

Philosophising Lexical Meaning as an OntoLex-Lemon Extension

Veruska Zamborlini* and Jiaqi Zhu and Marieke van Erp

DHLab, KNAW Humanities Cluster, Amsterdam, the Netherlands

*Ontology & Conceptual Modeling Research Group, Federal Univ. of Esp rito Santo, Brazil

{veruska.zamborlini, jiaqi.zhu, marieke.van.erp}@dh.huc.knaw.nl

Arianna Betti

Universiteit van Amsterdam, Amsterdam, the Netherlands

a.betti@uva.nl

Abstract

OntoLex-Lemon is a model for representing lexical information, focusing on the use of lexical entries in texts rather than their definitions. This work proposes an extension to the model that aims to capture the definition of senses attributed to lexical entries. We explicitly represent a conceptual setup authored by an agent that operates on lexical content. It either proposes new senses for existing lexical entries in a language or coins new terms to express proposed senses. It provides textual and/or formal definitions to senses/concepts, and can serve as an interpretation of other senses/concepts through rephrasing, translation, formalization, or comparison. Because a conceptual setup and its interpretations may not be unanimously accepted, it is important to support the selection of relevant meanings, as for example, those proposed by a certain author. We illustrate the application of our proposed extension with two case studies, one about the philosophical definition of the concept of *idea* and its interpretations, and one about historical attributions of meaning to the Dutch East India Company (VOC).

1 Introduction

The OntoLex-Lemon¹ W3C recommendation for representing lexical information focuses on the various usages of lexical entries in texts. While this approach has proven effective in many contexts, it was not designed to capture the definitions that underpin the lexical senses attributed to the entries. Several extensions have been proposed to enhance the expressiveness of the OntoLex-Lemon model in different aspects, such as capturing morphological decomposition (*decomp* module), representing translations and lexical variation (*vartrans* module), describing metadata about lexical resources (*lime* module), and also linking multilingual linguistic resources (through Linguistic Linked Open

Data (LLOD) initiatives) (Khan et al., 2022; Gromann et al., 2024). However, to the best of our knowledge, none of them directly addresses the need to represent the definitional and interpretative foundations of lexical senses or concepts.

To address this gap, we propose an extension to the OntoLex-Lemon model at a conceptual level, that is, not yet implemented. The extension enables explicit representation of a conceptual setup providing meaning attributed to lexical entries as textual or formal definitions by original authors and/or by other authors interpreting the original ones. In this work we use the term *conceptual setup* as generic label for a (loose) view/conceptualization (e.g., a term coined in journalism), an expert-level conceptualization/theory (e.g., a domain-specific definitions in a scholarly text), or a fully developed theory (e.g., a formal philosophical framework). We also refer to definition of *lexical sense* and *lexical concept* somewhat interchangeably, as the latter is typically lexicalized through the former in a particular language. Finally we consider interpretation as rephrasing, translating, explaining or formalizing someone else’s conceptual setup with the intention of preserving the intended meaning, as opposed to (i) intentionally changing the meaning (as in correcting or complementing it) or (ii) directly/originally describing a conceptual set up (as in "interpreting reality").

While not all lexical senses have a specific source/author for their definitions, and usage may diverge from original definitions, our proposed extension aims to systematically capture those definitions and their interpretations for which there is traceable and verifiable evidence. This approach thus aims to enrich representations of lexical meaning, ultimately supporting the analysis and understanding of how concepts evolve over time.

To illustrate and motivate our proposal, we present two case studies from the digital humanities domain in Section 3. The parallel of these two

¹www.w3.org/2016/05/ontolex/

cases is that both combine computational methods and digital humanities expertise to deal with the challenge of how concepts evolve or change over time in their specific domain, namely philosophy and history. They address similar research questions, such as: How have certain concepts changed/evolved? What kind of changes have they undergone? And how to model these conceptual changes in a way that is computationally manageable and interpretable for answering humanities research questions? By introducing our extension to the OntoLex-Lemon model applied to case studies from the domains of philosophy and history, we demonstrate its potential for supporting and enriching digital humanities research in general. However, the extension is broadly conceived and therefore applicable to other domains where lexical definitions also evolve or diverge, such as medicine, law, and science (Oortwijn et al., 2021).

The first example is from the *eIdeas* project within the "Concepts in Motion" lab,² a group aiming to trace computationally how concepts evolve over time. In philosophy, (re)interpreting a theory, or a concept within a theory, often requires a profound and concrete understanding of the implications behind the words that are used to formulate the concept or theory in question. During the interpreting process, interpreters usually need to assign meanings to the word/lexicon that is core to the concept/theory according to their understanding. Based on different underlying assumptions or philosophical perspectives, interpreters can have various interpretations and applications of the original concept/theory. To have new insights or approaches to a concept/theory, philosophers usually need to engage with the arguments and counterarguments that have been proposed about this concept/theory over time. In this background, our extension to the OntoLex-Lemon model can help trace the evolution and (re)interpretations of a philosophical concept/theory by providing a dynamic and multifaceted perspective on its meaning and applications.

The second example is from the "Trifecta" project,³ which combines computational linguistics and semantic web technologies to extract and model, from the maritime and food history domains, concepts in their contexts, such as the Dutch East India Company, slavery, coffee, and cinnamon. (van Erp, 2023) points out that Large Knowledge

Graphs (KGs) such as Wikidata and DBpedia only express a limited representation of the concepts and entities they represent. For instance, at the time of writing (van Erp, 2023), DBpedia focuses in representing the concept *coffee* on the food dimension, while it could be explored through multiple aspects, such as a plant, the activity of drinking the drink, a colonial good, and more. The project aims to automatically capture different dimensions of concepts in various contexts and represent this multi-dimensionality in Knowledge Graphs. Towards this goal, Trifecta focuses on dealing with key challenges: a. identity (what the concept is and how it is perceived), b. change (how this concept evolved over time), and c. the long tail (what low-frequency contexts are connected to this concept). Linguistic information supported by the OntoLex-Lemon ontology can play a role in tackling these challenges. The schema in Figure 1 that illustrates scholarly and historical texts representing different meanings and interpretations attached to the concept of the VOC.

The remainder of this paper is organised as follows. In Section 2 we discuss related work, followed by two case studies stemming from digital humanities scenarios and competency questions for the proposed model to address in Section 3. In Section 4 we present our proposed extension to the OntoLex-Lemon model by providing UML representation, and we illustrate our extension with two schemas that instantiate the model for our case studies. We present our discussion revolving around the relationship to other modules and models in Section 5 and to what extent the competency questions are addressed in Section 6. Our conclusions and directions for future work respectively are presented in Section 7.

