Journal of Digital Terminology and Lexicography - ISSN:3103-3601
Vol. 2, No. 1 (2026) DOI: 10.25430/pupj.jdtl.1784885954

Ontology-Based Multilingual Terminological Resource for Biomedical Abbreviations

Margarida Ramos[1,*], Matilde Canelas[1], Chiara Barbero[1], Ana Mouro[1], and Rute Costa[1]

[1]NOVA CLUNL - Linguistic Research Centre of NOVA University Lisbon, Av. de Berna, 26 C, 1069-061, Lisbon, Portugal.

*Corresponding author(s). E-mail(s): mvramos@fcsh.unl.pt
Contributing authors: amcanelas@fcsh.unl.pt; chiarabarbero@fcsh.unl.pt; almouro@fcsh.unl.pt; rute.costa@fcsh.unl.pt

Abstract

Biomedical abbreviations play a central role in scientific and clinical communication, allowing textual economy and easing expert interaction. However, their widespread use can generate ambiguity, especially in multilingual and interdisciplinary contexts, where a single abbreviation may designate different concepts, while multiple abbreviations may refer to the same concept. Although biomedical abbreviation ambiguity has been significantly investigated, the conceptual organisation of abbreviation usage across biomedical knowledge resources remains underexplored. This study investigates patterns of intra- and interlingual denominative variation, as well as mono- and polyreferential ambiguity involving abbreviated and full forms. A comparative analysis of conceptual representation, preferred labels, indexed variants, and abbreviation coverage across authoritative biomedical knowledge resources was conducted, being supported by a multilingual comparable corpus of specialised texts that focused on neurological, neurodegenerative, gastrointestinal, psychiatric, and metabolic disorders linked to gut–brain axis dysfunction. Data were processed through automated extraction techniques and conceptual alignment procedures to compare usage found in corpora with structured knowledge representations. The results revealed systematic variations in how abbreviations are used across languages and recurring differences between textual evidence and ontological representations, reflecting the challenges of integrating biomedical knowledge from resources built according to different perspectives and purposes. In response to these challenges, the study introduced an ontology, designed for accurately representing the conceptual and linguistic dimensions of biomedical abbreviation ambiguity, together with a multilingual dictionary integrating corpus evidence and terminological data. Overall, this work supports the use of ontology-driven and corpus-informed approaches to improve semantic interoperability and structured biomedical knowledge representation across languages.

Keywords: abbreviations, biomedical terminology, domain-ontology, terminology, terminology resource

1. Introduction

Biomedical texts are characterised by a high degree of terminological density and extensive use of abbreviated forms (see Wermuth & Verplaetse, 2019). In scientific and clinical communication alike, abbreviations contribute to textual economy and facilitate specialised interaction among experts (see Bowker & Hawkins, 2006). However, their widespread use also introduces significant challenges regarding interpretation, conceptual clarity, and knowledge representation. These challenges become particularly evident in multilingual and interdisciplinary contexts, where the same abbreviation may designate different concepts, or different abbreviations may refer to the same concept across languages and resources.

The ambiguity associated with biomedical abbreviations has been extensively documented in previous research, particularly in relation to natural language processing, information retrieval, and clinical communication (Liu et al., 2001; Grossman Liu et al., 2021; Jaber & Martínez, 2021). While many studies focus on abbreviation expansion and disambiguation in corpora or clinical records, less attention has been devoted to the conceptual organisation underlying abbreviation usage and variation across biomedical knowledge resources. In particular, discrepancies between specialised text and the representation of concepts in biomedical ontologies and thesauri may compromise terminological consistency, semantic interoperability, and the accurate interpretation of specialised texts.

Within this context, the present study addresses abbreviation ambiguity from a terminological and knowledge-organisation perspective. Grounded in the theoretical framework of Terminology, particularly the principle of its double – conceptual and linguistic – dimension, this paper analyses the relations established between abbreviations, full forms, and concepts in multilingual biomedical text.

The analysis is based on multilingual comparable corpora in English, Portuguese, and Italian, from which abbreviations and corresponding full forms were semi-automatically extracted through corpus query methodologies and manually curated. We compared corpus-attested usages with the representation of the corresponding concepts in MeSH,[1] NCIt,[2] SNOMED CT,[3] and IOBC,[4] assessing the extent to which these resources capture the conceptual and terminological variation evidenced in the corpora.

