Journal of Digital Terminology and Lexicography
Vol. 1, No. 1 (2025) DOI: 10.25430/pupj.jdtl.1747980096

The Art of Modelling Terminology Resources with Ontolex-lemon and Semantic Web Standards: the TERMCAT Usecase

Paula Diez-Ibarbia[1*], Patricia Martín-Chozas[1†] and Elena Montiel-Ponsoda[1†]

[1*]Ontology Engineering Group, Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla Del Monte, 28660, Madrid, Spain.

*Corresponding author(s). E-mail(s): paula.diez@upm.es;
Contributing authors: patricia.martin@upm.es; elena.montiel@upm.es;
†These authors contributed equally to this work.

Abstract

Despite the constant growth of language resources, there remains a lack of terminological and lexicographic data available in interoperable formats like RDF. To palliate this deficiency, efforts must focus on developing guidelines for modelling and sharing such data in formats that facilitate interoperability and seamless integration across platforms and applications. To achieve this objective, this study analyses the modelling requirements of the glossaries published by the Catalan Terminology Centre (TERMCAT) for their transformation into RDF, following, among others, Ontolex-lemon, the *de facto* standard for the modelling of lexicographic resources in RDF. In this contribution, we explore extensions and adaptations of previous modelling strategies to accommodate the specific requirements of the TERMCAT glossaries, discuss the modelling issues encountered in the process, and propose alternative modelling options.

Keywords: Terminology, Ontolex-lemon, Linked Data, TERMCAT

1 Introduction

The advantages of converting language resources into the Resource Description Framework (RDF) format are widely acknowledged. This Semantic Web standard, developed by the World Wide Web Consortium (W3C), is designed to enable data exchange by semantically defining relationships between data elements. RDF is the basis of the Linked Data paradigm,[1] which enables the integration, sharing, and reuse of structured data across different systems and domains. The Language Resources community has actively embraced this paradigm, as evidenced by the Linguistic Linked Open Data (LLOD) Cloud initiative.[2] Standard vocabularies and models, such as SKOS[3] (Miles, Matthews, Wilson, & Brickley, 2005) and Ontolex-lemon[4] (J. McCrae et al., 2012) (henceforth, Ontolex), have facilitated the adoption of RDF for the transformation of thesauri, terminology resources, lexicons, and dictionaries into interoperable data formats.

[1] https://www.w3.org/DesignIssues/LinkedData.html
[2] https://linguistic-lod.org/llod-cloud-jan2011
[3] https://www.w3.org/2004/02/skos/
[4] https://www.w3.org/2016/05/ontolex/

Copyright ©2025 The Author(s)

Despite these advancements, existing models do not always meet the specific representation requirements of every language resource. Structural differences may occur even between resources within the same typological category, posing a challenge for language professionals when transforming their resources to Semantic Web standards.

This need is contemplated in the national project TeresIA,[5] that aims to provide an access point to linked terminology resources in Spain, and IA services for terminology management. In this project, one of the objectives is to build an automatic converter adapted to the most used data schemas to model terminology resources that help language professionals with little or no Semantic Web knowledge model their resources.

Therefore, this paper specifically focusses on the modelling of terminology resources. To develop this automatic converter, we have selected a set of authoritative and well-known terminology resources at the national and European levels as representative examples. In this work, we analyse the challenges encountered when converting into RDF the resources produced by the TERMCAT Terminology Centre, which is a highly relevant institution at a national level. In addition, we propose a methodological approach for modelling the data found in these terminology resources. By examining TERMCAT's openly available resources, covering different languages and domains, we assess the capacity of Ontolex and other Semantic Web standards to accommodate the specific requirements of terminology modelling.

2 Related Work

The community of researchers dedicated to publishing lexicographic and terminology resources in Semantic Web formats is small but well-established. As a result, although the related body of literature is not extensive, it includes several significant and influential works. In this section, we review practical initiatives, such as conversion experiments and tools, as well as theoretical analysis of standards for representing lexicographic and terminological data, including recognised vocabularies and proposals.

In terms of lexicographic resources, one of the most important works is the conversion of the English lexicon WordNet (J. McCrae, Fellbaum, & Cimiano, 2014) to the *lemon* model (the predecessor of the Ontolex model). A similar work was the publication of the Apertium dictionaries (Gracia, Villegas, Gómez-Pérez, & Bel, 2018) following the same vocabulary. In the same line, the series of multilingual KDictionaries were transformed (Bosque Gil, Lonke, Gracia del Río, & Kernerman, 2019) taking the Ontolex-lemon as a reference (J.P. McCrae, Bosque-Gil, Gracia, Buitelaar, & Cimiano, 2017). Another relevant work in this area is the conversion of the Diccionario da Lingua Portugueza, which studies the equivalences between TEI Lex-0 encoding and Ontolex (Almeida et al., 2022). Finally, it is worth mentioning the work reported in (Bosque-Gil, Gracia, Montiel-Ponsoda, & Gómez-Pérez, 2018), which serves as a key reference document for researchers in the field, as it provides a comprehensive survey of various existing vocabularies for the modelling of lexicographic resources according to Semantic Web standards.

Regarding the conversion of terminology resources, one of the main areas of research has been the conversion of resources structured in TBX, an ISO standard for terminology information exchange (Melby, 2015), to Semantic Web standards. Several conversion efforts have been undertaken in this vein, including the work presented in Cimiano et al. (2015), which transforms a simplified version of the well-known InterActive Terminology for Europe (IATE) term base and the European Migration Network (EMN) glossary to the *lemon* format. This work also introduces a tool for transforming TBX into RDF.[6] A similar work is described in di Buono, Cimiano, Elahi, and Grimm (2020), which also proposes the conversion of IATE and other term bases hosted by the GENTERM centre,[7] making use of Ontolex and related vocabularies. This work also relies on the Terme-à-LLOD service,[8] a conversion tool from TBX to Ontolex, which additionally supports the hosting and browsing of the converted data, and offers a SPARQL endpoint.

In this regard, the conversion of TBX resources remains an active area of research. One of the latest studies delves deeply into the specifications of TBX and Ontolex models to identify their needs and requirements (Bellandi, Di Nunzio, Piccini, & Vezzani, 2023), with the aim of building an automatic converter that is subsequently presented in Bellandi, Di Nunzio, Piccini, and Vezzani (2024). More tools dealing with language resources in RDF are found in the literature, such as VocBench (Stellato et al., 2020), which

[5] https://proyectoteresia.org/
[6] http://tbx2rdf.lider-project.eu/converter/
[7] https://cvt.ugent.be/downloads.htm
[8] https://github.com/ag-sc/terme-a-llod

is a well-established tool to collaboratively model language resources supported by the Publications Office of the European Union.[9] Finally, one of the most recent tools is LexO, a collaborative web-based editor for creating and managing lexical and terminological resources based on the OntoLex model (Bellandi, 2021). This tool is particularly beneficial for non-expert users, as it requires no technical expertise, thereby facilitating broader adoption of these standards.

Finally, our research is supported by previous efforts towards the conversion of TERMCAT terminology resources (Bosque-Gil, Montiel-Ponsoda, Gracia, & Aguado-De-Cea, 2016). In this work, the TerminotecaRDF portal was proposed as a gathering point for multilingual terminology resources in Spain, which also included the conversion of Terminesp[10] to Ontolex. Taking this lead, we have followed a similar methodology to adapt TERMCAT terminology resources to the current Ontolex specification.

3 TERMCAT terminology resources

To identify the potential modelling needs of terminology resources, a set of terminology resources was analysed, which are published on the Terminologia Oberta platform (open terminology platform)[11] of the TERMCAT. The terminology resources in this collection are available in three formats: XML, HTML, and PDF. For this use case, over 150 terminology resources in XML were examined, covering a wide range of domains such as science, gastronomy, and tourism, to mention but a few.

The TERMCAT XML terminology resources share a homogeneous structure with respect to the nodes and attributes in which information is organised, which simplifies automatisation processes. As shown in Figure 1, the root element (`cessiodades`) is a node that contains three subnodes that provide information about the resource:

1. The node `autor` (‘author’) informs about the author of the terminology resource, which tends to be ‘TERMCAT, Centre de Terminologia’.
2. The node `titol` (‘title’) provides the name of the resource, such as ‘Diccionari d’atletisme’ (Dictionary of Athletics).
3. The node `fitxes` (‘cards’) groups all the concepts of the resource, along with the information related to those concepts (e.g., definitions, terms, etc.).

As shown in Figure 1, while the nodes `autor` and `titol` do not have further subnodes; inside the node `fitxes`, multiple subnodes named `fitxa` (‘card’) can be found. Each of those `fitxa` subnodes is used to represent a concept, which is identified with a numeric ID through the attribute `num` (‘number’). Additionally, each `fitxa` (‘card’) node can have four subnodes, two out of which are always present (`areatematica` or domain, and `denominacio` or designation) and the other two are optional (`definicio` or definition and `nota` or note).

The node `areatematica` (‘domain’ or ‘thematic area’) informs about the domain of the concept and may appear several times in the same `fitxa` node (i.e., concept). On the other hand, the node `denominacio`, which can also appear multiple times within a `fitxa`, is used to represent a single term. To describe the term, the node `denominacio` takes four different attributes:

1. The attribute `llengua` (‘language’) identifies the language of the term. In the analysis conducted, 46 unique values were found for this attribute. Although Catalan is the predominant language in the different glossaries, the presence of Spanish and English is also strong. In addition, the resources include languages from various territories and countries, such as Portuguese, French, Italian, Chinese, and Japanese. Notably, the presence of minority languages, such as Basque, Galician, and Welsh, was also observed. In some resources, such as the *Diccionari de l’activitat parlamentària*[12] (Dictionary of Parliamentary Activity), terms in Catalan Sign Language were found. However, occasionally the attribute `llengua` is not used to designate a language, but to indicate that the terms are language independent, such as symbols, formulas, codes (without any further specification), CAS numbers,[13] or even authors.

[9] https://op.europa.eu/en/
[10] https://www.wikilengua.org/index.php/Wikilengua:Terminesp
[11] https://www.termcat.cat/ca/terminologia-oberta
[12] https://www.termcat.cat/ca/diccionaris-en-linia/289
[13] A specific type of code used in chemistry: https://www.cas.org/cas-data/cas-registry

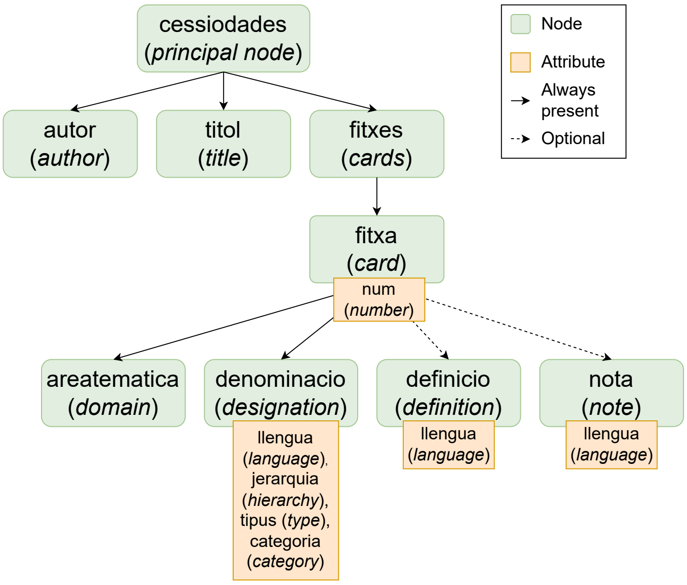

Fig. 1 XML structure of TERMCAT Terminologia Oberta terminology resources

2. The attribute `tipus` ('type') can take three values: *principal*, *equivalent*, and *remissio*. Although no information was found about this feature within the TERMCAT documentation,[14] these three values appear to indicate the recommended use. They are consistent across all resources and each concept appears to have only one 'principal' value, usually a Catalan term. Additionally, 'remissio' means 'remission' which suggests that the attribute `tipus` is related to the frequency of use of the term.
3. The attribute `jerarquia` ('hierarchy') can take 8 unique values to represent the relation between the terms under the same concept: *terme pral.* ('principal term'), *abrev.* ('abbreviation'), *sigla* ('initialism'), *den. com.* ('commercial designation'), *den. desest.* ('dismissed designation'), *sin. compl.* ('complementary synonym'), *alt. sin.* ('alternative synonym'), and *var. ling.* ('linguistic variant'). It should be also noted that according to TERMCAT, Centre de Terminologia (2022a) a 'principal term' is a term that is adequate in all contexts. In other words, it can be considered an absolute synonym. On the other hand, a 'linguistic variant' consists of a form that differs from another solely in spelling, while maintaining identical pronunciation, such as 'water-resistant' and 'water resistant', or 'druggability' and 'drugability' (TERMCAT, Centre de Terminologia, 2022c). As for 'complementary synonyms', these synonyms refer to secondary terms that are adequate but have a more restricted validity. Lastly, 'alternative synonyms' are unrecommended documented forms (TERMCAT, Centre de Terminologia, 2022a).
4. The attribute `categoria` ('category') stores information about the part-of-speech (noun, adjective, interjection, locution, etc.). In addition to part-of-speech information, other types of grammatical features may also be provided. For example, a noun may be accompanied by details about its grammatical number (e.g. plural) and/or grammatical gender (masculine, feminine, and neuter). Similarly, verbs may include information about their valency (transitive/intransitive). They can also be labelled as prepositional or pronominal verbs. Furthermore, this attribute can also be used to indicate that the term is not a full form, but a prefix or a suffix.