2 Related Work

OntoLex-Lemon (McCrae et al., 2017) results from an effort of the Ontology Lexicon (OntoLex) community group becoming a W3C standard model for providing rich linguistic grounding to ontologies. It provides means to connect ontology entities to lexical entries with their morphological and syntactic properties. It is designed to be combined with the other four OntoLex modules: syntax and semantics (*synsem*), decomposition (*decomp*) variation and translation (*vartrans*) and linguistic metadata (*lime*). Several modules and extensions to OntoLex are reviewed in (Gromann et al., 2024) (and similar

²<https://concepts-in-motion.org/>

³<https://trifecta.dhlab.nl/>

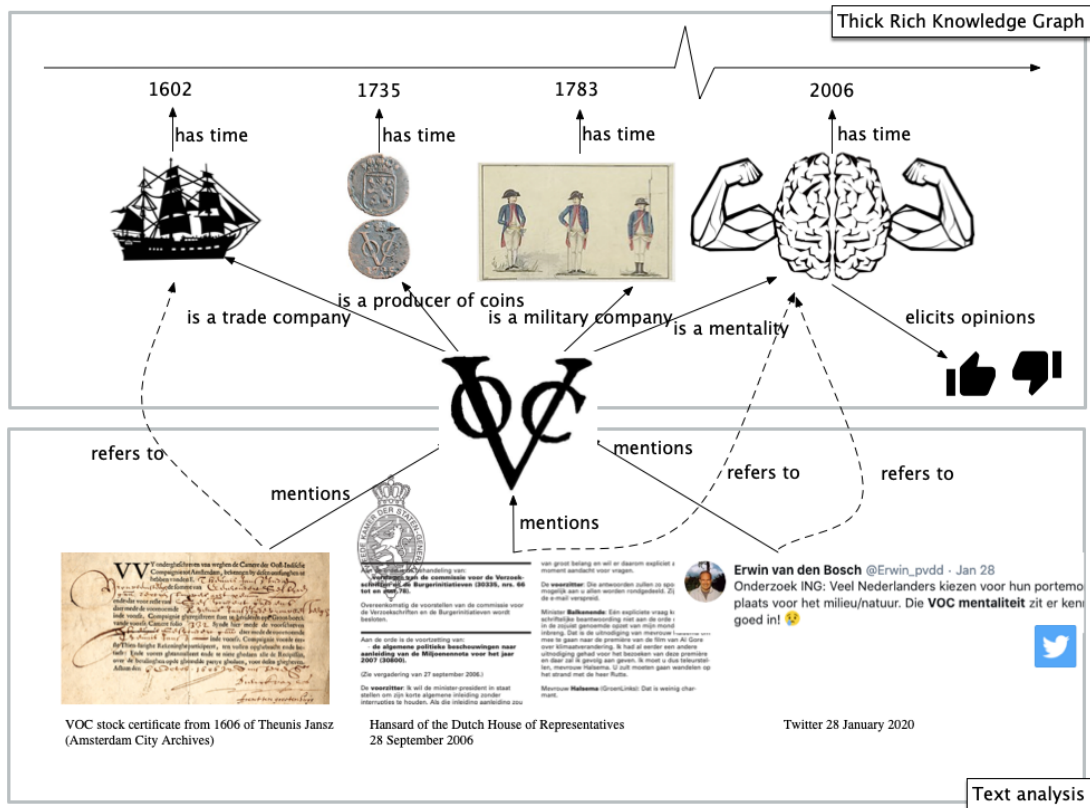


Figure 1: Schematic illustration of meanings VOC can take on over time

examples throughout).

The *synsem* module is concerned with providing semantics to the lexical entries by connecting them to existing ontologies that may provide a formal specification to constrain the meaning of a concept. The goal of our proposal is (i) not to rely on the existence/adequacy of an OWL⁴ ontology, but on documented conceptualizations/theories backing the senses attributed to the entries; (ii) to allow for formal definitions to be provided in formal languages beyond OWL and (iii) to allow for an OWL ontology to be derived from well-annotated theoretical sources ultimately providing also a detailed provenance for the concepts in the ontology.

The *vartrans* module models translation as a relation between senses, defining an exact (non-questionable) correspondence between them. In contrast, we propose an interpretation relation among lexical concepts that accommodates subjectivity and variation. It could ultimately serve as translations when the provided interpretation is accepted as valid.

The *lime* module (Linguistic MetaData) provides a standardized way to describe metadata about linguistic resources, such as the lexicon or

the conceptualization set. It includes information about the language(s) covered, the number of lexical entries, the structure of the lexicon, and its linkage to other resources. This module supports interoperability and facilitates the discovery and integration of lexical datasets on the Web. However, it does not suffice to address the requirements for evidence supporting the attribution of meaning to lexical entries as envisaged in our proposal.

(Khan et al., 2022) provides an overview of research projects that use linguistic linked data vocabularies to create and publish lexical resources in various languages using the OntoLex-Lemon model (and its extensions). The paper introduces representative projects across various domains and use cases, including digital humanities, and discusses the influence of these projects on the use or definition of linguistic linked data models and vocabularies in detail. Two examples of related projects designed for the lexical modeling of historical domain-specific vocabulary are DitMAO–LexO–MAIA (Giovannetti et al., 2024) and ALMA (Tittel, 2023). Both initiatives share a similar goal: to capture the context in which certain senses were used (or proposed), particularly within historical or scholarly sources. However, their so-

⁴<https://www.w3.org/OWL/>

lutions do not explicitly model the provenance, interpretation, and formalization of definitions.

3 Case Studies and Competency Questions

This section presents two case studies that were devised with domain experts in the humanities domain including a set of competency questions for the proposed model to address.

3.1 Bolzano's Theory of Ideas & Interpretations

In his book *Wissenschaftslehre* (1837) (Bolzano, 1837), the Bohemian philosopher Bernard Bolzano proposes a theory in which he defines, among others, the term *Vorstellung* as "*Vorstellung [ist] dasjenige, was als Bestandtheil in einem Satze vorkommen kann, für sich allein aber noch keinen Satz ausmacht.*" (Bolzano, 1837) §. 48, which is translated as "that which can occur as a component in a sentence, but which on its own does constitute a sentence".

More than a century later, the Italian philosopher Betti, in their book chapter "Bolzano's Universe: Truth, Logic and Metaphysics (2012)" (Betti, 2012), renders Bolzano's *Vorstellung* as *Idea*, and rephrases the definition as "*an idea is that part of a proposition that is not itself a proposition*".