Particular attention is devoted to cases of intralingual and interlingual denominative variation and ambiguity. In terminology science, denominative variation belongs to the linguistic dimension of terminology. It refers to the presence of multiple terms designating the same concept within a specific domain or across different contexts. A single concept may thus be expressed through more than one designation, and these alternative designations are known as denominative variants. Such variation can arise from several factors, including synonymy, the coexistence of abbreviated and full forms, and historical or diachronic evolution, among others. As noted by Daille (2017, p. 30), "denominative variants are lexicalised forms representing the same concept, but adopting different lexical forms according to different contexts". Ambiguity, in contrast, occurs when a single linguistic form admits more than one interpretation. When those interpretations are semantically related, the phenomenon is more specifically termed polysemy. In terminological work, both ambiguity and polysemy are generally viewed as obstacles to precision and univocity. At the conceptual level, monoreferentiality and polyreferentiality describe complementary referential patterns. Monoreferentiality occurs when different forms, whether full forms or abbreviations, designate a single concept. Polyreferentiality, by contrast, occurs when one identical form, particularly an abbreviation, designates multiple distinct concepts. This distinction is important for analysing conceptual organisation and referential precision in specialised texts.

To address these issues, this study proposes two complementary resources: (i) OntoAbbMed, an OWL ontology that systematises conceptual and linguistic information on biomedical abbreviation ambiguity, and (ii) a *Multilingual Dictionary of Biomedical Abbreviations* integrating corpus evidence,

[1] https://www.ncbi.nlm.nih.gov/mesh/
[2] https://ontobee.org/ontology/NCIT
[3] https://termbrowser.nhs.uk/?
[4] https://bioportal.bioontology.org

conceptual relations, and multilingual terminological data. By combining corpus-based terminology work, biomedical resources, and ontology engineering, this study aims to contribute to specialised knowledge organisation and reduce ambiguity in biomedical communication.

The article is organised as follows. Section 2 presents the theoretical framework and the scope of the study. Section 3 describes the multilingual corpora and compilation methodology. Section 4 addresses terminology extraction procedures, while Section 5 systematises the variation patterns identified in the corpora. Section 6 analyses the representation of the corresponding concepts in biomedical resources. Sections 7 and 8 present OntoAbbMed and the multilingual dictionary respectively. Finally, Section 9 summarises the main findings and outlines future work.

2. Theoretical Framework and Scope of Application

2.1. The Project

This article is developed within the framework of the HEREDITARY project[5] and contributes to continuing research on the integration and representation of knowledge related to neurodegenerative diseases associated with the gut–brain axis. The project adopts a multidisciplinary approach, combining linguistic and clinical data with computational methods to support both specialised knowledge organisation and accessibility to diverse audiences.

In this paper, we focus exclusively on healthcare expert audiences, with the proposed ontology and dictionary designed for them, while also serving as a starting point for the future development of definitions aimed at non-expert users. In this context, the present work focuses on the extraction, analysis, and structuring of terminological data from specialised multilingual corpora, with the aim of structuring domain knowledge in a systematic and concept-oriented manner. In this setting, abbreviations emerge as a particularly relevant phenomenon, given their frequent use in biomedical text and their potential to generate ambiguity at both intra- and interlinguistic levels.

2.2. Double Dimension of Terminology

To address this phenomenon, this work is grounded in the theoretical framework of terminology, particularly the principle of its double dimension (conceptual and linguistic) (Costa, 2013; Santos & Costa, 2015). Concepts belong to structured systems of knowledge and are defined through their relations to other concepts, while terms correspond to their linguistic expressions in text (International Standard Organization [ISO], 2019; Ramos & Costa, 2024). This twofold perspective implies that terminological work involves not only identifying and organising concepts, but also analysing the terms used to designate them in specialised communication. Methodologically, this is reflected in the articulation of semasiological and onomasiological approaches, moving respectively from terms to concepts and from concepts to terms. This allows continuous interaction between linguistic data and conceptual structuring. This framework is particularly relevant to abbreviation analysis, as abbreviations operate at the linguistic level and often create ambiguity in term–concept relations, for instance when one abbreviation designates multiple concepts or several abbreviations refer to the same concept, especially in multilingual and specialised contexts.