Lastly, regarding the optional subnodes of `fitxa`, the node `definicio` ('definition') provides the definition of the concept, while the node `nota` ('note') contains notes about the use or origin of the concept. Both nodes have an attribute named `llengua` ('language') to indicate the language of the definition or note. Therefore, although the original structure of the resource may seem simple, the analysis of the data across different glossaries raised challenging decisions that needed to be carefully examined when modelling the data into Ontolex.

[14] https://www.termcat.cat/es/recursos/criteris

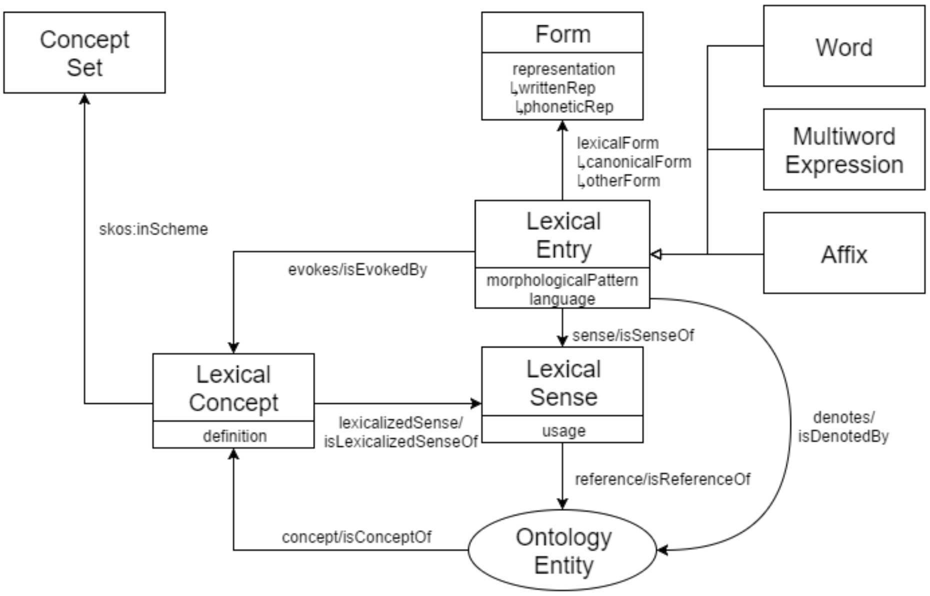

Fig. 2 Ontolex Core Diagram

4 Implemented Ontologies

Despite Ontolex being originally intended to add lexical information to ontologies, and subsequently being used to model lexicographical resources, we have selected this vocabulary to represent TERMCAT data following the guidelines of previous work on the topic (Cimiano et al. 2015; di Buono et al. 2020), Since this model has come to be widely accepted as the standard for representing language resources as Linked Data. As shown in Figure 2, Ontolex revolves around four main classes: Lexical Entry, Form, Lexical Concept, and Lexical Sense. The class Lexical Entry is used to represent a word, phrase, or lexical unit in a specific language. The different morphological realisations of the entry are expressed through the class Form. Regarding Lexical Sense, it is used to provide a semantic connection between the Lexical Entry class and a concept in the ontology. Lastly, a Lexical Concept represents an abstract concept or idea that unifies meanings across Lexical Senses and languages.

Since not all types of information in TERMCAT terminology resources could be covered by this model, complementary ontologies and models were used, as listed below.

1. LexInfo (Cimiano, Buitelaar, McCrae, & Sintek, 2011): a model that provides data categories to represent, for instance, grammatical information (part-of-speech, gender, and number) or to provide the normative authorisation of a term (such as ‘deprecated’).
2. Simple Knowledge Organisation System (SKOS):[15] a W3C standard, commonly used to represent hierarchical data in thesauri, classification systems and other types of organisation systems.
3. Vartrans:[16] an Ontolex complementary module that proposes a way to model semantic relationships such as translation or term variant.
4. Synsem:[17] an Ontolex complementary module used to represent the syntactic behaviour of certain entries, such as verbs.
5. Ontologies of Linguistic Annotation (OLiA):[18] a model focused on linguistic annotation such as the valency of verbs (e.g. transitive or intransitive).
6. Easy-tv (etv):[19] an Ontolex-based ontology for the representation of signed terms.

[15] https://www.w3.org/2009/08/skos-reference/skos.html
[16] http://www.w3.org/ns/lemon/vartrans#
[17] http://www.w3.org/ns/lemon/synsem#
[18] https://github.com/acoli-repo/olia
[19] https://w3id.org/def/easytv

7. Termlex:[20] (Martín-Chozas, Declerck, Montiel-Ponsoda, & Rodríguez-Doncel, 2024) a proposal based on Ontolex that aims to represent terminological data. Amongst other aspects, this proposal allows grouping the definitions and notes referred to a concept by language and source.
8. DBpedia Ontology (DBO):[21] a cross-domain model used to represent codes and CAS numbers.[22]
9. DCMI Metadata Terms:[23] a widely used vocabulary for describing metadata about resources, such as documents, images, datasets, and any other kind of digital or physical entity. For example, it allows representing data such as the title or language of an element. This ontology is also known as Dublin Core Terms (DCTerms).

5 Modelling Issues

The purpose of this section is to present the main issues encountered in the modelling of TERMCAT terminology resources. Section 5.1 discusses the difficulties involved in symbol modelling. Then, Section 5.2 presents the approach followed for the modelling of codes and CAS numbers. After that, Section 5.3 analyses the representation of specific subfields across terminological entries. In Section 5.4, we focus on the representation of authorship in sea mammal glossaries. Section 5.5 addresses the representation of forms of different genders, and, lastly, Section 5.6 centres on the modelling of prepositional and pronominal verbs.

5.1 Symbols

As outlined in Section 3, TERMCAT resources indicate the language of a term through the attribute `llengua` in the node `denominacio`. However, in certain cases, this attribute contains values that do not correspond to any natural language. For example, the value ‘sbl’ is used to denote symbols. According to TERMCAT, Centre de Terminologia (2022b), a symbol is a graphical representation composed of elements of either the same or different nature (such as alphabetic characters, numerals, superscripts, or punctuation marks), which is conventionally assigned a specific meaning. Notably, symbols are valid across multiple languages. Examples of terms classified as symbols include:

1. Φ: In the area of telecommunications,[24] it can stand for three different concepts: i) magnetic flux, ii) electric potential or scalar potential, and iii) electrostatic potential. However, if the focus is shifted to chemistry,[25] this symbol can be used to represent a couple of concepts: i) gravitational chemical potential and ii) centrifugal chemical potential.
2. K-2: in water sports, it can represent the concept denoted by kayak doubles, kayak pair or tandem.[26]
3. °C: the measurement unit for temperature ‘degree Celsius’.[27]
4. Au: in chemistry, it represents the chemical element gold.[28]
5. Rad: it can stand for the measurement ‘radiant’.[29]
6. IFN: in the area of health, IFN can stand for ‘interferon’, a type of protein.[30]
7. F: this character can take several meanings. To begin with, in chemistry, it can stand for the chemical element ‘fluorine’. Moreover, it can also be used to refer to the unit of electrical capacitance ‘farad’.[31] Additionally, F can also be used to indicate the ‘noise factor’ in telecommunications.[32]

The examples show that TERMCAT symbols can vary in their forms. At first glance, some of the terms could be regarded as an initialism or abbreviation, such as the term ‘Rad’, which could be represented with LexInfo instances (`lexinfo:initialism` and `ontolex:abbreviation`, respectively). Nonetheless, TERMCAT, Centre de Terminologia (2022b) distinguishes between those type of terms and symbols;

[20] https://termlex.oeg.fi.upm.es/
[21] http://dbpedia.org/ontology/
[22] https://www.cas.org/es-es/cas-data/cas-registry
[23] http://purl.org/dc/terms/
[24] https://www.termcat.cat/Thor/files/diccionaris/cadfdltelecomunicacions.xml
[25] https://www.termcat.cat/Thor/files/diccionaris/cadfdlquimicaqoqiqfqaeq.xml
[26] https://www.termcat.cat/Thor/files/diccionaris/cadfdlesport2025.xml
[27] https://www.termcat.cat/Thor/files/diccionaris/cadfdlfisica2aed.xml
[28] https://www.termcat.cat/Thor/files/diccionaris/cadfdlquimicaqoqiqfqaeq.xml
[29] https://www.termcat.cat/Thor/files/diccionaris/cadfdlfisica2aed.xml
[30] https://www.termcat.cat/Thor/files/diccionaris/cadfdlcovid19.xml
[31] https://www.termcat.cat/Thor/files/diccionaris/cadfdltelecomunicacions.xml
[32] https://www.termcat.cat/Thor/files/diccionaris/cadfdltelecomunicacions.xml

consequently, labelling symbols as abbreviations or initialisms seemed to be an unfaithful representation of the original data. In fact, TERMCAT, Centre de Terminologia (2022b) provides three points to distinguish between these three types of terms:

1. Abbreviations that do not include a full stop to indicate truncation are classified as symbols.
2. Initialisms are considered symbols if they incorporate lowercase letters where only uppercase letters would typically be expected.
3. An abbreviation or initialism is also classified as a symbol if it adheres to a specific structural pattern (e.g. the presence of non-existent characters in the designation or a non-canonical shortening) or if its international validity has been confirmed.

Due to the distinction between these three types of terms, the options of initialism and abbreviation were discarded for the modelling of terms labelled as 'sbl'. As an alternative, the use of the instance `lexinfo:symbol` was suggested, which is described as a "character or glyph representing an idea, concept or object".[33] Therefore, LexInfo's Symbol instance could be suitable for terms such as 'Φ' or 'F'. Nevertheless, it could be considered an inadequate way of representing other terms such as 'Au', 'K-2' or 'IFN'. The use of `lexinfo:internationalScientificName` was also taken into account for the representation of this phenomenon since most of the symbol terms appear to be from the scientific domain (chemistry, health, physics, maths…). Nonetheless, it could be argued that the use of certain symbols can be local and not international. Moreover, the symbols could belong to domains that may not fit in the scope of science, such as sports. Alternatively, the OLiA class Symbol was suggested (`olia:Symbol`). This class is defined as "a single graphical sign that occurs in a written text with a conventionalized meaning but that does not represent a phoneme (like ordinary characters), an orthographic sign (punctuation) or a number".[34] Even though this description could encompass most TERMCAT symbols, a few could fall outside the scope of this classification, such as 'Rad', for instance, which constitutes a phoneme. Therefore, taking all these considerations into account, the final proposal for modelling these terms is based on the use of DBpedia Categories, in particular, the class `dbc:Symbols`.[35] This class lacks a formal description, yet it appears to be generic enough to encompass all the terms, as it is considered to be broader than other concepts such as national symbols, consumer symbols, heart symbols, flags, pythagorean symbols, or diacritics.

Ideally, when representing TERMCAT symbols, each term should be manually and individually analysed to determine the most appropriate representation. However, the modelling work detailed in this paper is intended to be used for an automated transformation of the TERMCAT resources; therefore, a general modelling approach needs to be settled for all the cases. For this reason, although `dbc:Symbols` may seem generic, we came to the conclusion that this is the most suitable option to accommodate the different types of symbols contained in these resources.

Once the representation of a symbol has been determined, it is necessary to model its relationship with other terms provided for the same concept. Although such relationships are implicit in the original resources, this work considers making them explicit to traverse the resulting graph with simpler and more straightforward queries. Usually, terms pointing to a shared concept and language would be considered synonyms, while the ones with a shared concept but a different language would be regarded as translations. However, some previous studies in terminology recognise a synonymy (or term variation) relation between a term (e.g., Spanish 'grados centígrados' or English 'degree Celsius') and the symbol (such as °C) that represents it (Cabré, 1999), instead of a translation one. For this reason, the representation displayed in Figure 3 was proposed.

Although symbols may not always be universally recognised, TERMCAT, Centre de Terminologia (2022b) stipulates that symbols must be valid in all languages of the file. Therefore, this work assumes that the languages encompassed by the concept were contemplated when the symbol term was included. If there are concerns regarding the consideration of these languages, the synonymy relation could be restricted to Catalan, as it is the reference language in these resources.

