

OntoLex and Onomasiological Ordering: Supporting Topical Thesauri

Sander Stolk

Leiden University, Leiden, The Netherlands,
s.s.stolk@umail.leidenuniv.nl

Abstract. The OntoLex vocabulary has been designed to capture lexicons and to add their lexicographical knowledge to ontologies in the Semantic Web. Although the specification of the vocabulary posits that OntoLex allows lexicons to be ordered onomasiologically, it does so for a very specific kind of onomasiological ordering only. As a consequence, the vocabulary is currently insufficient for capturing a large proportion of the existing topical thesauri. This paper demonstrates the current expressivity and this shortcoming of OntoLex through two case studies: *The Historical Thesaurus of the Oxford English Dictionary* and *The Scots Thesaurus*. In order for OntoLex to offer full support for topical thesauri and their ordering principles, this paper proposes the addition of a single property to the vocabulary: `ontolex:isSenseIn`. . . .

Keywords: OntoLex, Lemon, onomasiological ordering, thesaurus

1 Introduction

The Lexicon Model for Ontologies vocabulary has been designed to capture lexicons and to add their lexicographical knowledge to ontologies in the Semantic Web [1]. The vocabulary has seen a number of updates, and was published as a W3C vocabulary by the OntoLex community group in May 2016 [2]. This version, henceforth OntoLex, has since been picked up by a number of bodies, including the Global WordNet Association, to represent and link existing lexical resources on the Semantic Web [3].

The specification of OntoLex puts forward a manner in which “lexicons can be ordered onomasiologically, that is by meanings rather than by lemmas” [2]. For publishers of topical thesauri, this is good news indeed. Such support is essential for these lexicographical works, which order their words by meaning instead of from a to z as is common in typical dictionaries. Yet the OntoLex vocabulary supports a very specific kind of onomasiological ordering only. As a consequence, the vocabulary is currently insufficient for capturing the knowledge from a large proportion of the existing topical thesauri. The current paper demonstrates this shortcoming of OntoLex and proposes a way forward for the vocabulary.

2 Methodology

In order to provide insight into the current support of OntoLex for the onomasiological ordering of topical thesauri, this paper will present two case studies. The first is based on the *Historical Thesaurus of the Oxford English Dictionary* [4]; the second on *The Scots Thesaurus* [5]. Both lexicographical works employ an onomasiological ordering for their lexicon. The first-mentioned thesaurus is considered to be a distinctive one and contains sets of synonyms. The second is not distinctive but cumulative and refrains from indicating synonymy [6].

This paper expresses samples from both thesauri in the OntoLex vocabulary. The manner in which OntoLex is applied is in line with the specification of the vocabulary [2] and the approach outlined by the Global WordNet Association [3]. This approach has been adopted by several projects, amongst which the Open Dutch Wordnet [7]. Namespaces relevant for this paper are provided in Listing 1. The RDF snippets in subsequent listings are specified in the Turtle RDF syntax [8]. Sample data from the case studies correspond with resources between angular brackets in the RDF snippets (that is to say, their namespace is left unspecified for the present purpose).

Listing 1. Namespaces

```
@prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
@prefix wn: <http://wordnet-rdf.princeton.edu/ontology#> .
```

3 Case study *Historical Thesaurus of the OED*

The first case study presented here is that of the *Historical Thesaurus of the Oxford English Dictionary* (HTOED). HTOED captures the English lexis that has existed throughout its 1300-year history, from Old English up to Modern English. This topical thesaurus groups together lexical items that are considered near-synonymous and provides insight into their use in time and place. HTOED was first published in print in 2009 [4] and in the following year also electronically [9].

Figure 1 depicts a sample from HTOED. This sample contains six categories from the topical system of the thesaurus (here represented by circles), which are organized in a hierarchy. A category that is displayed lower than another category to which it is connected by means of a line is subordinate to that connected category. On the right, a number of lexical senses are displayed (some of which are obsolete, conveyed by a dagger sign). These senses are considered synonyms, or rather, near-synonyms, in HTOED and are members of the “Freedom/liberty” category.