This interpretative chain continues: Betti's student Hungerbühler, in his thesis "A computational method for philosophical interpretation (2018)" (Hungerbühler, 2018), offers yet another layer by formalizing the concepts from Betti's interpretation using OWL Description Logics, such as the formal definition of *Idea* described in Listing 1. By reasoning over these formal definitions, Hungerbühler's thesis provides interesting insights on the definitions of concepts and their interpretations.

```
Class: Idea
SubClassOf: partOf some Proposition
DisjointWith: Proposition
```

Listing 1: Manchester OWL Syntax Example

This chain of provenance is essential, for example, when discrepancies arise: if an inconsistency is found using formalisations such as Hungerbühler's, the issue can be traced back to verify if it stems from his own reinterpretation, from Betti's interpretation, or from Bolzano's original theory. This case study illustrates possible benefits of a representation that allows for keeping the provenance of the original documents from which the definitions

are taken along with the chain of interpretations. Furthermore the ability to use formal syntaxes (besides OWL) to describe concepts allows for later extraction of a formal model as input for reasoners and analysis of the results.

3.2 VOC as a "company-state"

In the 17th and 18th centuries, the Vereenigde Oost-indische Compagnie (Eng., *Dutch East India Company*) (VOC) played an important role in early modern world history. The VOC was set up in the Dutch Republic as a trading company to trade with and in Asia, and soon created a trading network of colonies and settlements in Asia and Africa (Gaasstra, 2003). Relying on the archives of the VOC as source material, historians discuss the VOC and its role in history from different perspectives, for example, in early modern global trade (Israel, 1989), in cross-cultural encounters (Blussé, 1986), and also in colonisation in Asia and Africa (Schrikker, 2007; Emmer, 2003). However, the VOC is a complex concept as it has conducted various kinds of activities and thus can be interpreted in various ways. In this paper, we focus on a certain historical perspective which understands the VOC as both commercial and political for its functions in both, and show how our extension to the model can help represent the relationship between different interpretations.

Inspired by historian Philip Stern's analysis of the English East India Company (EIC) as a "company-state" in the book (Stern, 2011), historian Arthur Weststeijn argued that the VOC should also be considered as a "company-state" as in (Weststeijn, 2014). Reinterpreting Stern and Weststeijn's "company-state" arguments, historian Erik Odegard further applied this perspective of understanding the VOC both as a ruler and a merchant in formulating his argument in (Odegard, 2020). Similar to the first case study on Bolzano's Theory of Ideas and Interpretations, our extension to the OntoLex model provides a structure that allows various interpretations or perspectives on the VOC to be presented and compared, which historians could benefit from. We propose this structure to enable researchers to work with interpretive complexity rather than flattening concepts under investigation. Our extension links interpretations to their authors and sources, which we expect will enable computational tracking of how arguments develop and circulate. The intended outcome is that researchers will be able to analyze not just what previous schol-

ars claim about a concept, but how these claims relate to different theoretical frameworks and textual sources. We anticipate this will create new possibilities for understanding historical knowledge.

3.3 Competency Questions

We defined the following competency questions with the domain experts. These questions were chosen based on their relevance for the type of research the domain experts want to conduct and serve as guidance and evaluation for the types of information our extension needs to cover.

- CQ1** What are all the definitions of a given lexical entry, along with their direct or indirect authors? (*e.g., Idea as defined by Betti or Hungerbühler and VOC as defined by Weststeijn and Odegard*)
- CQ2** Which concepts have been (re)defined by a particular author? (*e.g., all concepts defined by Bolzano or Odegard*)
- CQ3** What are the various interpretations that have been proposed for a specific conceptual setup? (*e.g., Betti's interpretation of Bolzano's theory; or Odegard interpretations of Stern and Weststeijn's theories.*)
- CQ4** What are all the interpretations proposed by a specific author? (*e.g., all interpretations by Betti, for example, for Bolzano, as well as all interpretations by Odegard for example, for Stern and Weststeijn.*)
- CQ5** What is the formal representation of all concepts included in a conceptual setup? (*e.g. Hungerbühler's formal interpretation of the Theory of Ideas in Manchester OWL syntax*)
- CQ6** How has a concept evolved over time, both in general and through contributions by particular authors? (*e.g. how the several definitions provided for the concept VOC have evolved in time and through different narratives, or have an author such as Bolzano provided different definitions or refinements for the concept of Idea in different works*)
- CQ7** Which definitions of terms are closer or farther away from each other? How close are they? (*e.g. how is the definition of Bolzano for Idea close the one by Aristotle, or is it closer to Aristotle's than Locke's definition? Or yet, how the definitions of the VOC relate to each other, as in agreement, complementarity, contradiction or others*).

4 OntoLex-Lemon Extension

The proposed extension is depicted in Figure 2 using a UML⁵ diagram. The classes and relations from OntoLex-Lemon and its modules are prefixed accordingly (*olex* as short for *OntoLex*) and depicted in shades of green and yellow, while the proposed ones are not prefixed and are depicted in purple color. Our proposal is intended as a modular extension, specializing or complementing the entities of the OntoLex-Lemon framework, enabling the representation of the provenance of lexical entries, senses and concepts, and the modeling of interpretive or derivational relationships between them.

A *View/Conceptualization* is composed of *Defined Lexical Concepts* lexicalized as *Defined Lexical Senses*. It is *authored* by an *Agent* and *authored at* a certain point in time (*temporal extension*) in an *Creation Event* possibly *described in* a *Document*. It is *expressed as* a *Lexicon* and may also coin a *CoinedLexicalEntry*. When the author is an *Specialised Agent* (for the subject in question), then the *Conceptualization* can be considered as a *Theory*. If it provides *FormalisedLexicalConcept* with a *formal definition*, it is then a *FormalisedTheory*. A formalization can be provided in any language/syntax, such as OWL Manchester Syntax⁶ or SWI-Prolog⁷. As long as they are provided with the appropriate "language" annotation, e.g. `@manchester` or `@swiprolog` or `^^:manchester` or `^^:swiprolog`, a script can select the formal definition of selected concepts to compose an output description that can be input for a proper reasoning service.

Moreover, a *View/Conceptualization* can be an *Interpretation* of one or more *Views/Conceptualizations*, if it provides *LexicalConcepts* that are *interpretations of* concepts in other *Views/Conceptualizations*. If it provides interpretations for all the entries in another *View/Conceptualization*, the *Lexicon* expressing the interpreting one can be modeled as providing an *interpretation of* the *Lexicon* expressing the other under interpretation.

To illustrate the proposed extension, we present two schemas that instantiate the model for our case studies. The color code refers to the respective classes according to Figure 2, having the lexical entries and their forms grouped in a gray box. First, Figure 3 illustrates the case study **Bolzano's Theory of Ideas & Interpretations** (subsection 3.1).