A different option was considered when modelling the semantic relations of symbols, which is based on a direct relation between the symbol and the Lexical Concept. This connection can be established with the properties available in the Semiotics[36] ontology (see Figure 4). However, as previously stated, this

[33] http://www.lexinfo.net/ontology/3.0/lexinfo#
[34] http://purl.org/olia/olia.owl#
[35] https://dbpedia.org/page/Category:Symbols
[36] http://www.ontologydesignpatterns.org/cp/owl/semiotics.owl

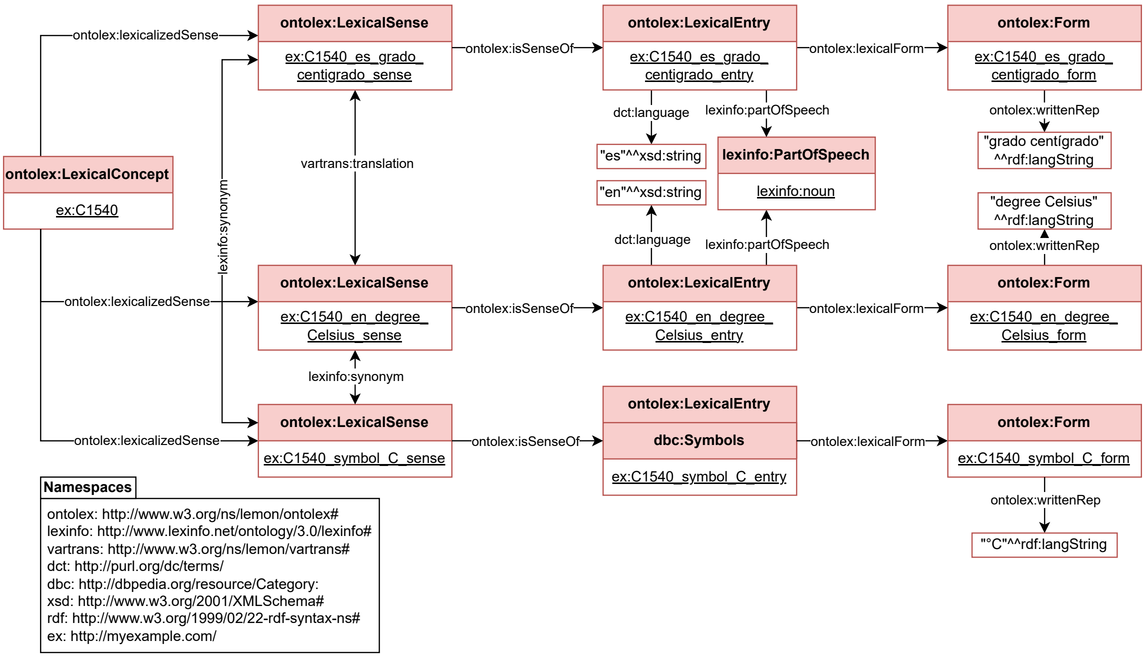
ontolex:LexicalConcept
ex:C1540
ontolex:lexicalizedSense
ontolex:LexicalSense
ex:C1540_es_grado_centigrado_sense
ontolex:isSenseOf
ontolex:LexicalEntry
ex:C1540_es_grado_centigrado_entry
ontolex:lexicalForm
ontolex:Form
ex:C1540_es_grado_centigrado_form
ontolex:writtenRep
"grado centígrado"^^rdf:langString
dct:language
"es"^^xsd:string
lexinfo:partOfSpeech
lexinfo:PartOfSpeech
lexinfo:noun
vartrans:translation
lexinfo:synonym
"en"^^xsd:string
"degree Celsius"^^rdf:langString
ontolex:LexicalSense
ex:C1540_en_degree_Celsius_sense
ontolex:LexicalEntry
ex:C1540_en_degree_Celsius_entry
ontolex:Form
ex:C1540_en_degree_Celsius_form
ontolex:LexicalSense
ex:C1540_symbol_C_sense
ontolex:LexicalEntry
dbc:Symbols
ex:C1540_symbol_C_entry
ontolex:Form
ex:C1540_symbol_C_form
"°C"^^rdf:langString
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
vartrans: http://www.w3.org/ns/lemon/vartrans#
dct: http://purl.org/dc/terms/
dbc: http://dbpedia.org/resource/Category:
xsd: http://www.w3.org/2001/XMLSchema#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 3 Modelling of symbols with Lexical Entries, example from *Diccionari de física*, TERMCAT, fitxa 1540

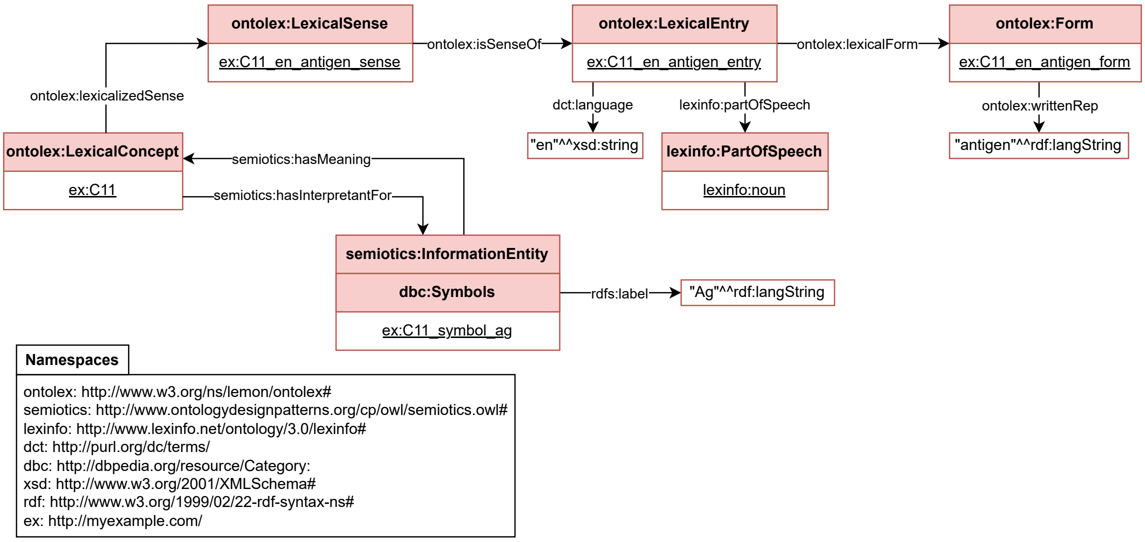
ontolex:LexicalSense
ex:C11_en_antigen_sense
ontolex:isSenseOf
ontolex:LexicalEntry
ex:C11_en_antigen_entry
ontolex:lexicalForm
ontolex:Form
ex:C11_en_antigen_form
ontolex:lexicalizedSense
dct:language
lexinfo:partOfSpeech
ontolex:writtenRep
ontolex:LexicalConcept
ex:C11
semiotics:hasMeaning
semiotics:hasInterpretantFor
"en"^^xsd:string
lexinfo:PartOfSpeech
lexinfo:noun
"antigen"^^rdf:langString
semiotics:InformationEntity
dbc:Symbols
ex:C11_symbol_ag
rdfs:label
"Ag"^^rdf:langString
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
semiotics: http://www.ontologydesignpatterns.org/cp/owl/semiotics.owl#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
dct: http://purl.org/dc/terms/
dbc: http://dbpedia.org/resource/Category:
xsd: http://www.w3.org/2001/XMLSchema#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 4 Modelling of symbols with semiotics, example from *Malalties metabòliques. Obesitat i diabetis*, TERMCAT, fitxa 11

work aims to respect and preserve the original structure of the TERMCAT resources; for this reason, since symbols in TERMCAT are presented at term level and not at concept level, the direct association to the Lexical Concept was discarded.

5.2 Codes and CAS Numbers

Apart from 'sbl', the values 'cod' and 'COD' can be found in the `llengua` attribute of the `denominacio` node. These values are used to indicate that the term is a code. Although no reference to codes was found in the TERMCAT documentation, a code can be defined as "a system of letters or digits used for identification or selection purposes" (Collins Dictionary, 2025). As for the representation of these terms, no instance of 'code' was found within the Ontolex, LexInfo, and OLiA ontologies. To address this issue, three different modelling options were proposed.

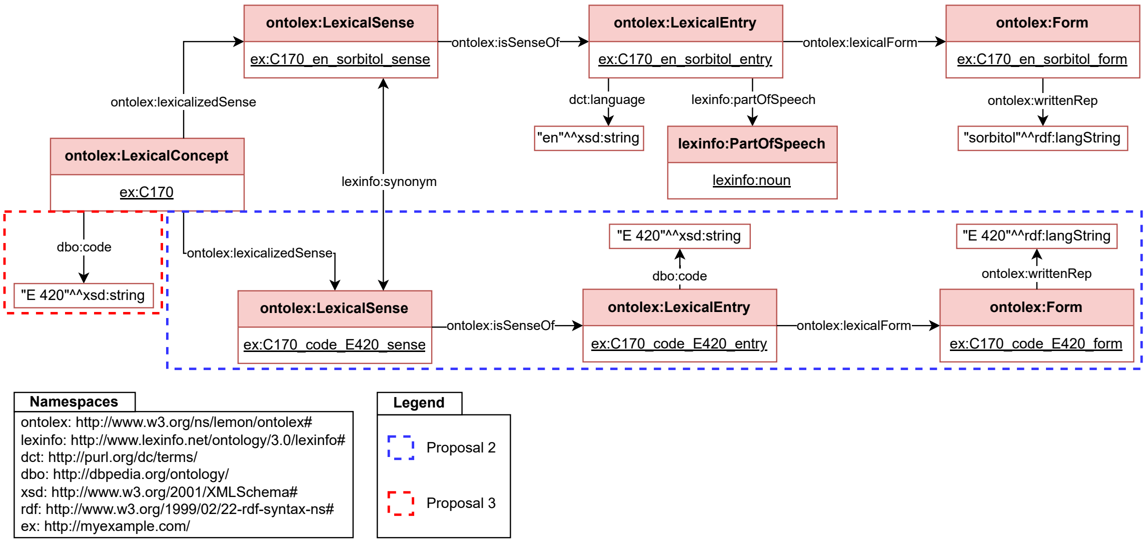

Fig. 5 Modelling of Proposal 2 and Proposal 3 for codes. Example from *Diccionari de malalties metabòliques. Obesitat i diabetis*, TERMCAT, fitxa 170

Proposal 1 initially suggested modelling codes as symbols, based on the definition provided by ISO/CD 704:2022 (2022), which states that "alphanumeric codes made up of combinations of letters, numbers or both shall be considered *symbols* if they do not represent words in a *natural language* or *abbreviations*". However, considering that TERMCAT already designates a distinct value for symbols ('sbl') and that terminologists explicitly refer to the terms in question as codes, Proposal 1 appears to be an inadequate representation. Consequently, this initial modelling approach was deemed unsuitable and subsequently discarded.

The remaining two modelling proposals are based on the code property found in the DBpedia Ontology (`dbo:code`). Proposal 2 advocated for maintaining the structure used for symbols. In other words, a Lexical Entry could be created, with the code property assigned to it as illustrated in Figure 5 (see Proposal 1 in blue). However, Proposal 3 claims that codes function as identifiers of a concept. Under this interpretation, the `dbo:code` property should be directly linked to the Lexical Concept, without the need to create a Lexical Entry or Form (see Proposal 3 in red in Figure 5). Since the original XML data considers codes as terms, Proposal 2 was ultimately adopted. In other words, a Lexical Entry and a Form are created, with the `dbo:code` property assigned to the Lexical Entry.

Finally, in addition to codes and symbols, TERMCAT can also use the `llengua` attribute in the `denominacio` node to introduce CAS numbers, which are identified by the value 'CAS' (see Listing 1). These numbers are assigned by the CAS Registry[37] and serve as unique identifiers for chemical substances. As such, they can be regarded as a specialised type of code. Consequently, the modelling of this information follows the schema established for codes (i.e., Proposal 2). However, rather than employing the general `dbo:code` property, it is recommended to use a more specific property, namely `dbo:casNumber`, as illustrated in Figure 6.

[37] https://www.cas.org/es-es/cas-data/cas-registry

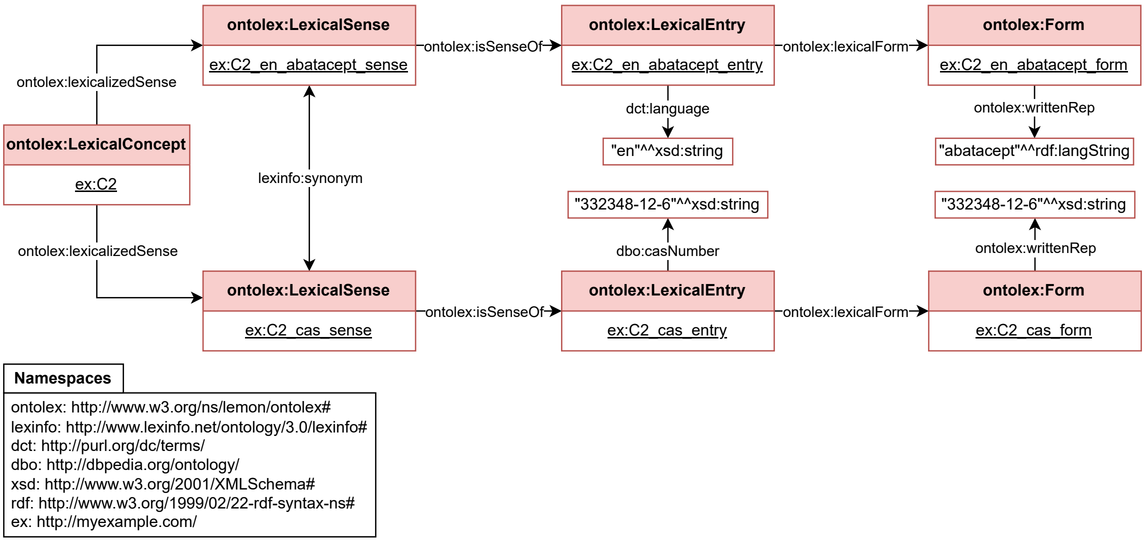

Fig. 6 Modelling of CAS Numbers, example from *Diccionari d'immunologia*, TERMCAT, fitxa 2

```
<denominacio
    llengua="en"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria=""
    ><![CDATA[abatacept]]>
</denominacio>
<denominacio
    llengua="CAS"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria=""
    ><![CDATA[332348-12-6]]>
</denominacio>
```

Listing 1 Example from *Diccionari d'immunologia*, TERMCAT, fitxa 2

5.3 Subfields in Terminological Entries

As mentioned in Section 5.1, the attribute `llengua` in the `denominacio` node is not always used to indicate the language of a term. Apart from identifying a term as a code or symbol, this attribute can also be used to specify the domain to which a term belongs. For instance, the value 'TA' designates anatomical terminology, while 'TH' refers to histological terminology[38] (TERMCAT, Centre de Terminologia, 2024). It is important to note that, in TERMCAT XML resources, the domain is typically indicated at the concept level within the `areatematica` node. For instance, in Listing 2, the domain of concept number 254 is specified as 'Otorrinolaringologia' (Otorhinolaryngology in English), which, by inference, applies to all terms associated with the same concept. However, in the case of the 'TA' value, it is assigned directly to a specific term rather than at the concept level.