2.3. Biomedical Abbreviations

Abbreviations are a recurrent and productive feature in biomedical texts, where they promote economy of expression and communicative efficiency. Although abbreviations function at the linguistic level, they remain associated with the underlying concepts they designate within a domain. However, this association is not always stable or unambiguous. Previous studies have shown that abbreviations frequently give rise to ambiguity, particularly when a single abbreviation corresponds to multiple concepts depending on context (see Pakhomov et al., 2002, p. 589), and that their understanding depends on identifying their expanded forms in text (Schwartz & Hearst, 2003). Conversely, a single concept may be designated by multiple abbreviations. This reflects denominative variation, whereby a single concept

[5] https://hereditary-project.eu/

may be associated with multiple designations (Cabré Castellví, 1999), including abbreviations as alternative reduced forms of designation.

Moreover, abbreviation use varies across biomedical text types: in scientific literature, abbreviations are typically introduced alongside their full forms upon first occurrence and subsequently used without expansion, whereas in clinical notes they are often used without such explicit introduction, making understanding more dependent on contextual information (Xu et al., 2007, p. 825). The complexity of abbreviation–concept relations is also reflected in UMLS,[6] where a single abbreviation may designate more than one concept. Empirical studies further underscore this issue, indicating that a substantial proportion of abbreviations are ambiguous and correspond to multiple concepts (Liu et al., 2001, p. 396; Liu et al., 2002, p. 467).

This ambiguity has also been documented in clinical practice, where abbreviations may be misinterpreted even among healthcare professionals, sometimes resulting in patient harm (Brunetti et al., 2007, p. 580). Even abbreviations assumed to be standard and universally accepted are frequently not recognised by their peers (Jayatilake & Oyibo, 2023). Within a cross-institutional context, the issue of misinterpretation assumes wider dimensions, since ambiguities in clinical notes and medical records may affect information transfer among expert teams and compromise the continuity and safety of patient care (Berger et al., 2026).

In response, organisations responsible for healthcare accreditation and patient safety, such as the Joint Commission,[7] have introduced recommendations restricting the use of error-prone abbreviations, including the 'Do Not Use' list, to improve communication and patient safety (The Joint Commission, 2003). These issues are particularly salient in biomedical texts, where abbreviations are highly productive and context-dependent. In multilingual contexts, such issues may be amplified by differences in linguistic conventions and terminological practices, posing challenges for the identification, alignment, and representation of concepts across resources and affecting terminological consistency and clarity in specialised communication.

3. The Corpora

Corpus compilation and corpus analysis methods for data extraction are broadly applied in terminological work (ISO, 2019; ISO, 2025). In line with a linguistic approach to terminology, anchoring in empirical data is widely accepted within the scientific community. Corpora constitute samples of natural language, including language varieties, formed by collecting and organising authentic 'pieces of language' (Sinclair, 1996), and reflecting actual language use according to internal and external criteria (McEnery et al., 2006; McEnery & Hardie, 2012).

Within the HEREDITARY project, domain-specific corpora of expert-authored texts for expert audiences (i.e., researchers, clinicians, and students), underpin the identification, extraction, systematisation, and analysis of terminological data. Semi-automatically extracted and manually reviewed terms and knowledge-rich contexts (Meyer, 2001) allow terminologists to infer expert conceptualisations, as described in Section 4.

3.1. Multilingual Comparable Corpora: Compilation Criteria

The multilingual framework is central to both the ongoing project, which integrates data, institutions, and audiences within an international context, and the ontology-based dictionary presented in this paper. It broadens the user base, helps to mitigate conceptual inconsistencies arising from linguistic gaps, and introduces challenges in aligning concepts across languages and disambiguating polysemous or ambiguous abbreviations.