⁵<https://www.uml.org/>

⁶www.w3.org/TR/owl2-manchester-syntax/

⁷www.swi-prolog.org/

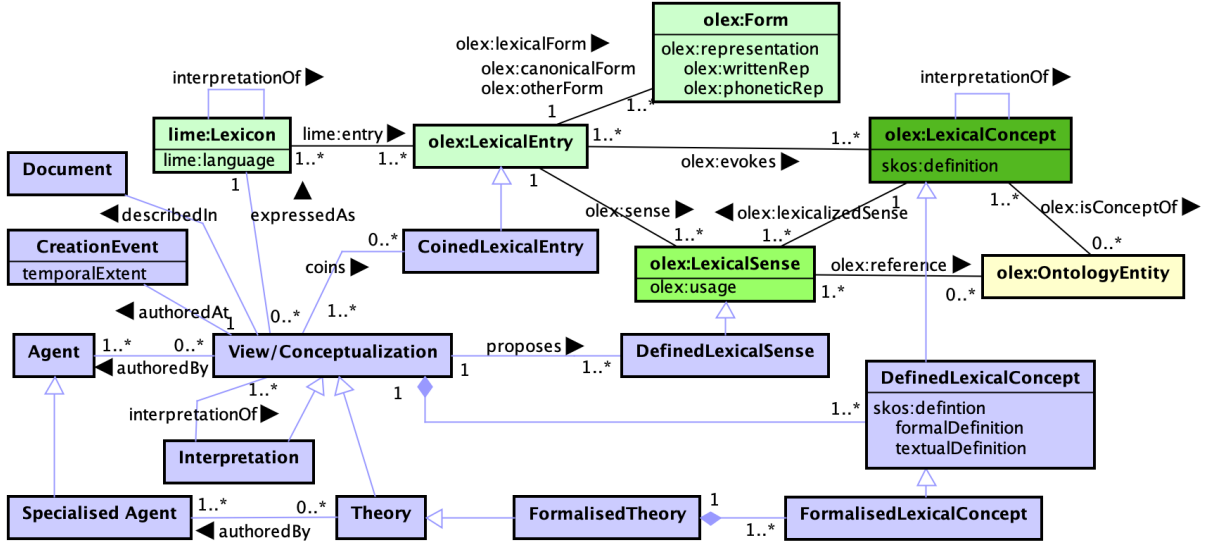


Figure 2: UML representation of (part of) OntoLex-Lemon model (indicated with prefixes and depicted in green shades and yellow) and the proposed extension (depicted in purple).

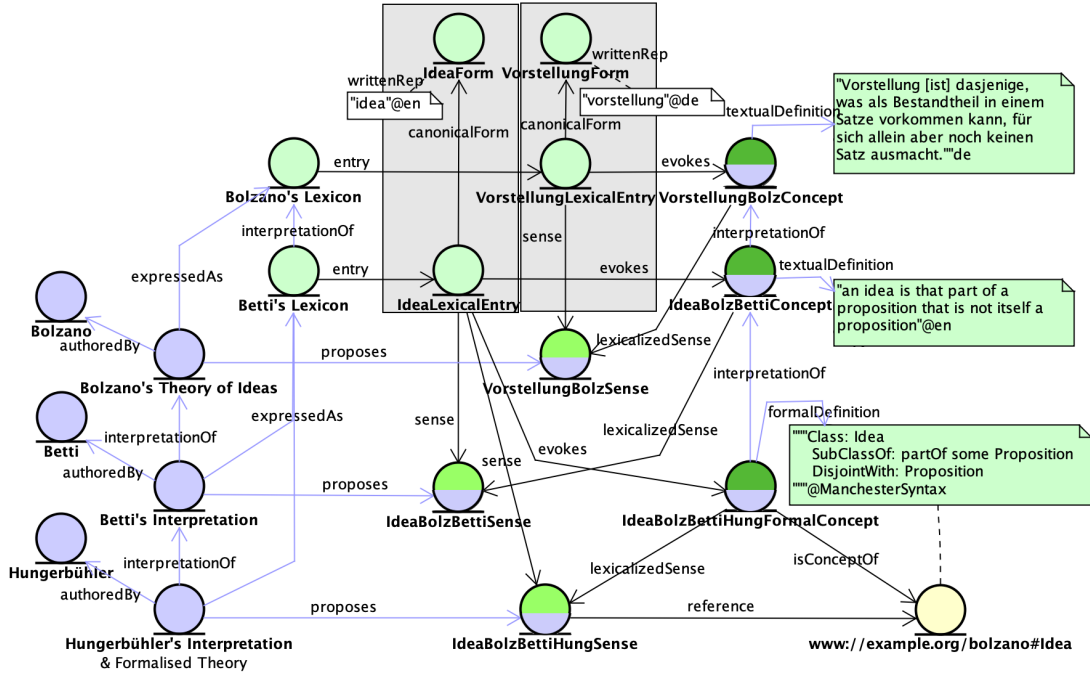


Figure 3: Schema representing an instantiation of the model for Bolzano's Theory of Ideas and its interpretation by Betti. Specifically for the entry *Vorstellung* in German, Betti proposes it as *Idea* in English rephrasing its definition, while Hungerbühler provides an interpretation of Betti's interpretation with a formalization in Description Logics using Manchester syntax (the color code refers to the respective classes in Figure 2).

The lexical entry for which the canonical form is *Vorstellung* in German, has its corresponding sense proposed by the mentioned theory, which is the lexicalized sense of a lexical concept having as definition the original text by Bolzano (detailed web annotation provided later). This provides provenance for the sense, namely, that it originates from Bolzano's theory.

Analogously, Betti's theory proposes a sense for the lexical entry with canonical form *Idea* that evokes the lexical concept whose textual definition in English is the rephrasing in the original text by Betti. It is an interpretation of the lexical concept lexicalized by the sense proposed by Bolzano. Rather than asserting that the senses for *Idea* and *Vorstellung* refer to exactly the same sense or are

a translation of each other, which would imply a perfect equivalence, we instead represent that *Idea*, as proposed in Betti's chapter, as linked to a distinct sense that is an interpretation of Bolzano's original sense. In this way, we preserve both the nuance of interpretation and the provenance of each contribution.

Furthermore, because Betti's entire chapter is dedicated to interpreting Bolzano's work, we model their theory as an interpretation of Bolzano's original theory. The lexicon that expresses Betti's theory thus provides interpretations of the lexical entries that express Bolzano's theory, or translations for them if Betti's interpretations are taken as valid.