[38] Histology is "the scientific study of the structure of tissue from plants, animals, and other living things" (Cambridge Dictionary, n.d.).

```
<fitxa
        num="254">
    <areatematica
        ><![CDATA[Otorrinolaringologia]]>
    </areatematica>
    <denominacio
          llengua="en"
          tipus="equivalent"
          jerarquia="terme pral."
          categoria=""
        ><![CDATA[pharyngeal tonsil]]>
    </denominacio>
    <denominacio
          llengua="TA"
          tipus="equivalent"
          jerarquia="terme pral."
          categoria=""
        ><![CDATA[tonsilla adenoidea]]>
    </denominacio>
    <denominacio
          llengua="TA"
          tipus="equivalent"
          jerarquia="terme pral."
          categoria=""
        ><![CDATA[tonsilla pharyngealis]]>
    </denominacio>
</fitxa>
```

Listing 2 Example from *Terminologia de ciències de la salut*, TERMCAT, fitxa 254

In terms of modelling, since both anatomy and histology fall within the domain of science and consist of Latinate terms that appear to be internationally standardised (e.g., abdomen), the use of `lexinfo:internationalScientificName` was suggested, an instance of the class `lexinfo:TermType`. However, this instance is overly generic, leading to a loss of significant information. Since this loss relates to the field of usage, it was proposed to incorporate the information as a domain specification. Therefore, the inclusion of an additional domain for terms was proposed, linked to the Lexical Sense through the `lexinfo:domain` (see Figure 7).

Regarding the representation of the anatomical and histological domains, DBpedia was identified as a suitable resource. Specifically, the instance `dbc:Anatomical_terminology` was proposed for terms associated with the 'TA' value (see Figure 7). However, no instance for 'histological terminology' was found. Consequently, a broader concept was selected, namely, `dbc:Histology`.

Lastly, the representation of the semantic relation between a term in a given language and a term with an additional subdomain was discussed. In particular, synonymy and translation were studied. Terms with additional subdomains could be regarded as part of the jargon used by the community of the term's domain. Although jargons and languages are not the same, jargons seem to be closer to languages than to symbols or codes. For this reason, the translation was chosen. In addition, some concepts may have two terms with specific subdomains. In these cases, a synonymy relation is established between terms with subdomains in common, as shown in Figure 7.

5.4 Authorship

Certain resources exhibit unique characteristics, such as the inclusion of authorship information, as in an terminology resource related to sea mammals.[39] This information is identified by the 'auct' value in the `llengua` attribute of the `denominacio` node (see Listing 3). The 'auct' value appears to be used to indicate data authorship, as the entries of this type contain a proper name and a year rather than a lexical term. As demonstrated in Listing 3, 'auct' entries are typically preceded by another entry specifying the scientific name of the corresponding animal. These scientific names are denoted by the 'nc' value in the `llengua` attribute of the `denominacio` node.

[39] https://www.termcat.cat/Thor/files/diccionaris/cdlmamifersmarins.xml

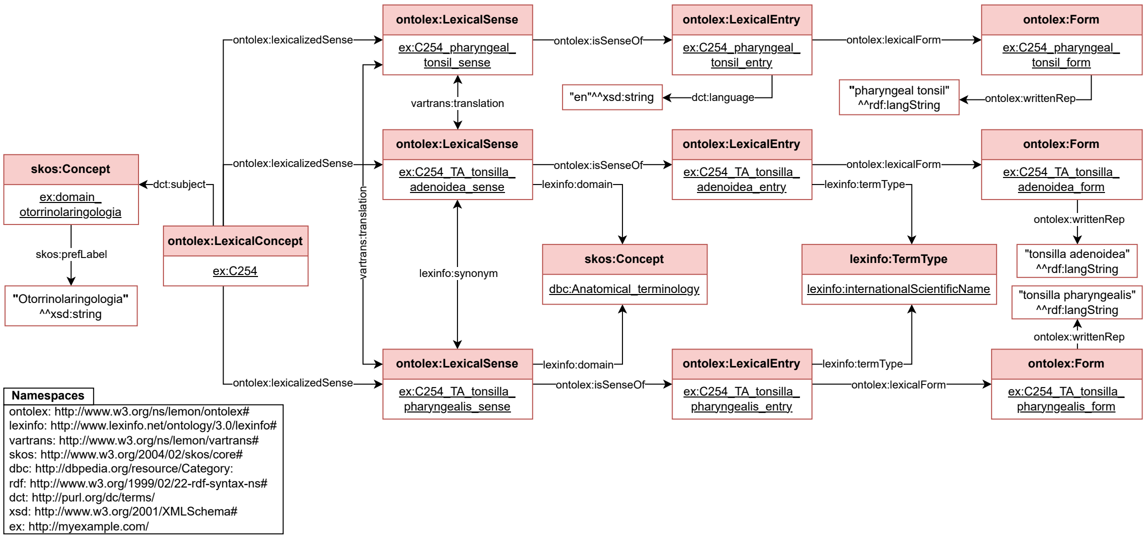
skos:Concept
ex:domain_otorrinolaringologia
skos:prefLabel
"Otorrinolaringologia" ^^xsd:string
dct:subject
ontolex:LexicalConcept
ex:C254
ontolex:lexicalizedSense
ontolex:lexicalizedSense
ontolex:lexicalizedSense
vartrans:translation
ontolex:LexicalSense
ex:C254_pharyngeal_tonsil_sense
vartrans:translation
ontolex:LexicalSense
ex:C254_TA_tonsilla_adenoidea_sense
lexinfo:synonym
ontolex:LexicalSense
ex:C254_TA_tonsilla_pharyngealis_sense
ontolex:isSenseOf
ontolex:isSenseOf
lexinfo:domain
lexinfo:domain
ontolex:isSenseOf
ontolex:LexicalEntry
ex:C254_pharyngeal_tonsil_entry
"en"^^xsd:string
dct:language
ontolex:LexicalEntry
ex:C254_TA_tonsilla_adenoidea_entry
skos:Concept
dbc:Anatomical_terminology
ontolex:LexicalEntry
ex:C254_TA_tonsilla_pharyngealis_entry
ontolex:lexicalForm
ontolex:lexicalForm
lexinfo:termType
lexinfo:TermType
lexinfo:internationalScientificName
lexinfo:termType
ontolex:lexicalForm
ontolex:Form
ex:C254_pharyngeal_tonsil_form
"pharyngeal tonsil" ^^rdf:langString
ontolex:writtenRep
ontolex:Form
ex:C254_TA_tonsilla_adenoidea_form
ontolex:writtenRep
"tonsilla adenoidea" ^^rdf:langString
"tonsilla pharyngealis" ^^rdf:langString
ontolex:writtenRep
ontolex:Form
ex:C254_TA_tonsilla_pharyngealis_form
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
vartrans: http://www.w3.org/ns/lemon/vartrans#
skos: http://www.w3.org/2004/02/skos/core#
dbc: http://dbpedia.org/resource/Category:
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
dct: http://purl.org/dc/terms/
xsd: http://www.w3.org/2001/XMLSchema#
ex: http://myexample.com/

Fig. 7 Modelling of anatomical terminology, example from *Terminologia de ciències de la salut*, TERMCAT, fitxa 254

```
<denominacio
    llengua="nc"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria=""
    ><![CDATA[<i>Balaenoptera musculus</i>]]>
</denominacio>
<denominacio
    llengua="auct"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria=""
    ><![CDATA[(Linnaeus 1758)]]>
</denominacio>
```

Listing 3 Example from *Noms de mamífers marins*, TERMCAT, fitxa 9

An effort was made to gain a deeper understanding of the information related to the author by searching for the animals by their scientific names. The technical data provided by various institutions suggest that in zoology the authorship information may be considered an integral part of the species name. In other words, the scientific name and the author information should be grouped together. This grouping can be observed in data provided by the Spanish Ministry of Environment (e.g. '*Balaenoptera musculus* (Linnaeus, 1758)') and the Global Biodiversity Information Facility (e.g. '*Lipotes vexillifer* Miller, 1918').

Based on the assumption that the author's name is part of the scientific or technical designation of the species, Proposal 1 in Figure 8 was suggested. In this representation, the scientific name and the authorship data are concatenated and modelled within the same Form. Furthermore, this Form is associated with an instance that specifies its scientific nature (`lexinfo:internationalScientificName`). However, since TERMCAT presents the scientific name and the author information in separate nodes, concerns were raised regarding the appropriateness of merging two distinct entries, as this approach might not faithfully reflect the original data structure.

Alternatively, the representation of authorship through provenance properties was suggested. Properties with the label 'author' were searched across several ontologies such as META-SHARE[40] or The Scientific Events Ontology (SEO).[41] However, the author properties were restricted to documents, which prevented their usage. In the end, a more generic property was selected: `dct:creator`. This property requires the use of a `dct:Agent` class. Consequently, Proposal 2 suggests to store the authorship information in a `dct:Agent` class, avoiding the creation of an `ontolex:LexicalEntry` (see Figure 9). However,

[40] http://www.meta-share.org/ontologies/meta-share/meta-share-ontology.owl/documentation/index-en.html#/author
[41] https://saidfathalla.github.io/SEOontology/Documentation/#Author

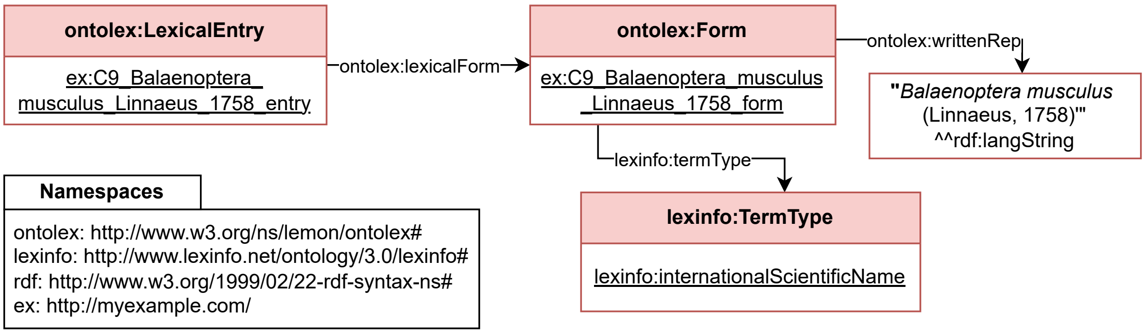
ontolex:LexicalEntry
ex:C9_Balaenoptera_musculus_Linnaeus_1758_entry
ontolex:lexicalForm
ontolex:Form
ex:C9_Balaenoptera_musculus_Linnaeus_1758_form
ontolex:writtenRep
"*Balaenoptera musculus* (Linnaeus, 1758)"^^rdf:langString
lexinfo:termType
lexinfo:TermType
lexinfo:internationalScientificName
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 8 Modelling of Proposal 1 for authorship representation, example from *Noms de mamífers marins*, TERMCAT, fitxa 9

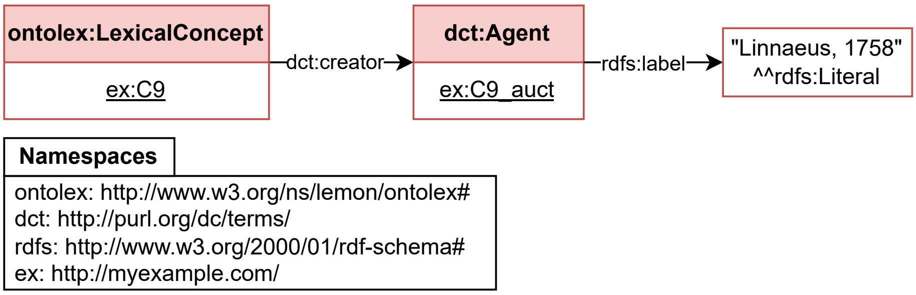
ontolex:LexicalConcept
ex:C9
dct:creator
dct:Agent
ex:C9_auct
rdfs:label
"Linnaeus, 1758"^^rdfs:Literal
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
dct: http://purl.org/dc/terms/
rdfs: http://www.w3.org/2000/01/rdf-schema#
ex: http://myexample.com/