Although the English corpus (HEREDITermCorpus_en)[8] was compiled first, serving as the

[6] https://www.nlm.nih.gov/research/umls/index.html
[7] The Joint Commission is an independent non-profit organisation based in the United States that accredits healthcare organisations and programmes according to quality and patient safety standards. Its recommendations and guidelines are widely recognised in the healthcare sector and aim to improve the safety and effectiveness of clinical communication and care. See https://www.jointcommission.org/en-us
[8] https://zenodo.org/records/16968962

reference corpus and proving to be the most productive, the findings presented here are based on a multilingual comparable corpus approach comprising English, Portuguese and Italian specialised corpora. The corpora cover two main disease categories within the gut–brain axis framework: (i) neurological and neurodegenerative disorders, and (ii) gastrointestinal, psychiatric, and metabolic conditions linked to gut–brain axis dysfunction. This framework reflects the growing recognition of the role of the microbiota-brain interaction in disease mechanisms.

The neurological and neurodegenerative group includes Alzheimer's disease (AD), Parkinson's disease (PD), multiple sclerosis (MS), amyotrophic lateral sclerosis (ALS), and frontotemporal dementia (FTD). The second group encompasses irritable bowel syndrome (IBS), inflammatory bowel disease (IBD), autism spectrum disorder (ASD), attention-deficit/hyperactivity disorder (ADHD), major depressive disorder (MDD), anxiety disorders (ANX), diabetes mellitus (DM), and obesity (OB). For the present analysis, only data from the neurological and neurodegenerative group were considered.

Texts were manually selected according to (i) relevance and suitability for the domain (and subdomain) under analysis, (ii) open-access availability, and (iii) compatibility with Sketch Engine[9] format requirements. Additionally, corpus coverage and balance across subcorpora were enhanced through bootstrapping, enabling the automatic, yet strictly parameterised, dataset expansion from selected "seed words".[10] Seed words were chosen to represent both the target disease and core concepts of the gut–brain axis framework. For instance, the frontotemporal dementia (FTD) subcorpus was expanded using keywords such as "frontotemporal dementia", "frontal-temporal dementia", "gut–brain", "brain–gut", "gut brain axis", "microbiota", and "microbiome". Comparable sets of seed words were used for the remaining disease-specific subcorpora. For the Portuguese and Italian corpora, the expansion relied on equivalent disease- and gut–brain-axis-related terms in each language. The resulting automatically retrieved texts were then manually revised and validated.

3.2. Corpus Description and Metadata Encoding

From a quantitative perspective, the three comparable corpora considered in this study are as follows:

- The British English dataset (HEREDITermCorpus_en) comprises 1,060 documents (704 abstracts[11] and 356 full papers), 234,215 sentences, 4,132,486 words and 6,029,603 tokens.
- The European Portuguese dataset (HEREDITermCorpus_pt)[12] comprises 126 full papers, 100,610 sentences, 1,999,301 words and 2,665,436 tokens.
- The Italian dataset (HEREDITermCorpus_it)[13] comprises 206 full papers, 144,009 sentences, 3,112,322 words and 4,140,747 tokens.

The corpus compilation process was rigorously documented following established good practices in Corpus Linguistics (see Baker et al., 2006). The documents were annotated with metadata concerning the topic or subdomain, text type, year of publication, and source identifier, e.g., AD-MT2020-UMAIA, where AD refers to Alzheimer Disease, MT to Master Thesis, 2020 to the year of publication and UMAIA to the MAIA University digital repository.

All metadata are available in open access. Transparent and systematic metadata encoding ensures resource reusability, facilitates result interpretation, and improves information filtering and retrieval efficiency.

For corpus compilation, annotation and analysis, Sketch Engine was used as a corpus management and NLP tool, chosen for its integrated functionalities, including POS tagging, customisable metadata annotation, statistical measures for assessing term relevance and distribution, and a flexible query language (CQL) for analysing co-occurrence patterns.

4. Trilingual Terminology Extraction

[9] https://www.sketchengine.eu/
[10] By "seed terms" we refer to domain-representative keywords used as search terms to guide and control the automatic expansion of the corpus.
[11] The subcorpus of abstracts was used as a complementary component of the main corpus, as abstracts are less terminologically dense due to their lower lexical density.
[12] https://zenodo.org/records/16969241
[13] https://zenodo.org/records/20328753

Table 1 Nine (9) variants and six (6) abbreviations of the term "alpha-synuclein"

Full form	Freq. in corpus	Abbreviation	Freq. in corpus
alpha-synuclein	1140	aSyn	373
α-synuclein	1952		
a-synuclein	452	a-syn	238
alpha-synuclein	1140		
α-synuclein	1952	αSyn	420
αSynuclein	71		
α-synuclein	1952	SNCA	331
alpha-synuclein	1140	α-syn	965
alphasynuclein	118	α-synuclein	1952

4.1. Terminology Extraction from HEREDITermCorpus_en

HEREDITermCorpus_en served as the source corpus for identifying abbreviations enclosed in brackets. To extract abbreviations appearing in brackets, we applied advanced searches formulated in Corpus Query Language (CQL), incorporating regular expressions (regex).