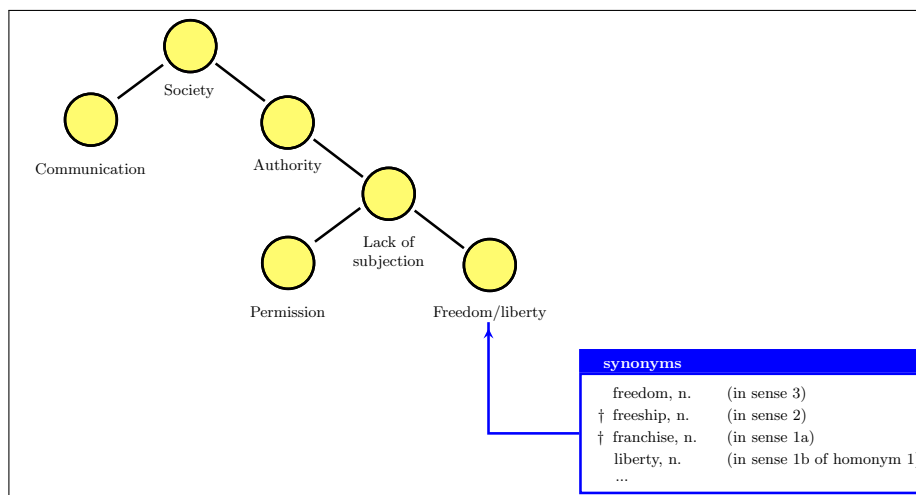


Fig. 1. Example HTOED content based on [9]

Expressing categories of the topical system of HTOED in OntoLex is relatively straightforward. Each HTOED category corresponds with a lexical concept in OntoLex. The latter is defined as a “mental abstraction, concept or unit of thought that can be lexicalized by a given collection of senses” [2]. This definition appears highly applicable to categories from topical thesauri. As lexical concepts are asserted to be specializations of SKOS concepts, it is possible to capture the hierarchy between categories using the **broader/narrower** relations from SKOS [10]. Listing 2 contains the RDF for expressing one of the HTOED categories in OntoLex, “Freedom/Liberty”, and the relation to its superordinate category “Lack of subjection”.

Listing 2. HTOED category “Freedom/liberty” expressed in OntoLex

```

<category-FreedomLiberty> a ontolex:LexicalConcept ;
  skos:prefLabel "Freedom/liberty"@en ;
  skos:broader <category-LackOfSubjection> .
  
```

The OntoLex vocabulary also contains terminology to express lexical senses and the lexical entries to which they belong. In order to state that a given lexical sense from HTOED belongs to one of its categories, the property `ontolex:isLexicalizedSenseOf` can be used. This property relates a lexical sense to a lexical concept, stating that it “lexicalizes” that concept. According to the section on Lexical Nets in the OntoLex specification, lexical senses that lexicalize the same concept are considered synonymous [2]. In other words, the relation of synonymy is not explicitly asserted in OntoLex, but can be inferred from the use of the `ontolex:isLexicalizedSenseOf` property. The resulting

RDF for the sense of *freedom* from the HTOED sample and its relation to the “Freedom/liberty” category is provided in Listing 3.

Listing 3. HTOED sense of *freedom* expressed in OntoLex

```
<sense-freedom-n-3> a ontolex:LexicalSense ;
  skos:prefLabel "freedom n. (sense 3)"@en ;
  ontolex:isSenseOf <entry-freedom-n> ;
  ontolex:isLexicalizedSenseOf <category-FreedomLiberty> .

<entry-freedom-n> a ontolex:LexicalEntry ;
  skos:prefLabel "freedom, n."@en ;
  wn:partOfSpeech wn:noun .
```

As shown, capturing the onomasiological ordering of the HTOED lexicon presents no issues with the OntoLex vocabulary. The vocabulary enables one to express categories and their hierarchy, lexical senses and their relation to a lexical entry, and the relation between the senses from HTOED and the categories to which they belong.

4 Case study *The Scots Thesaurus*

The second case study in this paper concerns *The Scots Thesaurus* (ScT) [5]. ScT captures the Lowland Scots lexis available throughout history, from its twelfth-century beginnings to the present. This thesaurus, published in 1990, categorizes its lexical items but does not indicate synonymy. Figure 2 depicts the sample taken from ScT, encompassing five categories and four lexical senses.