Fig. 9 Modelling of Proposal 2 for authorship representation. Example from *Noms de mamífers marins*, TERMCAT, fitxa 9

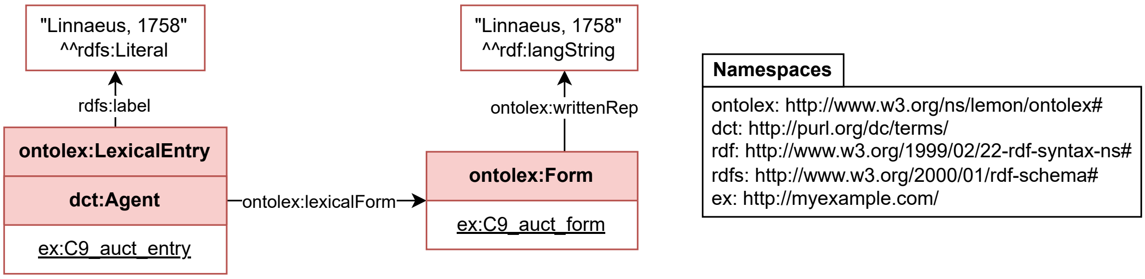
"Linnaeus, 1758"^^rdfs:Literal
rdfs:label
ontolex:LexicalEntry
dct:Agent
ex:C9_auct_entry
ontolex:lexicalForm
"Linnaeus, 1758"^^rdf:langString
ontolex:writtenRep
ontolex:Form
ex:C9_auct_form
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
dct: http://purl.org/dc/terms/
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
rdfs: http://www.w3.org/2000/01/rdf-schema#
ex: http://myexample.com/

Fig. 10 Modelling of Proposal 3 for authorship representation. Example from *Noms de mamífers marins*, TERMCAT, fitxa 9

as previously emphasised, this study seeks to maintain the highest fidelity to the original data. For this reason, a last proposal was suggested (Proposal 3) whereby a Lexical Entry (along with a Form and a Lexical Sense) is generated to represent authorship, as illustrated in Figure 10. Although this approach may not constitute a fully accurate terminological representation, it ensures a closer alignment with the original data.

After establishing the creation of a Lexical Entry (together with an Agent) for the authorship data, its semantic relations with the rest of the elements were explored, especially the placement of the property `dct:creator`. Proposal A considered the association between the Lexical Sense of the author and the Lexical Sense of the scientific term, as illustrated in Figure 11 (Proposal A, blue arrow). This would directly link the authorship data to the scientific term. Alternatively, according to Proposal B, the authorship information could be connected to the Lexical Concept (see Proposal B in Figure 11, red arrow). Lastly, Proposal C advocated for the absence of the creator property despite the loss of information. For now, this latter approach (Proposal C) has been chosen while further discussion takes place. Expert input would be highly valuable in finalizing this decision.

5.5 Grammatical Gender Representation

As previously mentioned in Section 4, TERMCAT may indicate the part-of-speech of a term through the `categoria` attribute in the `denominacio` node. With regard to nouns (denoted by 'n'), TERMCAT

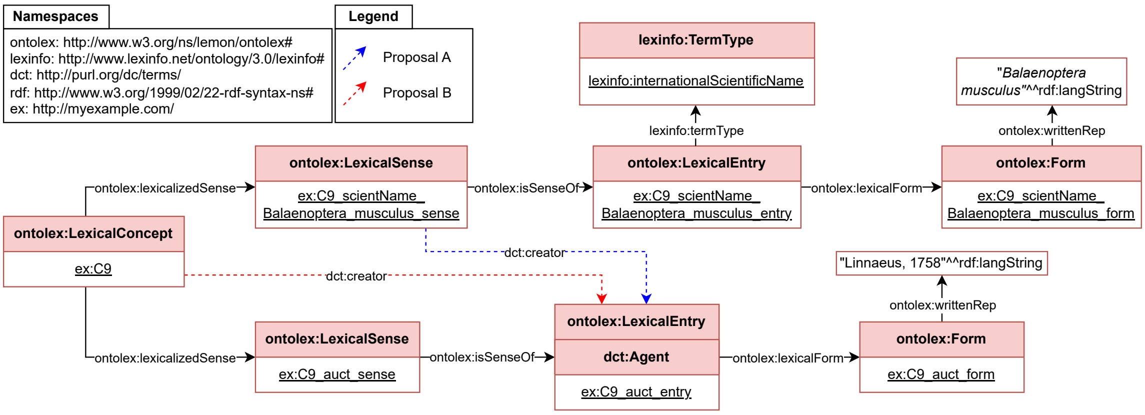
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
dct: http://purl.org/dc/terms/
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/
Legend
Proposal A
Proposal B
lexinfo:TermType
lexinfo:internationalScientificName
lexinfo:termType
"Balaenoptera musculus"^^rdf:langString
ontolex:writtenRep
ontolex:lexicalizedSense
ontolex:LexicalSense
ex:C9_scientName_Balaenoptera_musculus_sense
ontolex:isSenseOf
ontolex:LexicalEntry
ex:C9_scientName_Balaenoptera_musculus_entry
ontolex:lexicalForm
ontolex:Form
ex:C9_scientName_Balaenoptera_musculus_form
ontolex:LexicalConcept
ex:C9
dct:creator
dct:creator
"Linnaeus, 1758"^^rdf:langString
ontolex:writtenRep
ontolex:LexicalEntry
dct:Agent
ex:C9_auct_entry
ontolex:lexicalizedSense
ontolex:LexicalSense
ex:C9_auct_sense
ontolex:isSenseOf
ontolex:lexicalForm
ontolex:Form
ex:C9_auct_form

Fig. 11 Modelling proposals A and B for semantic relations in authorship representation. Example from *Noms de mamífers marins*, TERMCAT, fitxa 9

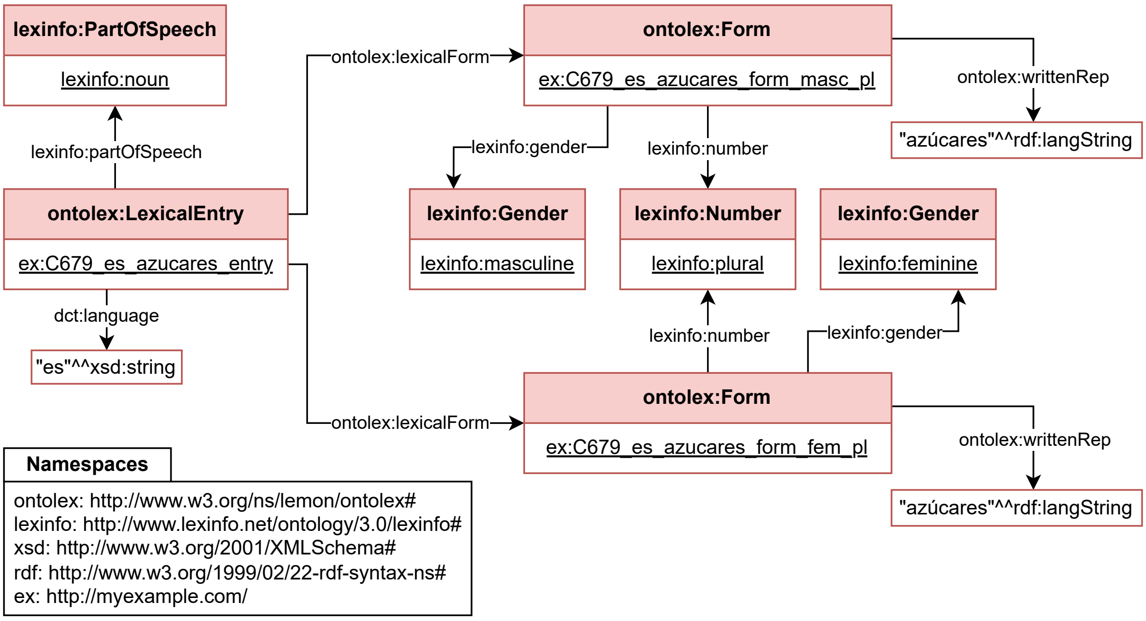

Fig. 12 Modelling of 'azúcares', example from *Diccionari de seguretat alimentària*, TERMCAT, fitxa 679

resources may also provide additional grammatical information, such as number and gender. Certain terms are inherently plural, as exemplified by 'reproductive rights' (TERMCAT, Centre de Terminologia, 2015–2024), a phenomenon denoted by the value 'pl' in TERMCAT. Furthermore, in languages such as Spanish, grammatical gender plays a significant role. In languages with grammatical genders, terms may have distinct forms depending on gender; for instance, the Spanish equivalent of 'teacher' can be 'maestro' (masculine) or 'maestra' (feminine). Although the representation of gender in terminology remains an open research question (Ralli & Evers, 2024), some TERMCAT resources provide multiple forms of terms according to gender. With regard to gender representation in Ontolex, the proposed approach involves the creation of multiple Ontolex Forms linked to a single Lexical Entry. However, TERMCAT introduced form variants in three different ways, which affect the representation with Ontolex.

To begin with, some TERMCAT terms contain a single word that is declared to be both masculine and feminine (see Listing 4); in other words, the masculine and feminine forms function as homonyms. Regarding the Ontolex representation, the creation of two Forms for the same Lexical Entry was suggested. Homonymous forms are duplicated and each `ontolex:Form` is assigned a distinct gender attribute, as illustrated in Figure 12.

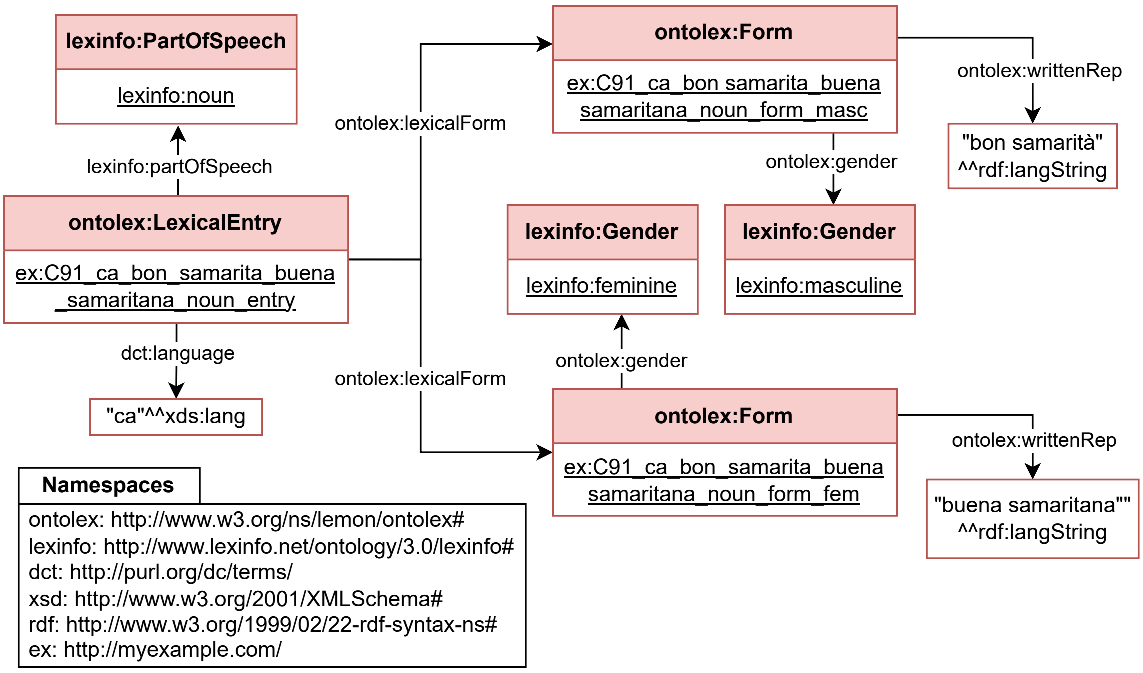
lexinfo:PartOfSpeech
lexinfo:noun
lexinfo:partOfSpeech
ontolex:LexicalEntry
ex:C91_ca_bon_samarita_buena _samaritana_noun_entry
dct:language
"ca"^^xds:lang
ontolex:lexicalForm
ontolex:lexicalForm
ontolex:Form
ex:C91_ca_bon_samarita_buena samaritana_noun_form_masc
ontolex:writtenRep
"bon samarità" ^^rdf:langString
ontolex:gender
lexinfo:Gender
lexinfo:feminine
lexinfo:Gender
lexinfo:masculine
ontolex:gender
ontolex:Form
ex:C91_ca_bon_samarita_buena samaritana_noun_form_fem
ontolex:writtenRep
"buena samaritana"" ^^rdf:langString
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
dct: http://purl.org/dc/terms/
xsd: http://www.w3.org/2001/XMLSchema#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 13 Modelling of 'bon samarità | bona samaritana', example from *Diccionari de bioètica,* TERMCAT, fitxa 91

```
<denominacio
    llengua="es"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria="n m pl/f pl">
    ><![CDATA[azúcares]]>
</denominacio>
```

Listing 4 Example from *Diccionari de seguretat alimentària,* TERMCAT, fitxa 679

Secondly, in the original resources, masculine and feminine forms can be presented by a separation through a vertical bar (|). For instance, 'bon samarità | bona samaritana' correspond to the masculine and feminine forms of 'good Samaritan' in Catalan (see Listing 5). In this case, preprocessing would be required to separate the two forms, using the bar (|) as reference. This way, two separated forms would be modelled, following the structure used previously in the representation of homonymous forms (see Figure 13), whereby two distinct Form classes are associated with the same Lexical Entry.

```
<denominacio
    llengua="ca"
    tipus="principal"
    jerarquia="terme pral."
    categoria="n m, f"

    ><![CDATA[bon samarità | bona samaritana]]>
</denominacio>
```

Listing 5 Example from *Diccionari de bioètica,* TERMCAT, fitxa 91

Thirdly, TERMCAT resources can introduce the feminine form with a suffix following the complete masculine form (see Listing 6). Taking the Catalan entry 'amfitrió -iona' ('host -ess' in English) as an example, the feminine form 'amfitriona' can be derived by applying the feminine suffix to the masculine form. Proposal 1 suggested following previous modelling structures and creating individual forms (see Figure 12). This proposal implies the automatic generation of the feminine form, which may introduce errors. To avoid word formation (Proposal 1), two other proposals were suggested: single-string representation (Proposal 2), and morphological representation (Proposal 3).