The regex developed for this study aims to capture polylexical terms whose full form precedes the corresponding abbreviation in brackets. By polylexical terms, we refer to specialised lexical combinations whose language-dependent syntactical structures follow the grammatical restrictions of their respective languages and should be treated as single lexical units, as they each designate a concept. The search strategy reflects conventional writing practices, whereby polylexical terms are typically introduced in full, followed by their abbreviated form in parentheses at first occurrence, after which, the abbreviated form generally replaces the full form in subsequent mentions. Although this convention is highly productive in scientific writing, abbreviations and their full forms may also occur in other textual configurations not captured by the regex used in this study. The proposed strategy therefore does not aim to exhaustively identify all textual configurations in which abbreviations and/or their corresponding full forms may occur.

The syntax of regex (1) CQL `"\(""(.){3,4}[[:alpha:]]*""\)"` identifies bracketed abbreviations composed of at least three or four capitalised initials. Using regex (1) – parameterised to capture exact forms: `default attribute=word –`, we obtained a concordance of 36,218 alphanumeric forms enclosed in round brackets. A context filter was simultaneously added to capture occurrences of the form "is" within a span of five forms on either side of the KWIC (keyword in context) (= abbreviation in brackets). This filter reduced, although not entirely, the number of digit-only forms and retrieved 1,002 bracketed abbreviations together with their surrounding knowledge-rich contexts (KRC, according to Meyer, 2001) about the concept designated by the abbreviation of the candidate term.

Appendix A (see Figure 20) shows the first 26 lines of the concordance returned by CQL (1).

The 1,002 contexts were manually curated, identifying 759 multi-word forms to the left of the KWIC corresponding to the initials of the bracketed abbreviations. After the removal of duplicate forms, 447 candidate terms remained, exhibiting variation in both the full form and the acronym or abbreviation; e.g., the full form of the term "*alpha-synuclein*" displays nine variants and six abbreviations or short forms (see Table 1).

After identifying formal variants – e.g., with or without a hyphen between the elements constituting the morphosyntactic structure of the full form or the abbreviation – we identified 354 unique abbreviation forms (*types*[14]).

[14] According to ISO, *type* is an "unique sequence of characters (3.5) in a text corpus (3.25)" (ISO, 2025).

Table 2 Five (5) abbreviations and corresponding full forms, in English

en	Abbreviation	Freq. in corpus	Full form	Freq. in corpus
	DLB	49	dementia with Lewy bodies	63
	LBD	44	Lewy body dementia	62
	APP (1)	730	atypical antipsychotics	5
	APP (2)		β-amyloid precursor protein (1)	4
			amyloid precursor protein (2)	75
	PPA	116	primary progressive aphasia	159

For this study, we focus on five abbreviations and their corresponding full forms, which we consider to be candidate biomedical terms (see Table 2), i.e., provisional terms extracted from the corpus and subsequently subjected to terminological analysis (ISO, 2025) and expert validation. Their selection was informed by ongoing work[15] conducted within the HEREDITARY project. These examples are representative of the ambiguity types discussed in this paper, which occur consistently across the corpora.

The selection of these five forms was motivated by ambiguity arising from variation in both at the level of the full forms and in the sharing of the same abbreviation by terms whose semantic load suggests that they designate different concepts, as can be observed in Table 2.

4.2. Terminology Extraction from HEREDITermCorpus_pt

For abbreviation extraction from the Portuguese corpus, both simple and advanced searches were used. Although regex (1) returned results comparable to those from the English corpus, the advanced search was restricted to the forms of interest identified in the English corpus (Table 2).