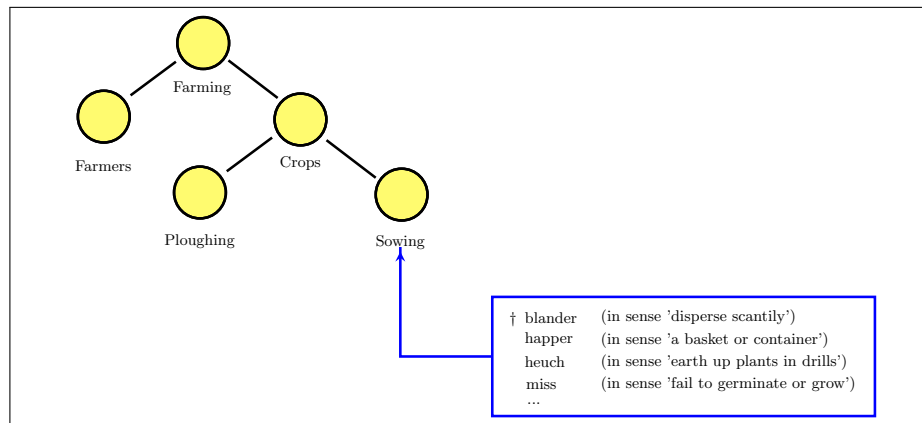


Fig. 2. Example ScT content

Expressing categories from ScT is possible in a manner identical to that used for HTOED. The result for the “Sowing” category from ScT, including its relation to the superordinate category “Crops”, is provided in Listing 4.

Listing 4. ScT category “Sowing” expressed in OntoLex

```
<category-Sowing> a ontolex:LexicalConcept ;
  skos:prefLabel "Sowing"@en ;
  skos:broader <category-Crops> .
```

As for the lexical senses from ScT, these too can be expressed in OntoLex comparable to how it has been done for HTOED. There is, however, a notable difference. The property `ontolex:isLexicalizedSenseOf` is unsuitable for relating the senses of ScT to the categories to which they belong. The lexical senses in ScT are not necessarily lexicalizations of the category in question. Moreover, senses that belong to the same category are not necessarily considered synonymous. In fact, they rarely are. A case in point are the senses of *happer* and *miss* from the sample. Both of these senses are members of the category “Sowing”, and indeed belong to that semantic domain, but can hardly be said to be synonymous or even to lexicalize the category.

What is missing, then, from the OntoLex vocabulary is terminology to express a looser manner of onomasiological ordering with categories than `ontolex:isLexicalizedSenseOf` does. The RDF snippet in Listing 5 contains the desired situation, where a tentative property `isSenseIn` is coined (see highlighted line) to express the relation between the sense of *blander* and the category to which it belongs.

Listing 5. ScT sense of *blander* expressed in OntoLex

```
<sense-blander-v-disperseScantly> a ontolex:LexicalSense ;
  skos:prefLabel "blander"@sco ;
  skos:definition "disperse scantily"@en ;
  ontolex:isSenseOf <entry-blander-v> ;
  :isSenseIn <category-Sowing> .

<entry-blander-v> a ontolex:LexicalEntry ;
  skos:prefLabel "blander, v."@sco ;
  wn:partOfSpeech wn:noun .
```

In short, OntoLex itself does not yet provide terminology to onomasiologically order the lexicographical content of ScT – and of other thesauri like it.

5 Discussion

The two case studies have shown that OntoLex is not yet expressive enough to indicate the relation between senses and categories for all topical thesauri. In fact,

the lack of a property like the tentative `isSenseIn` does not just affect conveying content from ScT and the great many existing cumulative thesauri like it. It also affects expressing these very relations found in thesauri such as HTOED. After all, senses in HTOED are not just lexicalizations of a category, they are also members of a number of categories. To illustrate, the assertion that the HTOED sense of *freedom* is a lexicalization of the category “Freedom/liberty” entails that this sense is a member of not just that category but also of its superordinate categories (see Listing 6).

Listing 6. HTOED sense of *freedom* and its relation to the categories of HTOED

```
<sense-freedom-n-3> a ontolex:LexicalSense ;
  ontolex:isLexicalizedSenseOf <category-FreedomLiberty> ;
  :isSenseIn <category-FreedomLiberty> ,
             <category-LackOfSubjection> ,
             <category-Authority> ,
             <category-Society> .
```

In order to truly express how senses are categorized according to topical systems in thesauri, then, additional terminology is required beyond what OntoLex currently offers. Properties from other vocabularies that might fill the gap, such as the `subject` property from Dublin Core Terms [11], tend to be too generic to be able to infer further knowledge from topical systems of thesauri. Moreover, the relation between such properties and `ontolex:isLexicalizedSenseOf` is not evident. As such, the required terminology is best captured in an update of the OntoLex vocabulary itself. The small addition of a single property such as `isSenseIn` (see Listing 7), then, and asserting its connection to the existing OntoLex property (see Listing 8) would enable onomasiological ordering of lexicons in topical thesauri of all varieties – distinctive or cumulative, and regardless of whether synonymy is indicated between senses.