Finally, Hungerbühler's theory provides a formalization in OWL-DL that is an interpretation of Betti's theory, which in turn interprets Bolzano's original theory. The sense proposed by Hungerbühler is thus expressed through an OWL class (identified here by the illustrative URI [www://example.org/bolzano#Idea](http://example.org/bolzano#Idea)) which is defined as equivalent to the formalization proposed by him. This formalization not only establishes a semantic anchor for the sense in question but also enables its use in automated reasoning tasks. By expressing the definition in a formal language such as OWL-DL (e.g., in Manchester Syntax), an OWL ontology can be generated and reasoned over using standard semantic web tools. It is important to note that although OWL-DL was selected for this particular case study, the proposed approach is not restricted to it; any other formal representation language could be employed to capture the definitions and support similar reasoning workflows.

Next, Figure 4 illustrates the case study **VOC as a "company-state"** (subsection 3.2). Here we have two lexical entries, for which the form is *EIC* in English (English abbreviation for the "English East India Company") and another one for which the form is *VOC* in Dutch (Dutch abbreviation for the "Dutch East India Company"). First, Stern proposes a sense for *EIC* entry that evokes the lexical concept whose textual definition in English describes it as a company-state. Next Weststeijn applies Stern definitions as an analogy to the *VOC* concept, actually proposing to it also a sense to the corresponding entry that evokes a similarly defined lexical concept. Finally, Odegard agree with them both, rephrases their definitions applied to both *EIC* and *VOC* entries, this providing a reinterpretation of the lexical concepts lexicalized by the senses proposed by Stern and Weststeijn. Important to

notice, an dotted red arrow connecting Stern and Weststeijn's concepts is meant to express the analogy relation between the concepts, which however has not being yet included in our proposal and is therefore object of investigation for future work.

Finally, Figure 5 illustrates how the Web Annotation Vocabulary⁸ can be used to document the provenance of both lexical entries and concepts. It describes two annotations having as source the same book of Bolzano. One has as body the lexical entry *Vorstellung* and the other has as body the lexical concept the entry evokes. They have selectors that describe the exact text referring to the bodies of the annotation (respectively the lexical entry and its definition) and indicating their location in the whole text by assigning a prefix and suffix.

5 Relation to other modules and models

In this section we discuss the possible relations of our proposal with two Ontolex modules, in particular *lime* and *vartrans*, and with *Prov-O*. We will further investigate the positive or negative consequences before incorporating them into the proposed extension.

The *lime* (The LInguistic MEtadata)⁹ module defines a *lime:ConceptualizationSet* as associating a *ontolex:ConceptSet* with a *lime:Lexicon*. One could consider a *View/Conceptualization* as a specialization of *ontolex:ConceptSet*, although the latter is clearly more than a just set of concepts. It may be a related to a *lime:ConceptualizationSet* since it is meant to bind the lexical concepts in the concept set and entries in the lexicon.

The relation to the *vartrans*¹⁰ module, which defines lexico-semantic relations such as *vartrans:Translation* between *ontolex:Senses* brings a more complex issue with respect to how to define the *Interpretation* and what is its relation between *vartrans:Translation* and *Interpretation*. We see a few possibilities: (i) a *vartrans:Translation* is a specialization of *Interpretation*, which would mean, among other things, that they would have to hold both between *ontolex:LexicalSenses* or between *ontolex:LexicalConcepts*; (ii) a variation of the

⁸<https://www.w3.org/TR/annotation-vocab/>

⁹[https://www.w3.org/community/ontolex/wiki/Final_Model_Specification#Metadata_\(lime\)](https://www.w3.org/community/ontolex/wiki/Final_Model_Specification#Metadata_(lime))

¹⁰[https://www.w3.org/community/ontolex/wiki/Final_Model_Specification#Variation_&_Translation_\(vartrans\)](https://www.w3.org/community/ontolex/wiki/Final_Model_Specification#Variation_&_Translation_(vartrans))

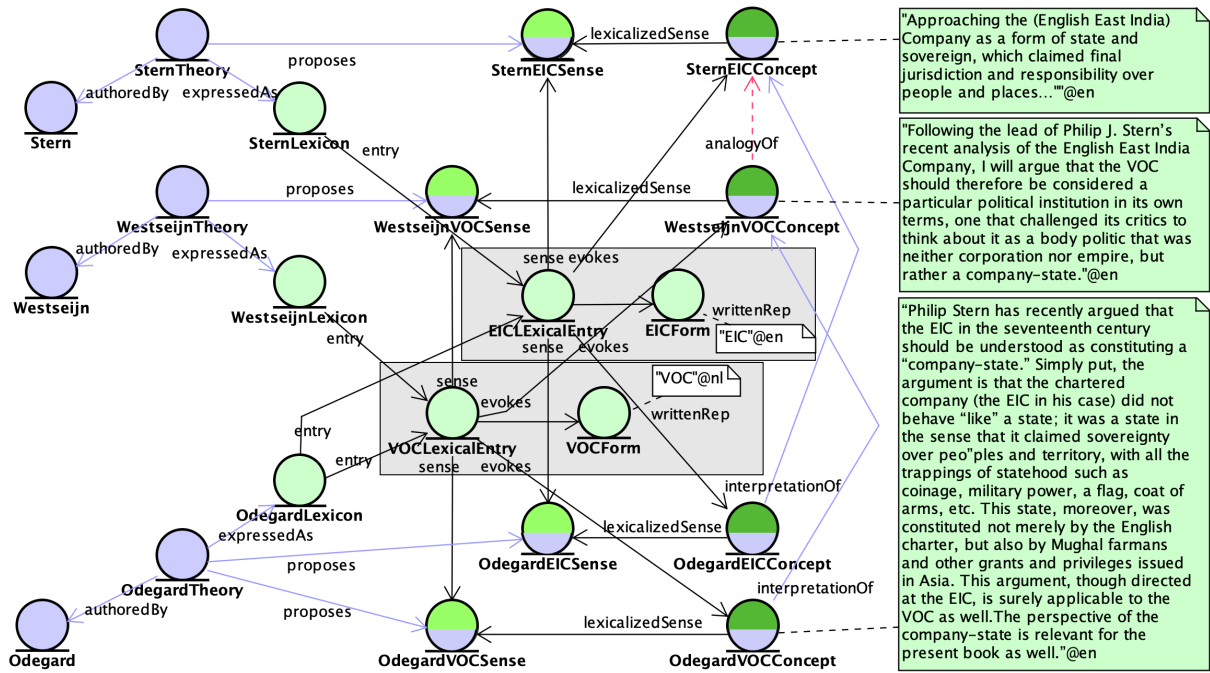


Figure 4: Schema representing an instantiation of the model for EIC and VOC concepts. Each of them have two senses by different historians, namely Stern, Weststeijn and Odegard (the color code refers to the respective classes in Figure 2).