The following advanced search (2), was designed to capture the abbreviation of the Portuguese equivalent of "dementia with Lewy bodies":

Advanced search (2):

```
Query type=word
Part-of-Speech = any
word = Lewy
Filter context=lemma context
Only keep lines with:
       all of demência
```

The concordance results obtained through Advanced search (2) revealed prepositional variation in the full forms of the terms "*demência com corpos de Lewy*" and "*demência dos corpos de Lewy*", as well as variation arising from the cross-linguistic use of the English abbreviation. Although the Portuguese term is written as "*demência de corpos de Lewy*", the English abbreviation is retained: "dementia with Lewy bodies" (DLB) (Table 3).

Further types of variation found in the HEREDITermCorpus_pt are discussed in Section 5.

[15] https://zenodo.org/records/18475733

Table 3 Example of prepositional variation in the Portuguese full form and cross-linguistic use of the English abbreviation

pt	Abbreviation	Freq. in corpus	Full form	Freq. in corpus
	DCL	104	demência com corpos de Lewy (1)	17
			demência dos corpos de Lewy (2)	2
	DLB	5	demência de corpos de Lewy	11

Table 4 Example of prepositional variation in the Italian full form and cross-linguistic use of the English abbreviation

it	Abbreviation	Freq. in corpus	Full form	Freq. in corpus
	LBD	217	demenza a corpi di Lewy (1)	57
	BDL	2		
	DLB	48		
			demenza con corpi di Lewy (2)	11
	-		demenza da/dei corpi di Lewy	2

4.3. Terminology Extraction from HEREDITermCorpus_it

The same methodology was applied to the Italian corpus. The advanced search parameters followed the discussion described in the Portuguese case study.

The Italian corpus likewise contains the terms "*demenza a corpi di Lewy*" and "*demenza con corpi di Lewy*", which, despite prepositional variation in their full forms, use English-derived abbreviations, namely "Lewy body dementia" (LBD) and "dementia with Lewy bodies" (DLB) (Table 4).

The advanced search also reveals, within the Italian corpus, the existence of:

- Additional full forms, including "*demenza da corpi di Lewy*" and "*demenza dei corpi di Lewy*", which, however, occur without any corresponding abbreviation.
- A single occurrence of the abbreviation BDL, associated with the full form "*demenza a corpi di Lewy*". As it appears only once within a single paper, we consider this instance is either an authorial usage or a typographical error.

Furthermore, considering the alternation "*dementia*" identified in the corpus, we also parametrised the advanced search filtering the co-occurrence of "*Lewy*" with the lemma "*malattia*" (rather than "*demenza*").

We identified prepositional variation in the full forms "*malattia a corpi di Lewy*" and "*malattia da/dei corpi di Lewy*". However, only the English-derived abbreviation LBD (borrowed from English "Lewy body dementia") was found, occurring once with "*malattia a corpi di Lewy*".

5. Systematisation of Data Extracted from Multilingual Corpora

In Appendix B (see Table 5), we systematised the full forms and their corresponding abbreviations to (i)

establish a typology of variation in the full forms and/or their corresponding abbreviations, and (ii) analyse the elements composing the morphosyntactic structure of each candidate term. This aims to detect ambiguity arising from intra- and interlingual denominative variation, as well as from mono- or polyreferentiality. For example, APP(1) in Portuguese establishes a mono-referential relation with PPA in English (@en) and APPL in Italian (@it), and a polyreferential relation with APP(2) in Portuguese (@pt) and APP(1) in English (@en). Thus, regardless of whether they designate the same concept, variation occurs in both abbreviated and full forms within and across intra and interlingual contexts, as shown in the following subsections.

The systematisation of the full forms of the candidate terms and their corresponding abbreviations reveals 8 patterns of abbreviation use:

(1) intralingual use in Portuguese of both Portuguese and English abbreviations;
(2) intralingual use in Italian of English abbreviations;
(3) interlingual use (pt-en) of the same abbreviation for terms designating different concepts;
(4) interlingual use (pt-en-it) of the same abbreviation for terms designating the same concept;
(5) interlingual use (pt-en-it) of different abbreviations and different full forms for the same concept;
(6) intralingual use, in all three languages, of the same abbreviation with different full forms referring to the same concept;
(7) intralingual use, in all three languages, of different abbreviations and different full forms referring to the same concept;
(8) interlingual use (pt-en) of the same abbreviation for terms designating different concepts.