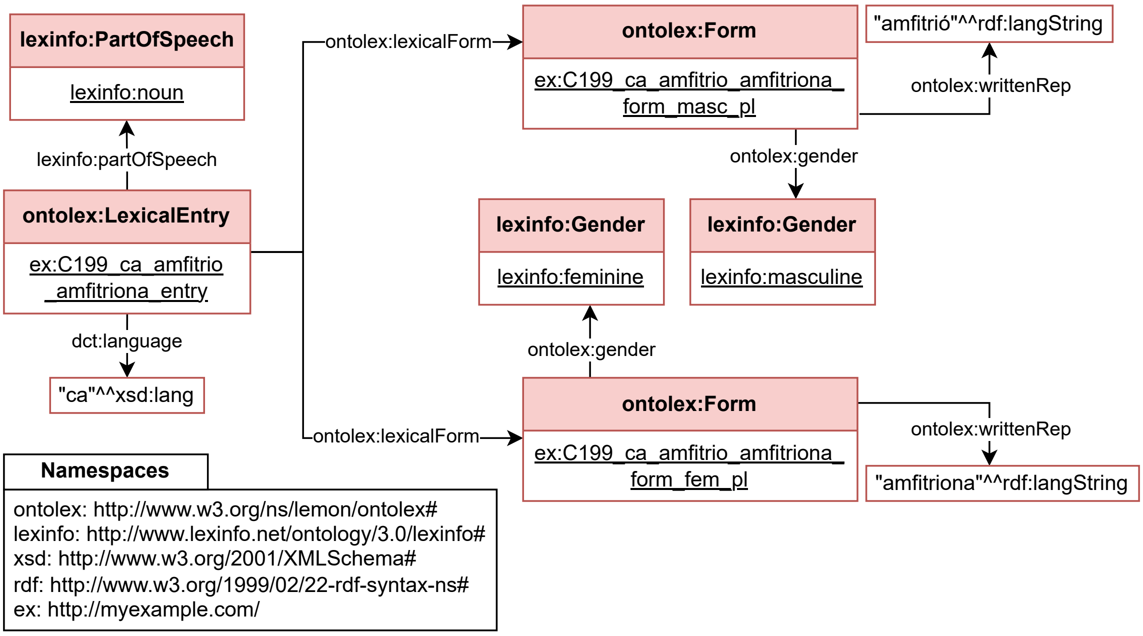
lexinfo:PartOfSpeech
lexinfo:noun
lexinfo:partOfSpeech
ontolex:LexicalEntry
ex:C199_ca_amfitrio_amfitriona_entry
dct:language
"ca"^^xsd:lang
ontolex:lexicalForm
ontolex:Form
ex:C199_ca_amfitrio_amfitriona_form_masc_pl
"amfitrió"^^rdf:langString
ontolex:writtenRep
ontolex:gender
lexinfo:Gender
lexinfo:feminine
lexinfo:Gender
lexinfo:masculine
ontolex:gender
ontolex:lexicalForm
ontolex:Form
ex:C199_ca_amfitrio_amfitriona_form_fem_pl
ontolex:writtenRep
"amfitriona"^^rdf:langString
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
xsd: http://www.w3.org/2001/XMLSchema#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 14 Modelling of Proposal 1 for 'amfitrió -iona', example from *Argot culinari i gastronòmic*, TERMCAT, fitxa 199

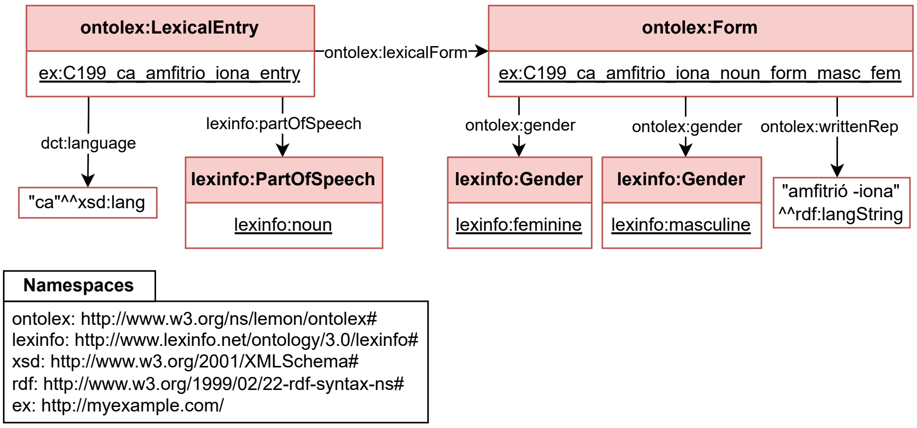
ontolex:LexicalEntry
ex:C199_ca_amfitrio_iona_entry
ontolex:lexicalForm
ontolex:Form
ex:C199_ca_amfitrio_iona_noun_form_masc_fem
dct:language
"ca"^^xsd:lang
lexinfo:partOfSpeech
lexinfo:PartOfSpeech
lexinfo:noun
ontolex:gender
lexinfo:Gender
lexinfo:feminine
ontolex:gender
lexinfo:Gender
lexinfo:masculine
ontolex:writtenRep
"amfitrió -iona"^^rdf:langString
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
xsd: http://www.w3.org/2001/XMLSchema#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 15 Modelling of Proposal 2 for 'amfitrió -iona', example from *Argot culinari i gastronòmic*, TERMCAT, fitxa 199

```
<denominacio
    llengua="ca"
    tipus="principal"
    jerarquia="terme pral."
    categoria="n m, f"
    ><![CDATA[amfitrió -iona]]>
</denominacio>
```

Listing 6 Example from *Argot culinari i gastronòmic*, TERMCAT, fitxa 199

Proposal 2 suggested mimicking the original data and retaining the entire string within a single Form. Additionally, two genders would be assigned to the Form as illustrated in Figure 15. This approach preserves the original data without requiring automatic processing, thus avoiding potential data errors.

Alternatively, Proposal 3 suggested representing suffixed feminine forms using the Morph ontology,[42] an Ontolex extension for morphological representation. This ontology allows a form to be decomposed into its root (`morph:RootMorph`) and the masculine or feminine suffixes (`morph:Suffix`), as shown in Figure 16. However, since TERMCAT does not specify either the root or the masculine suffix, this approach would

[42] https://www.w3.org/community/ontolex/wiki/Morphology#

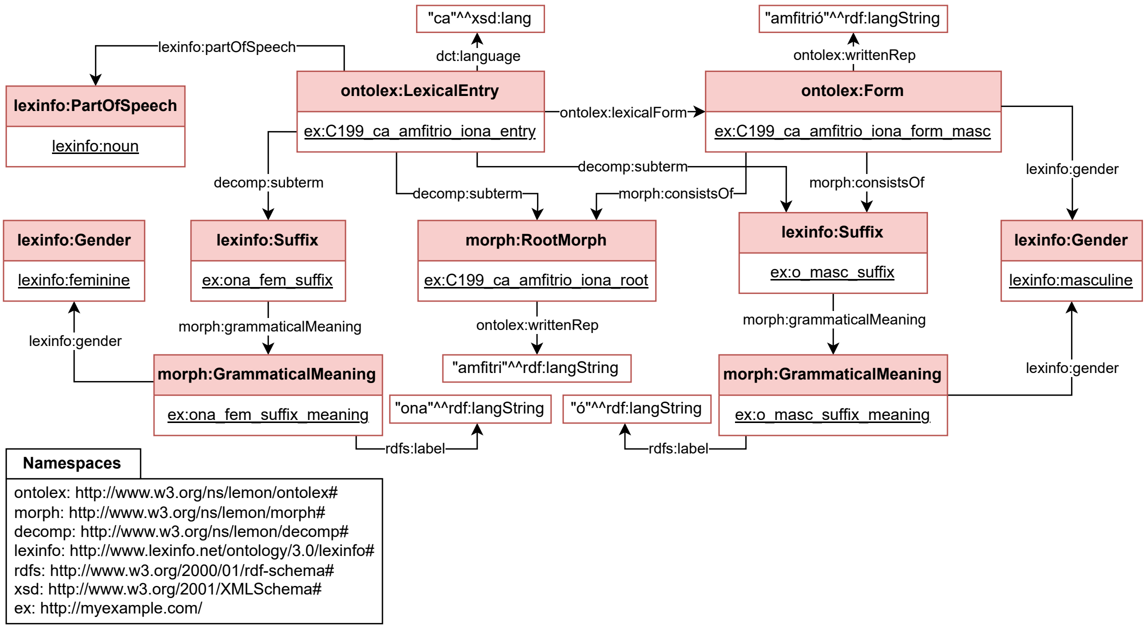

Fig. 16 Modelling of Proposal 3 for 'amfitrió -iona', example from *Argot culinari i gastronòmic*, TERMCAT, fitxa 199

need additional data preprocessing. Rather than generating the feminine form directly, the masculine form would need to be segmented into its root and masculine suffix. Although it would be possible to omit the explicit modelling of the root and masculine suffix, representing only the full masculine form and the feminine affix, this approach has limitations. Specifically, the absence of a root would prevent the retrieval of feminine forms through SPARQL queries.

After reviewing the alternative proposals, the word formation approach (Proposal 1) was ultimately selected. Although this paper advocates for a representation that remains as faithful as possible to the original data, the single-string representation suggested in Proposal 2 was deemed inadequate, as it would declare a single `ontolex:Form` for a given Lexical Entry, instead of distinguishing two separate forms. Furthermore, while gender distinctions are explicitly indicated in the original XML files through the gender and form order (e.g. 'n m, f'), this information would be lost in the RDF representation, as each gender is represented individually without a defined order. Consequently, Proposal 2 was not implemented. Similarly, Proposal 3 did not seem to be the best option, as it would require automatic preprocessing while introducing a more complex representation, thereby complicating SPARQL queries. For these reasons, the word formation approach (Proposal 1) was adopted (see Figure 14), despite acknowledging the potential errors it may introduce.

5.6 Verbs

Among the terms in TERMCAT glossaries, verbs are also found, identified by the 'v' value in the `categoria` attribute of the `denominacio` node. To model verbs, the instance `lexinfo:verb` from the class `lexinfo:PartOfSpeech` is used. This instance is linked to the Lexical Entry as illustrated in Figure 17. Additionally, some entries provide further information about the verb, such as valency (i.e., transitive or intransitive). In the original XML resources, transitive verbs are marked as 'v tr', while intransitive verbs are designated as 'v intr'. To model these values, OLiA was used, specifically the classes `olia:Intransitive` and `olia:Transitive` (see Figure 17).