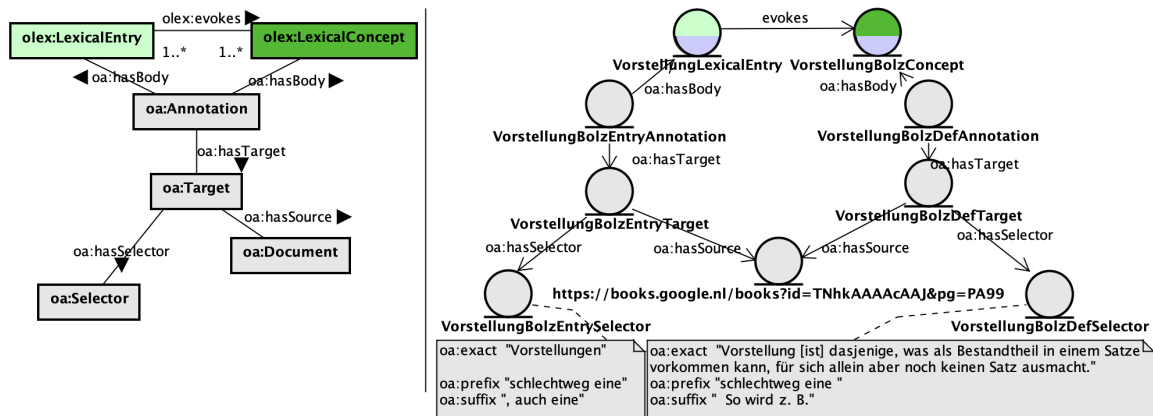


Figure 5: Provenance annotation for the lexical entry *Vorstellung* alongside the concept it evokes (the color code refers to the respective classes in Figure 2 while Open Annotation ones are prefixed with *oa* and depicted in gray).

previous one is that they could be overlapping classes, meaning not all interpretations are translations and not all translations are interpretations (in the sense that they are not questionable); (iii) more aligned with our current proposal is that *Interpretation* hold between *ontolex:LexicalConcepts*, *vartrans:Translation* hold between *ontolex:LexicalSenses*, and the former could be derived from the latter; and (iv) it could also be that *Interpretation* hold between *ontolex:LexicalSenses* AND between *ontolex:LexicalConcepts*.

The **Prov-O**¹¹ aims to support the representation of provenance information, either by being directly used or by serving as a reference model for creating domain specific provenance information. Its main entities are *prov:Entity*, *prov:Activity* and *prov:Agent* among which several provenance relations hold, for example, *prov:wasDerivedFrom* indicates that an entity is changed or created based on another, while *prov:wasAttributedTo* ascribes an entity to an agent. Since our domain does require more specific provenance, such as the inter-

¹¹www.w3.org/TR/prov-o/

Table 1: Competency Questions and Support by Models

Competency Question	OntoLex Base	OntoLex+ Others	Proposed Extension
CQ1.1 What are all the definitions of a given lexical entry?	✓	n.a.	n.a.
CQ1.2 What are all the definitions of a given lexical entry, along with their authors?	✗	~	✓
CQ1.3 What are all the definitions of a given lexical entry, along with their direct or indirect authors?	✗	~	✓
CQ2.1 Which concepts have been defined by a particular author?	✗	~	✓
CQ2.2 Which concepts defined by a particular author have also the term coined by him?	✗	~	✓
CQ2.3 Which concepts have been interpreted by a particular author?	✗	~	✓
CQ3 What are all the interpretations proposed by a specific author?	✗	~	✓
CQ4 What are the various interpretations that have been proposed for a specific view/conceptualization/theory?	✗	~	✓
CQ5.1 What is the (formal) definition of a concept given a (formal) syntax?	✓	n.a.	n.a.
CQ5.2 What is(are) the formal definition(s) of a concept?	✗	✗	✓
CQ5.3 What are the formal definitions of all concepts in a given theory?	✗	✗	✓

pretation of a concept as another one, or an analogy among them, we consider that Prov-O should not be used as is, but it can be a reference model from which our proposed extension can specialize.

6 Addressing the Competency questions

In this section, we discuss whether the competency questions and some variations can be addressed by the OntoLex-Lemon Base model, by combining it with other modules or vocabularies, or by the proposed extension. Table 1 indicates if the questions are fully, partially or not addressed using, respectively, the symbols ✓, ~, ✗. Moreover, we use n.a. when OntoLex-Base address the issue and therefore no extension is necessary.

It turns out that a combination with the vocabularies *Prov-O* and *SKOS*¹² can partially simulate the semantics intended in the proposed extension if a *View/Conceptualization* is taken as a *ontolex:ConceptSet*, which is connected to *ontolex:LexicalConcept* via *skos:inScheme*, also if *prov:wasDerivedFrom* connects a *ontolex:ConceptSet* to an-

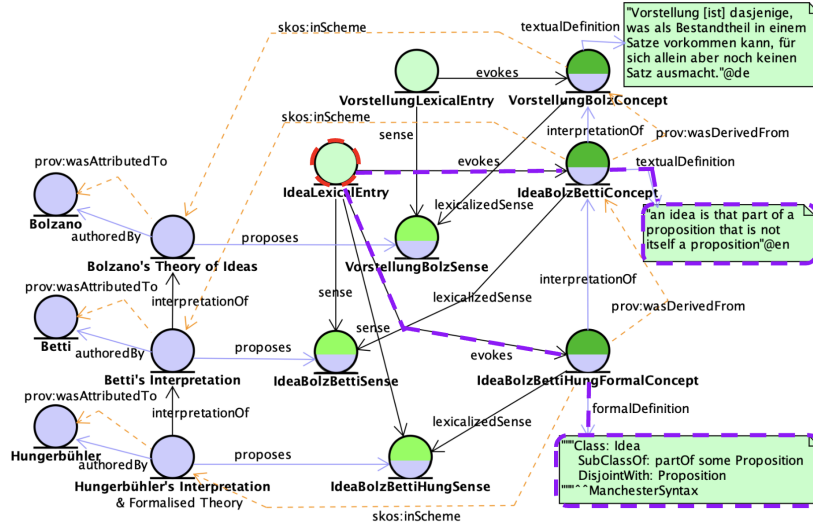
other one as *interpretation-of*, and if *prov:wasAttributedTo* indicates both the authorship of a *ontolex:ConceptSet* by *prov:Agent* and the coining of *ontolex:LexicalEntry* by a *prov:Agent*. However, the meaning may not be as clear, and therefore we consider the competency question to be partially addressed.