5.1. Intralingual Denominative Variation

The examples provided in Appendix B (see Table 5) are identified in this subsection according to the type of variation observed in the same language. Out of the intralingual denominative variation, we have identified two types of referential ambiguity and one type of lexical ambiguity, as follows.

5.1.1 Monoreferential Ambiguity (en-pt-it)

(7) In all three languages, the intralingual use of **different abbreviations** and **different full forms** designating the **same concept**:

DLB@pt =*demência **de** corpos de Lewy*
DCL@pt =*demência **com** corpos de Lewy* (1); *demência **dos** corpos de Lewy* (2)

DLB@en = dementia with Lewy bodies
LBD@en = Lewy body dementia

LBD@it= *demenza **a** corpi di Lewy*
DLB@it= *demenza **con** corpi di Lewy* (2)
BDL@it= *demenza **a** corpi di Lewy*

(6) In all three languages, the intralingual use of the **same abbreviation** with **different full forms** for terms designating the **same concept**:

DLC@pt
(1) *demência com corpos de Lewy*
(2) *demência dos corpos de Lewy*

APP(2)@en
(1) β-amyloid precursor protein
(2) amyloid precursor protein

APP @it
(1) *proteina precursore dell'amiloide*
(2) *precursore della proteina beta-amiloide*
(3) *precursore di Aβ*

DLB@it
(1) *demenza a corpi di Lewy*
(2) *demenza con corpi di Lewy*

Although monoreferentiality generally does not generate conceptual ambiguity, as each term designates a single concept, the examples above reveal clear inconsistencies, affecting both abbreviation and full forms. Such variation occurs not only across languages (see Section 5.2.1), but also within the same language. This lack of standardisation in clinical and scientific texts may result in inconsistencies in disease recording and monitoring, leading to inaccuracies in clinical data and potentially compromising patient treatment.

5.1.2 Polyreferential Ambiguity (pt)

(3) In an intralingual context, the use of the **same abbreviation** for **different full forms** designating **different concepts**:

APP(1)@pt = *afasia primária progressiva*
APP(2)@pt = *proteína precursora amilóide*

APP(1)@en= atypical antipsychotics
APP(2)@en=amyloid precursor protein (2)

5.1.3 Lexical Ambiguity (pt-it)

(1) In an intralingual context (pt), the use of **both abbreviations** corresponding to the **Portuguese and English terms** entails **different abbreviations** and **different full forms** designating the **same concept**:

DLC@pt = *demência com corpos de Lewy* (1)
DLB@pt = *demência de corpos de Lewy*

(2) In intralingual context (it), the use of two abbreviations corresponding to English terms:

DLB@it = *demenza a corpi di Lewy* (1); *demenza con corpi di Lewy* (2)
LBD@it = *demenza a corpi di Lewy*

5.2. Interlingual Denominative Variation

The examples provided in Appendix B (see Table 5) are now identified according to the type of variation observed across the three languages. Regarding interlingual denominative variation, we have identified two types of referential ambiguity, as follows.

5.2.1. Monoreferential Ambiguity (pt-en-it)

(4) In an interlingual context (pt-en-it), the use of the same abbreviation for different full forms of terms designating the same concept:

APP(2)@pt = *proteína precursora amilóide*
APP(2)@en = β-amyloid precursor protein (1)
APP@it = *precursore di Aβ* (3)

(5) In an interlingual context (pt-en-it), the use of different abbreviations and different full forms for terms designating the same concept:

APP(1)@pt = *afasia primária progressiva*
PPA@en = primary progressive aphasia
APPL@it = *sindrome dell'afasia primaria progressiva* → different full form

DCL@pt = *demência dos corpos de Lewy* (2)
DLB@en = dementia with Lewy bodies
BDL@it = *demenza a corpi di Lewy*

5.2.2. Pluri-referential Ambiguity (pt-en)

(8) In an interlingual context (pt-en), the use of the same abbreviation for different full forms of terms designating different concepts:

APP(1)@pt = *afasia primária progressiva*
APP(2)@en = β-amyloid precursor protein (1)

6. Conceptual Representation of Ambiguous Biomedical Abbreviations in Knowledge Resources

This section presents a comparative analysis of how concepts associated with variation in abbreviation use, including potentially ambiguous cases, are represented across selected biomedical resources. The selection is based on recurrent patterns of abbreviation use identified in the corpus, namely (i) different abbreviations designating the same concept, (ii) a single abbreviation designating different concepts, and (iii) cross-linguistic variation in abbreviation use.