In addition to valency, other verb characteristics may also be specified. For example, certain verbs are classified as prepositional verbs, indicated by the value 'v prep'. According to TERMCAT, Centre de Terminologia (2022d), prepositional verbs are those that typically require a complement introduced by a preposition. In the XML terminology resources, prepositions are enclosed within italicised HTML tags (<i> and </i>) and square brackets ([]), as illustrated in Listing 7. To model this phenomenon, `lexinfo:PrepositionFrame` can be used to indicate that the verb requires a complement introduced by a preposition (see Figure 18). Additionally, the specific prepositions that a verb may take (e.g. 'from')

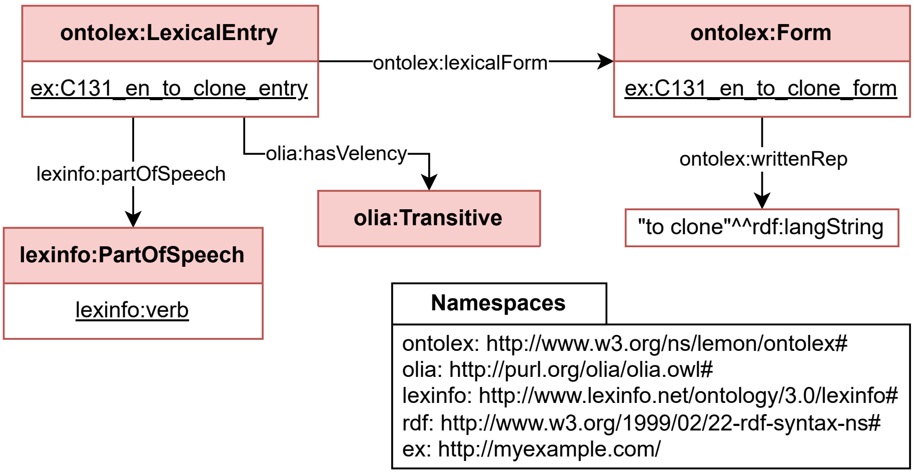
ontolex:LexicalEntry
ex:C131_en_to_clone_entry
ontolex:lexicalForm
ontolex:Form
ex:C131_en_to_clone_form
olia:hasVelency
lexinfo:partOfSpeech
ontolex:writtenRep
olia:Transitive
"to clone"^^rdf:langString
lexinfo:PartOfSpeech
lexinfo:verb
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
olia: http://purl.org/olia/olia.owl#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 17 Modelling of transitive verbs, from *Diccionari de bioètica*, TERMCAT, fitxa 131

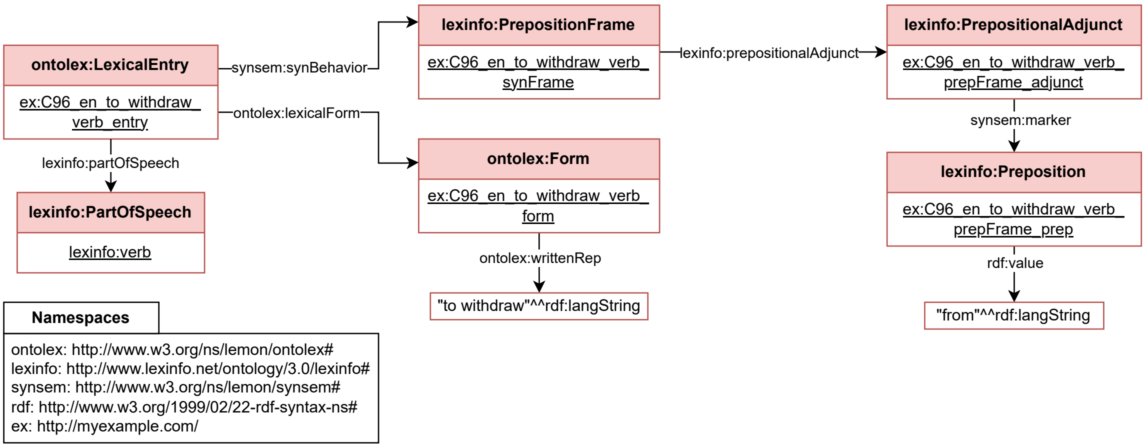
ontolex:LexicalEntry
ex:C96_en_to_withdraw_verb_entry
synsem:synBehavior
ontolex:lexicalForm
lexinfo:PrepositionFrame
ex:C96_en_to_withdraw_verb_synFrame
lexinfo:prepositionalAdjunct
lexinfo:PrepositionalAdjunct
ex:C96_en_to_withdraw_verb_prepFrame_adjunct
synsem:marker
lexinfo:Preposition
ex:C96_en_to_withdraw_verb_prepFrame_prep
rdf:value
"from"^^rdf:langString
ontolex:Form
ex:C96_en_to_withdraw_verb_form
ontolex:writtenRep
"to withdraw"^^rdf:langString
lexinfo:partOfSpeech
lexinfo:PartOfSpeech
lexinfo:verb
Namespaces
ontolex: http://www.w3.org/ns/lemon/ontolex#
lexinfo: http://www.lexinfo.net/ontology/3.0/lexinfo#
synsem: http://www.w3.org/ns/lemon/synsem#
rdf: http://www.w3.org/1999/02/22-rdf-syntax-ns#
ex: http://myexample.com/

Fig. 18 Modelling of prepositional verbs, example from *Diccionari general de l'esport*, TERMCAT, fitxa 96

can be represented using `lexinfo:Preposition`. This preposition can then be designated as a marker of an adjunct (`lexinfo:PrepositionalAdjunct`) through the module for the representation of Syntax and Semantics (synsem).

```
<denominacio
    llengua="en"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria="v prep">
    ><![CDATA[withdraw <i>[from]</i>, to]]>
</denominacio>
```

Listing 7 Example from *Diccionari general de l'esport*, TERMCAT, fitxa 96

Lastly, verbs may also be classified as 'pronominal', denoted by the label 'pron' (see Listing 8). According to TERMCAT, Centre de Terminologia (2022e), pronominal verbs are those that are accompanied by a pronoun that agrees with the subject but does not fulfil any specific syntactic function within the sentence. Consequently, `lexinfo:ReflexiveFrame` was proposed to represent pronominal data. However, in certain languages, such as Spanish, pronominal and reflexive forms do not appear to be entirely equivalent. For instance, as noted by the Royal Spanish Academy, pronominal verbs should not be categorised as reflexive verbs due to grammatical nuances. In reflexive usage, the pronoun functions as a direct object, representing the person affected by the action. In contrast, in pronominal usage, the pronoun is merely part of the structure of the verb and does not act as an argument (that is, it does not represent a separate participant in the action). For instance, the first person singular pronoun 'me' has a reflexive use in the sentence "Me mojé a mí mismo" (I got myself wet) but has a pronominal use in the sentence "Empezó a llover y me mojé" (It started raining, and I got wet). (Real Academia Española and Asociación de Academias de la Lengua Española, n.d.)

```
<denominacio
    llengua="es"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria="v pron"
    ><![CDATA[registrarse]]>
</denominacio
```

Listing 8 Example from *Diccionari de turisme*, TERMCAT, fitxa 515

To avoid the classification of a pronominal verb as reflexive, an alternative approach was considered, namely, the representation of pronominal verbs using `olia:Cliticization`. This process is defined as "a process by which a complex word is formed by attaching a clitic to a fully inflected word".[43] This approach could be applied to languages such as Spanish and Catalan, where pronominalisation appears to occur through the attachment of 'se' and '-se' to the verb, respectively. However, in French, pronominalisation morphemes may either be affixed at the beginning of the verb (e.g., s'enregistrer) or appear as separate elements (e.g., se loger). In cases where the pronominal morpheme is not directly attached to the verb, the use of `olia:Cliticization` may not be appropriate.

Ultimately, in the absence of a more suitable approach for modelling pronominal verbs, the use of `lexinfo:ReflexiveFrame` was adopted until a more precise solution is identified. While acknowledging minor inaccuracies in the data, it is important to note that pronominal pronouns have traditionally been classified as reflexive pronouns (Real Academia Española and Asociación de Academias de la Lengua Española, n.d.).

6 Complete Modelling Proposal

TERMCAT terminology resources cover a wide range of domains, which often present different modelling needs. Taking into account all those necessities, the modelling schema in Figure 19 was proposed. This model is created by reusing existing vocabularies and ontologies. In particular, this schema is based on the Ontolex model, together with several other ontologies and vocabularies introduced in Section 4.

As can be observed in Figure 19, the core of the model is composed of three Ontolex classes: Lexical Concept, Lexical Sense, and Lexical Entry. These classes are connected in all directions. Following the TERMCAT structures in the XML terminology resources, a Lexical Concept usually has more than one associated Lexical Entry, and each Lexical Entry has one associated Lexical Sense.

Lexical Concepts count with `dct:identifier` that registers the original TERMCAT id, marked on the `num` attribute of the `fitxa` node. Moreover, each Lexical Concept is associated with a Concept Set (`ontolex:ConceptSet`), which corresponds to a TERMCAT resource. The URL of the corresponding TERMCAT resource is linked to the Concept Set through the property `dct:source`. Additionally, definitions and notes are associated to a Lexical Concept through `termlex:Definition` and `termlex:Note`. These classes allow for grouping definitions and notes by language or/and source, for instance. Moreover, domains are also bound to the Lexical Concept, following the hierarchy of the TERMCAT XML terminology resources. This connection is marked by `rdfs:domain`. Each domain constitutes a `skos:Concept` and a hierarchical representation is achieved through the properties `skos:narrower` and `skos:broader`. All domains are grouped in a `skos:ConceptScheme` through the property `skos:inScheme`, and the name of the schema is indicated with the property `rdfs:label`. Additionally, the broadest domains are marked through the property `skos:topConcept`.

Furthermore, each TERMCAT `denominacio` node constitutes a term entry, represented with `ontolex:LexicalEntry`. If the TERMCAT entries are marked as prefix or suffix (identified with 'pfx' and 'sfx' in the attribute `categoria`), the subclasses `ontolex:Prefix` and `ontolex:Suffix` are used instead, respectively. The type of term (e.g., abbreviation, formula, scientific name, etc.) can also be modelled with the instances in `lexinfo:TermType`. A term entry can have some associated part-of-speech information through the instances declared in `lexinfo:PartOfSpeech`. As discussed in Section 5.6,

[43] http://purl.org/olia/olia.owl#

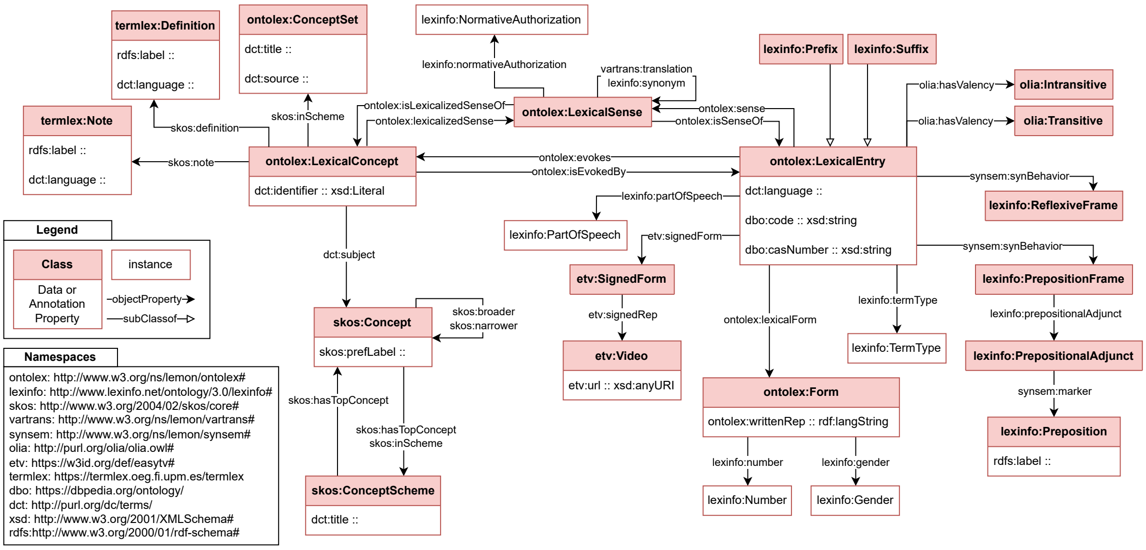

Fig. 19 Complete modelling proposal of TERMCAT resources

verbs may contain more information about valencies (represented with the property `olia:hasValency`) or syntactic behaviour (through the property `synsem:synBehavior`). Lexical Entries also have a form, whether written or signed. Signed forms (in Catalan Sign Language) are identified with the value 'SC' in the `llengua` attribute of the `denominacio` node. These types of entries tend to have the URL of a YouTube video as displayed in Listing 9. Signed forms are represented with the class `etv:SignedForm`, which can have an associated `etv:Video`. The URL of the video is indicated with the property `etv:url`.

```
<denominacio
      llengua="en"
      tipus="equivalent"
      jerarquia="terme pral."
      categoria="n"
      ><![CDATA[abstention]]>
</denominacio>
<denominacio
    llengua="SC"
    tipus="equivalent"
    jerarquia="terme pral."
    categoria=""
    ><![CDATA[https://youtu.be/wkAx3wXJViY]]>
</denominacio>
```

Listing 9 *Diccionari de l'activitat parlamentària*, TERMCAT, fitxa 3

On the other hand, written forms are modelled with the class `ontolex:Form`, which uses the property `ontolex:writtenRep` to represent the written form. Moreover, the term's grammatical number (e.g., plural) can be specified with the instances declared in `lexinfo:Number`. Similarly, the gender of the term (masculine, feminine, or neuter) can be indicated with the instances in `lexinfo:Gender`.

Lastly, Lexical Senses (`ontolex:LexicalSense`) are used to indicate the lexical meaning of a Lexical Entry. Lexical Senses can be related to each other by a translation or synonym relation, depending on the language of the term. A relation of synonymy (`lexinfo:synonym`) is created between the terms of a concept that share the same language, while a relation of translation (`vartrans:translation`) is created between the terms of different languages linked to the same concept. Additionally, information about the status of the term (e.g., deprecated) can be included in the instances of `lexinfo:NormativeAuthorization`.

