerably fleshed out.

One of the most important issues relates to how to represent the time intervals associated with the periods corresponding to a lexical item's having a given meaning. It is often the case that due to a lack of evidence we cannot give an exact year, decade or even century specifying when a certain meaning first started to be used, nor for when it stopped being commonly used. Fortunately, there has been work done recently on representing so called qualitative time intervals, namely intervals which do not have specified start or end dates, defining relationships between them using Allen's temporal relations (e.g., Before, After, Meet), and on building reasoning tools using SWRL and special querying tools (Batsakis and Petrakis, 2011). This kind of work seems to be an important starting point for the further development of *lemonDIA*.

It would also be useful to add further properties that specify for a given time period which of the senses of a word are used predominantly, which of them are rarely used, though not yet obsolete, and which senses are at least still understood if not used. In addition it is important to be able to specify information relating to context or literary genre, especially when it comes to working with resources such as ancient Greek or Latin wordnets, where certain words may be obsolete or rarely used in one literary genre or in every day speech but still common in another.

At this stage of the development of the *lemonDIA* model these issues need to be explored in much greater depth. The most important thing of course is to see how the model works in practice, namely, when it is used on an actual lexical resource, something we have still to undertake.

6. Acknowledgements

The research in this paper was undertaken within Memorata Poetis, an Italian national project and part of the Programma di Ricerca di Rilevante Interesse Nazionale program for 2010/2011.

7. References

Batsakis, S. and Petrakis, E. (2011). Representing temporal knowledge in the semantic web: The extended 4d fluents approach. In Hatzilygeroudis, I. and Prentzas, J., editors, *Combinations of Intelligent Methods and Applications*, volume 8 of *Smart Innovation, Systems and Technologies*, pages 55–69. Springer Berlin Heidelberg.

Bittner, T. and Donnelly, M. (2004). The mereology of stages and persistent entities. In de Mántaras, R. L. and Saitta, L., editors, *ECAI*, pages 283–287. IOS Press.

Bizzoni, Y., Boschetti, F., Del Gratta, R., Diakoff, H., Monachini, M., and Crane, G. (2014). The making of Ancient Greek WordNet. In Calzolari, N., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014), Reykjavik, Iceland, May 28-30, 2014*, Reykjavik, Iceland. European Language Resources Association (ELRA).

Cimiano, P., McCrae, J., Buitelaar, P., and Montiel-Ponsoda, E. (2013). On the role of senses in the ontology-lexicon. In *New Trends of Research in Ontologies and Lexical Resources*, pages 43–62. Springer.

Hollmann, W. B., (2009). *English language: description, variation and context*, chapter Semantic change. Basingstoke: Palgrave.

Kulkarni, M., Dangarikar, C., Kulkarni, I., Nanda, A., and Bhattacharya, P. (2010). Introducing sanskrit wordnet. In *The 5th International Conference of the Global WordNet Association (GWC-2010)*.

Mccrae, J., Aguado-De-Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., and Wunner, T. (2012). Interchanging lexical resources on the semantic web. *Lang. Resour. Eval.*, 46(4):701–719, December.

Welty, C. and Fikes, R. (2006). A reusable ontology for fluents in owl. In *Proceedings of the 2006 Conference on Formal Ontology in Information Systems: Proceedings of the Fourth International Conference (FOIS 2006)*, pages 226–236, Amsterdam, The Netherlands, The Netherlands. IOS Press.