Figure 6 is a variation of the instantiation in Figure 3 including, in orange dashed lines, some of the aforementioned properties as alternatives to the proposed extension. It highlights the paths that could provide answers to the complementary questions CQ1.1, CQ1.2 and C1.3. The dashed purple paths illustrate the paths using the extension, while the dashed-dotted orange paths illustrate the alternative paths. It illustrates that, although similar results can be obtained with existing vocabularies, the proposed extension offers greater domain specificity, making it more suitable for guiding consistent and semantically accurate use.

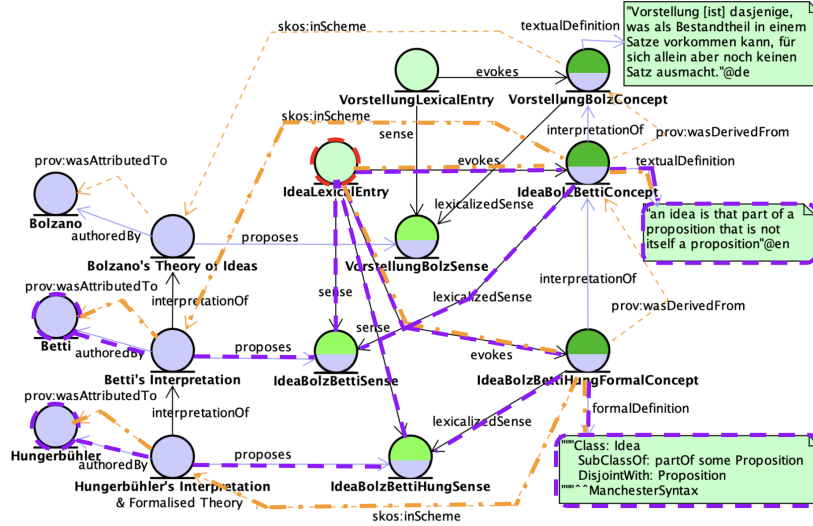
7 Conclusion and Future Work

We propose a conceptual extension to OntoLex-Lemon with the purpose of representing the provenance of senses with evidence. It allows for ex-

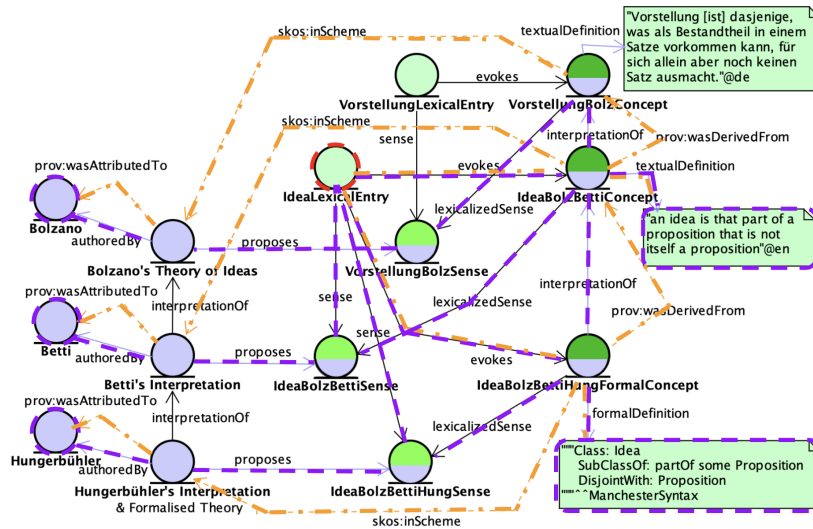
¹²<https://www.w3.org/2009/08/skos-reference/skos.html>



(a) CQ1.1 Achieves the definitions for the lexical entry *Idea* (red dashed circle) using only OntoLex-Lemon Base (purple dashed paths/lines).



(b) CQ1.2 Achieves the definitions for the lexical entry *Idea* and the authors using both OntoLex-Lemon extended and an alternative path using *Prov-O* and *SKOS* (orange dashed-dotted paths/lines).



(c) CQ1.3 Achieves the definitions for the lexical entry *Idea* and the direct and indirect authors using both OntoLex-Lemon extended and an alternative path using *Prov-O* and *SKOS*.

Figure 6: Visualization highlighting paths providing answers to the complementary questions CQ1.1, 1.2 and 1.3.

pressing conceptual setups as well as their interpretations, as well as expressing the textual or formal definitions of the concepts, accompanied by annotations leading to the excerpt of original text where the definition is provided. The current proposal addresses Competency Questions 1 to 5. Competency Questions 6 and 7 are challenging regarding evolution of concepts and comparison among them and will be addressed in future work, as well as new competency questions.

As our aim is to outline, at a conceptual level, how the OntoLex-Lemon model could be extended to address the proposed competency questions, the implementation is still to be investigated. For that we would consider the reuse of existing vocabularies, such as Prov-O, SKOS and DC-Terms¹³, as well as structured representations like nanopublications¹⁴. It is also important to further investigate the connections to other modules and extensions of OntoLex. Next we will conduct a practical evaluation of our proposal by applying it not only to extended versions of our case studies but to related cases from the literature.

We furthermore plan to investigate the alignment of our proposal with existing models to address upcoming challenges. One promising direction is the integration with the Linguistic Annotation Scheme GRaSP (van Son et al., 2016), a framework that adopts a multilayered approach in four layers, namely events, attribution, factuality, and opinion. We aim to explore how our proposed model for representing definitions and interpretations can be integrated with GRaSP. In particular, the opinion layer offers a promising space to explore differing theoretical perspectives, conceptual interpretations, and scholarly disagreements.

Another relevant direction is the alignment with the LMM (Linguistic Meta-Model) (Picca et al., 2008), meant for representing heterogeneous lexical knowledge, providing a semiotic-cognitive representation of linguistic knowledge grounded in DOLCE foundational ontology (Gangemi et al., 2002). In particular, it considers different ways of assigning meaning to an expression, expliciting the ontological nature of the "meaning definitions" and the relations between them, which can be relevant for understanding how to relate and compare the definitions of lexical concepts.

In terms of modelling concept evolution, we also

want to explore geographical and temporal dimensions because the meanings and the interpretations of concepts can vary across time and geography. Geographical factors can influence how a concept is understood and used in a specific place. For example, the VOC might be perceived differently in the former Dutch colonies from a postcolonial perspective than in the Netherlands from a perspective of Dutch national history.¹⁵ Temporal dimensions, including historical periods, cultural-societal shifts, and technological advancements, also reveal how concepts evolve over time and how their interpretations change. For example, the concept of "privacy" has undergone significant transformation in the digital age, evolving from Warren and Brandeis's 1890 conception of "the right to be let alone" to contemporary debates between individual autonomy-based approaches versus social relational frameworks that "surpass the perspective of the individual" (Becker, 2019). Survey data demonstrates measurable temporal shifts in privacy attitudes, with older adults more concerned about their online security and privacy compared to the younger generations, reflecting broader cultural-societal shifts in how privacy is conceptualized in digital contexts (Holmes, 2022). Highlighting the geographical and temporal contexts from which a concept or interpretation emerges is likely to promote historiographical practices, and representing geographical and temporal information along with lexical information can contribute to this advancement. The documentation of our extension can be found at <https://github.com/trifecta-project/lexical-sense-definition>.