Listing 7. Suggested OntoLex property `isSenseIn`

```
ontolex:isSenseIn a owl:ObjectProperty ;
  rdfs:label "is sense in"@en ;
  rdfs:comment "This property relates a lexical sense to a
               concept that captures its meaning to some
               extent (that is, partially or even fully)."@en ;
  rdfs:domain ontolex:LexicalSense ;
  rdfs:range ontolex:LexicalConcept .
```

Listing 8. Connection between existing OntoLex property and the suggested one

```

ontolex:isLexicalizedSenseOf
  rdfs:subPropertyOf ontolex:isSenseIn .

```

6 Conclusion

This paper has shown, by means of two case studies, to what extent the OntoLex vocabulary currently supports relating lexical senses to the concepts that facilitate an onomasiological ordering. Such an ordering is (by their very definition) used in lexicographical works known as topical thesauri. As it stands, the OntoLex vocabulary offers some support for those thesauri considered to be distinctive and that capture synonymy. Such thesauri ensure that lexical senses displayed at a certain category do not just belong to that category, but also express (or lexicalize) that category. Those thesauri that do not have that same level of specificity, but merely use their categories to organize lexical senses into semantic domains, are not yet supported by the terminology in OntoLex.

The small addition of a single property, as suggested in this paper, would have a big impact on the expressivity of OntoLex. The onomasiological ordering of both distinctive and cumulative thesauri – regardless of whether these thesauri indicate synonymy – could then properly be conveyed on the Semantic Web. As a result, the variety of lexicographical resources that sit comfortably in OntoLex would not be limited to dictionaries and lexical nets, as is presently the case, but would also include thesauri. Increased support in OntoLex for onomasiological ordering, then, would allow all these resources to truly shine on the Web. In short, ordering by meaning through the new `ontolex:isSenseIn` is both meaningful and sensible.

References

1. McCrae, J., Aguado-de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gmez-Prez, A., Gracia, J., Hollink, L., Montiel-Ponsoda, E., Spohr, D., Wunner, T.: Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation* **46**(4), 701–719 (2012)
2. Lexicon Model for Ontologies: Community report, 10 May 2016 (2016). URL <http://www.w3.org/2016/05/ontolex/>
3. Global WordNet Association: Global Wordnet formats. URL <http://globalwordnet.github.io/schemas/>
4. Kay, C., Roberts, J., Samuels, M., Wotherspoon, I. (eds.): *Historical thesaurus of the Oxford English Dictionary: with additional material from "A thesaurus of Old English"*. Oxford University Press, Oxford (2009)
5. Macleod, I., Cairns, P., Macafee, C., Martin, R. (eds.): *The Scots thesaurus*. Aberdeen University Press, Aberdeen (1990)

6. Kay, C., Alexander, M.: Diachronic and synchronic thesauruses. In: P. Durkin (ed.) *The Oxford handbook of lexicography*, pp. 367–380. Oxford University Press, Oxford (2016)
7. Postma, M., van Miltenburg, E., Segers, R., Schoen, A., Vossen, P.: Open Dutch WordNet. In: *Proceedings of the Eighth Global Wordnet Conference*. Bucharest, Romania (2016)
8. Beckett, D., Berners-Lee, T., Prud’hommeaux, E., Carothers, G.: *RDF 1.1 Turtle: W3C recommendation 25 February 2014* (2014). URL <http://www.w3.org/TR/turtle/>
9. *Historical thesaurus of the Oxford English Dictionary* (2010). URL <http://oed.com/thesaurus>
10. *SKOS Simple Knowledge Organization System reference: W3C recommendation 18 August 2009* (2009). URL <http://www.w3.org/TR/skos-reference/>
11. *DCMI metadata terms* (2012). URL <http://purl.org/dc/terms/>