7 Conclusion

This study examines the modelling of terminology resources using the Ontolex framework, ensuring that the representation meets both semantic requirements and the needs of automated processing. For this purpose, a collection of terminology resources from the Catalan Terminology Centre (TERMCAT) portal was examined, as this is a highly relevant resource at a national level. This collection comprises more than 150 terminology resources that cover a wide range of domains, including health, law, gastronomy, and sports. In addition to the diversity of subject areas, the terminology resources contain various types of term entries, such as chemical formulas, codes, symbols, and signed forms. As Ontolex alone does not fully accommodate all these data types, complementary ontologies, such as LexInfo, were considered.

To gain familiarity with the data, a preliminary analysis was conducted, revealing that certain elements of the XML are used inconsistently. For instance, the `llengua` attribute within the `denominacio` node is not solely employed to indicate the language of a term but also to denote whether the term represents a symbol, a code, a formula, a scientific name, or an anatomical or histological term. This irregular use may result from the constraints within the TERMCAT format when structuring terminological data. Such inconsistencies complicate automation processes, as the appearance of new values in this attribute would require manual verification and updates. For example, if a new value such as 'Arch' were introduced to designate archaeological terminology, since such value would not be registered within the exceptions, the automatic conversion would associate such information to the `ontolex:LexicalEntry` with the `dct:language`, resulting in a misrepresentation.

During the process of modelling the TERMCAT resources, several challenges and uncertainties emerged. One of the main difficulties was the identification of appropriate classes and properties to accurately represent certain data, as seen in the cases of pronominal verbs and symbols. In some instances, the selected representation was not entirely suitable due to its overly generic nature.

Another challenge involved the adequacy of representation in relation to semantic relationships. For instance, it was necessary to determine whether symbols and codes should be directly linked to the Lexical Concept or assigned their own Lexical Sense, connecting them to other terms through a synonymy relation.

Similarly, in the domain of sea mammals, it remained unclear whether authorship information should be (i) associated with the Lexical Concept, (ii) linked to the Lexical Sense of the scientific name, or (iii) represented as an independent term with no explicit semantic relation.

In all cases, the selected approach adhered to the principle of maintaining the closest possible alignment with the original XML data structure. However, in certain cases, minor modifications were proposed. For instance, in languages with grammatical gender, a term may have several forms, according to their gender. In TERMCAT, these form variants are registered in a single entry. For example, a single entry may contain a masculine and feminine form together. These dual forms may appear in one of three ways: (i) as a single word, indicating that both masculine and feminine forms are homonyms; (ii) as two full forms separated by a vertical bar; or (iii) as the full masculine form followed by the feminine suffix. To ensure accurate representation, we propose modelling these as two distinct `ontolex:Form` instances associated with the same `ontolex:LexicalEntry`. In order to extract the written representation of each form, the grouped information must be processed. Specifically, homonyms are duplicated, full forms are separated using the vertical bar (|), and the complete feminine form is generated automatically.

In terms of directions for future research, further analysis of terminology resources is recommended to enrich the modelling schema presented, such as the Interactive Terminology for Europe (IATE) resource.[44] This resource offers various features, including information about the source of a term, a note, or a definition. Additionally, concepts may encompass relationships with other concepts (referred to as cross-references), including but not limited to: "has capital city," "is narrower than," or "is not to be confused with." Furthermore, additional part-of-speech values, such as *nominal phrase*, may be encountered. In addition, terms in IATE are assigned a reliability code, which could provide valuable insights for further modelling improvements.

The art of modelling terminology resources therefore requires a thorough analysis that brings together experts in Semantic Web standards and domain specialists, ensuring that the proposed solution effectively balances technical interoperability with domain-specific requirements.

Acknowledgements

This work has received funding from the INESData project (Infrastructure to Investigate Data Spaces in Distributed Environments at UPM - TSI-063100-2022-0001, under the UNICO I+D CLOUD call) and TeresIA

[44] https://iate.europa.eu/home

(Terminology and AI), both funded by the Ministry for Digital Transformation and the Civil Service, within the framework of the recovery plan PRTR financed by the European Union (NextGenerationEU).

References

Almeida, B., Costa, R., Salgado, A., Ramos, M., Romary, L., Khan, F., … Tasovac, T. (2022). Modelling Usage Information in a Legacy Dictionary: From TEI Lex-0 to Ontolex-Lemon. Y. Rochat, C. Métrailler, & M. Piotrowski (Eds.), *Proceedings of the Workshop on Computational Methods in the Humanities 2022* (Vol. 3602, pp. 5–21). Lausanne, Switzerland: CEUR. Retrieved from https://ceur-ws.org/Vol-3602/#paper1 (ISSN: 1613-0073)

Bellandi, A. (2021). LexO: an open-source system for managing OntoLex-Lemon resources. *Language Resources and Evaluation*, *55*(4), 1093–1126, https://doi.org/10.1007/s10579-021-09546-4

Bellandi, A., Di Nunzio, G.M., Piccini, S., Vezzani, F. (2023). From TBX to Ontolex Lemon: Issues and Desiderata. G.M.D. Nunzio, R. Costa, & F. Vezzani (Eds.), *Proceedings of the 2nd International Conference on Multilingual Digital Terminology Today (MDTT 2023)* (Vol. 3427). Lisbon, Portugal: CEUR. Retrieved from https://ceur-ws.org/Vol-3427/#paper4 (ISSN: 1613-0073)

Bellandi, A., Di Nunzio, G.M., Piccini, S., Vezzani, F. (2024). LemonizeTBX: Design and Implementation of a New Converter from TBX to OntoLex-Lemon. *Digital Humanities Quarterly*, *18*(2), , Retrieved from https://www.digitalhumanities.org/dhq/vol/18/2/000745/000745.html

Bosque Gil, J., Lonke, D., Gracia del Río, J., Kernerman, I. (2019). Validating the OntoLex-lemon Lexicography Module with K Dictionaries' Multilingual Data. *Electronic lexicography in the 21st century. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal* (pp. 726–746).

Bosque-Gil, J., Gracia, J., Montiel-Ponsoda, E., Gómez-Pérez, A. (2018). Models to represent linguistic linked data. *Natural Language Engineering*, *24*(6), 811–859, https://doi.org/10.1017/S1351324918000347

Bosque-Gil, J., Montiel-Ponsoda, E., Gracia, J., Aguado-De-Cea, G. (2016). Terminoteca RDF: A gathering point for multilingual terminologies in Spain. *Proceedings of Term Bases and Linguistic Linked Open Data - TKE 2016, 12th International Conference on Terminology and Knowledge Engineering* (pp. 136–146).

Cabré, M.T. (1999). *Terminology: Theory, Methods and Applications.* John Benjamins Publishing Company.

Cambridge Dictionary (n.d.). *Histology.* Retrieved from https://dictionary.cambridge.org/dictionary/english/histology (Accessed: 2025-03-18)

Cimiano, P., Buitelaar, P., McCrae, J., Sintek, M. (2011). LexInfo: A declarative model for the lexicon-ontology interface. *Journal of Web Semantics*, *9*(1), 29–51, https://doi.org/10.1016/j.websem.2010.11.001

Cimiano, P., McCrae, J.P., Rodríguez-Doncel, V., Gornostay, T., Gómez-Pérez, A., Siemoneit, B., Lagzdins, A. (2015). Linked terminologies: applying linked data principles to terminological resources. *Proceedings of the eLex 2015 Conference* (pp. 504–517). Retrieved from https://elex.link/elex2015/proceedings/eLex_2015_34_Cimiano+etal.pdf (ISBN 978-961-93594-3-3)

Collins Dictionary (2025). *Code - Definition and Meaning.* Retrieved from https://www.collinsdictionary.com/dictionary/english/code (Accessed: 2025-03-18)

di Buono, M.P., Cimiano, P., Elahi, M.F., Grimm, F. (2020). Terme-à-LLOD: Simplifying the conversion and hosting of terminological resources as linked data. M. Ionov, J.P. McCrae, C. Chiarcos, T. Declerck, J. Bosque-Gil, & J. Gracia (Eds.), *Proceedings of the 7th workshop on linked data in linguistics (ldl-2020)* (pp. 28–35). Marseille, France: European Language Resources Association. Retrieved from https://aclanthology.org/2020.ldl-1.5/

Gracia, J., Villegas, M., Gómez-Pérez, A., Bel, N. (2018). The apertium bilingual dictionaries on the web of data. *Semantic Web*, *9*(2), 231–240, https://doi.org/10.3233/SW-170258

ISO/CD 704:2022 (2022). *International standard iso 704:2020: Terminology work - principles and methods.* Geneva: ISO/CD 704:2022. International Organization for Standardization.

Martín-Chozas, P., Declerck, T., Montiel-Ponsoda, E., Rodríguez-Doncel, V. (2024). Representing terminological data in the semantic web. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, , https://doi.org/https://doi.org/10.1075/term.22037.mar

McCrae, J., Aguado-de Cea, G., Buitelaar, P., Cimiano, P., Declerck, T., Gómez-Pérez, A., … Wunner, T. (2012). Interchanging lexical resources on the Semantic Web. *Language Resources and Evaluation*, *46*(4), 701–719, https://doi.org/10.1007/s10579-012-9182-3

McCrae, J., Fellbaum, C., Cimiano, P. (2014). Publishing and Linking WordNet using lemon and RDF. *Proceedings of the 3rd Workshop on Linked Data in Linguistics.*

McCrae, J.P., Bosque-Gil, J., Gracia, J., Buitelaar, P., Cimiano, P. (2017). The Ontolex-Lemon model: development and applications. *Proceedings of eLex 2017 conference* (pp. 19–21).

Melby, A.K. (2015). TBX: A terminology exchange format for the translation and localization industry. *Handbook of Terminology: Volume 1* (pp. 393–424). John Benjamins Publishing Company.

Miles, A., Matthews, B., Wilson, M., Brickley, D. (2005, September). SKOS core: simple knowledge organisation for the web. *Proceedings of the 2005 international conference on Dublin Core and metadata applications: vocabularies in practice* (pp. 1–9). Madrid, Spain: Dublin Core Metadata Initiative.

Ralli, N., & Evers, E. (2024). To gender or not to gender, that is the question: gender-inclusive language in the legal context. *Terminology Science & Research / Terminologie : Science et Recherche*, *27*, 75–92, Retrieved from https://journal-eaft-aet.net/index.php/tsr/issue/archive

Real Academia Española and Asociación de Academias de la Lengua Española (n.d.). *Glosario de términos gramaticales.* Online version. Retrieved from https://www.rae.es/gtg/verbo-pronominal (Accessed: 2025-03-21)

Stellato, A., Fiorelli, M., Turbati, A., Lorenzetti, T., Van Gemert, W., Dechandon, D., … others (2020). VocBench 3: A collaborative Semantic Web editor for ontologies, thesauri and lexicons. (Vol. 11, pp. 855–881). SAGE Publications Sage UK: London, England.

TERMCAT, Centre de Terminologia (2015–2024). *Terminologia de ciències de la salut.* Barcelona: TERMCAT, Centre de Terminologia. Retrieved from https://www.termcat.cat/Thor/files/diccionaris/wadfdlcienciesdelasalut2024.xml

TERMCAT, Centre de Terminologia (2022a). *Adequació: Termes preferents, sinònims complementaris, variació lingüística i alternatives sinònimes.* Retrieved from https://arxiu.termcat.cat/criteris/CRJER02_ADEQUACIO_TermePralSinComplVarLingAltSin.pdf (Accessed: 2025-03-18)

TERMCAT, Centre de Terminologia (2022b). *SÍMBOLS: Naturalesa, representació i restriccions.* Retrieved from https://arxiu.termcat.cat/criteris/CREQU06_SIMBOLS_NaturalesaRepresentacioRestriccions.pdf (Accessed: 2025-03-18)

TERMCAT, Centre de Terminologia (2022c). *VARIANTS LINGÜISTIQÜES: Caracterització i representació.* Retrieved from https://arxiu.termcat.cat/criteris/CRJER05_VARIANTSLINGUISTIQUES_CaracteritzacioRepresentacio.pdf (Accessed: 2025-03-18)

TERMCAT, Centre de Terminologia (2022d). *VERBS: Representació dels verbs prepósiciónals.* Retrieved from https://arxiu.termcat.cat/criteris/CRCAT05_VERBS_RepresentacioVerbsPreposicionals_accedira.pdf (Accessed: 2025-03-18)

TERMCAT, Centre de Terminologia (2022e). *VERBS: Representació dels verbs prónóminals.* Retrieved from https://arxiu.termcat.cat/criteris/CRCAT06_VERBS_RepresentacioVerbsPronominals_imaginar-seentrenar-se.pdf (Accessed: 2025-03-18)

TERMCAT, Centre de Terminologia (2024). *Llengües i equivalència: Ordenació de les fitxes terminològiques del TERMCAT.* Retrieved from https://arxiu.termcat.cat/criteris/CRQUEST03_LLENGUESEQUIVALENCIA_OrdenacioFitxesTERMCAT.pdf (Accessed: 2025-03-18)