Acknowledgments

This work was funded by the European Union under grant agreement No. 101088548 (TRIFECTA), and by the Dutch Research Council (NWO) through the projects Culturally Aware AI (KIVI.2019.005), e-Ideas (VICI grant 277-20-007), and Counting the Bees (Open Competition grant 406.22.FHR.031). Views and opinions expressed are, however, those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. We acknowledge our

¹³www.dublincore.org/specifications/dublin-core/dcmi-terms/

¹⁴<https://nanopub.net/>

¹⁵<https://internationaleonline.org/contributions/the-dutch-voc-mentality-cultural-policy-as-a-business-model/>

colleagues Manjusha Kuruppath for the input information to the VOC case study and Teresa Paccosi for valuable discussions about interpretations and translations, and the reviewers for their insightful comments and constructive criticism.

Author Contributions

Author contributions (by author initials) are listed according to the Contributor Roles Taxonomy (CRediT). Conceptualization: VZ, AB, JZ, MvE; Formal Analysis: VZ; Methodology: VZ, AB, JZ, MvE; Project administration: AB, MvE; Visualization: VZ, MvE; Writing (original draft): VZ, JZ; Writing (review and editing): VZ, JZ, MvE, AB

References

- Melanie Becker. 2019. [Privacy in the digital age: comparing and contrasting individual versus social approaches towards privacy](#). *Ethics and Information Technology*, 21(4):307–317.
- A. Betti. 2012. [Bolzano’s Universe: Truth, Logic and Metaphysics](#), pages 167–190. Oxford Univ. press.
- Leonard Blussé. 1986. *Strange Company: Chinese Settlers, Mestizo Women and the Dutch in VOC Batavia*. KITLV Press, Leiden.
- Bernard Bolzano. 1837. *Dr. B. Bolzanos Wissenschaftslehre: Versuch einer ausführlichen und größtentheils neuen Darstellung der Logik mit steter Rücksicht auf deren bisherige Bearbeiter*. Zweiter Band, volume 2. in der JE v. Seidelschen Buchhandlung.
- Pieter C. Emmer. 2003. *The Dutch in the Atlantic Economy, 1580–1880: Trade, Slavery and Emancipation*. Ashgate, Aldershot.
- Femme Gaastra. 2003. *The Dutch East India Company: Expansion and Decline*. Walburg Pers, Zutphen.
- Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. 2002. Sweetening ontologies with dolce. In *International conference on knowledge engineering and knowledge management*, pages 166–181. Springer.
- Emiliano Giovannetti, Davide Albanesi, Andrea Bellandi, Simone Marchi, Mafalda Papini, and Flavia Sciolette. 2024. [Maia: an open collaborative platform for text annotation, e-lexicography, and lexical linking](#). *Umanistica Digitale*, 8(18):27–52.
- Dagmar Gromann, Elena S. Apostol, Christian Chiarcos, Marco Cremaschi, Jorge Gracia, Katerina Gkirtzou, Chaya Liebeskind, Liudmila Mockiene, Michael Rosner, Ineke Schuurman, Gilles Sérasset, Purificação Silvano, Blerina Spahiu, Ciprian O. Truica, Andrius Utka, and Giedre V. Oleskeviciene. 2024. [Multilinguality and lloD: A survey across linguistic description levels](#). *Semantic Web*, 1.
- Lisa Holmes. 2022. [Generational attitudes and actions around data privacy](#). *Euromonitor International*. Data from Euromonitor International’s annual Voice of the Consumer: Digital Survey.
- Silvan Hungerbühler. 2018. A computational method for philosophical interpretation. Master’s thesis, University of Amsterdam.
- Jonathan Israel. 1989. *Dutch Primacy in World Trade, 1585–1740*. Clarendon Press, Oxford.
- Anas F. Khan, Christian Chiarcos, Thierry Declerck, Daniela Gifu, Elena González Blanco García, Jorge Gracia, Maxim Ionov, Penny Labropoulou, Francesco Mambrini, John P. Mccrae, Émilie Pagé-Perron, Marco Passarotti, Salvador R. Muñoz, and Ciprian O. Truica. 2022. [When linguistics meets web technologies. recent advances in modelling linguistic linked data](#). *Semantic Web*, 13:987–1050.
- John P Mccrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. [The ontolex-lemon model: development and applications](#). In *Proceedings of eLex 2017 conference*.
- Erik Odegard. 2020. *The Company Fortress: Military Engineering and the Dutch East India Company in South Asia, 1638-1795*. Leiden University Press.
- Yvette Oortwijn, Hein van den Berg, and Arianna Betti. 2021. [Ground truths for the humanities](#). *Preprint*, arXiv:2103.12841.
- Davide Picca, Alfio Gliozzo, and Aldo Gangemi. 2008. [Lmm: an owl-dl metamodel to represent heterogeneous lexical knowledge](#). In *Proc. of the Int. Conference on Language Resources and Evaluation*.
- Alicia Schrikker. 2007. *Dutch and British Colonial Intervention in Sri Lanka, 1780–1815: Expansion and Reform*, volume 3 of EMI (European Expansion and Indigenous Response). Brill, Leiden.
- Philip Stern. 2011. *The Company-State: Corporate Sovereignty and the Early Modern Foundations of the British Empire in India*. Oxford University Press.
- Sabine Tittel. 2023. Ceci n’est pas un dictionnaire. adding and extending lexicographical data of medieval romance languages to and through a multilingual lexico-ontological project. *Electronic lexicography in the 21st century*, pages 39–52.
- Marieke van Erp. 2023. [Unflattening knowledge graphs](#). In *Proceedings of the 12th Knowledge Capture Conference 2023, K-CAP ’23*, page 223–224, New York, NY, USA. Association for Computing Machinery.
- Chantal van Son, Tommaso Caselli, Antske Fokkens, Isa Maks, Roser Morante, Lora Aroyo, and Piek Vossen. 2016. [Grasp: A multilayered annotation scheme for perspectives](#). In *Tenth Int. Conference on Language Resources and Evaluation*, pages 1177–1184.
- Arthur Weststeijn. 2014. [The voc as a company-state: Debating seventeenth-century dutch colonial expansion](#). *Itinerario*, 38(1):13–34.