The analysis focuses on the full forms associated with the abbreviations DLB, LBD, APP, and PPA discussed in previous sections, examining how their corresponding concepts are represented across the selected resources. Data were collected from three major biomedical resources: the National Cancer Institute Thesaurus (NCIt), Medical Subject Headings (MeSH) and SNOMED CT. When a concept was not be found, additional searches were conducted via BioPortal[16] to determine whether it was represented in other ontologies. For each concept, the following elements were recorded whenever available: (i) natural language definition, (ii) preferred term, (iii) alternative designations associated with the concept corresponding to what this study considers denominative or lexical variation (e.g., synonyms, index terms, or term variants, depending on the resource), and (iv) its position within the resource's conceptual organisation.

By comparing term variants, definitions and structural relations across these resources, this section identifies convergences and divergences in concept representation. This comparison clarifies the conceptual relations underlying abbreviation ambiguity and supports the ontology-based approach proposed in this study.

6.1. dementia with Lewy bodies (DLB) and Lewy body dementia (LBD)

This subsection examines the concept designated by the terms "dementia with Lewy bodies" and "Lewy body dementia", associated with the abbreviations DLB and LBD in the HEREDITermCorpus_en. It analyses how this concept is represented across the selected resources.

6.1.1. MeSH

In MeSH, this concept is represented by the preferred term "Lewy body disease" (MeSH Unique ID: D020961).[17] Figure 1 provides an overview of its representation.

[16] https://bioportal.bioontology.org
[17] RDF Unique Identifier: https://id.nlm.nih.gov/mesh/M0015964.html

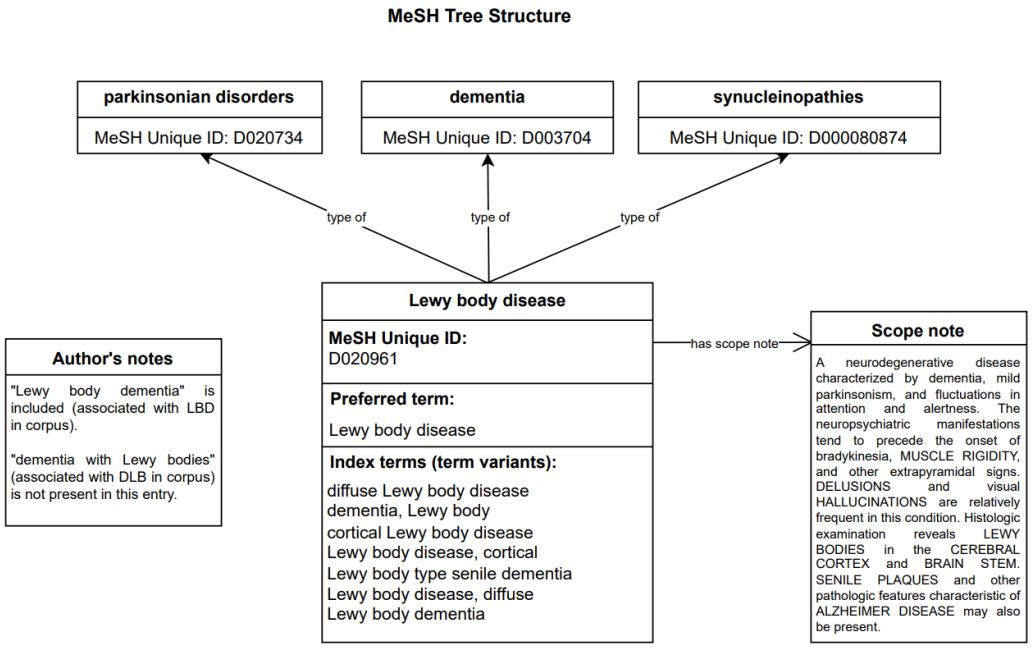

Fig. 1 <Lewy body disease> in MeSH

MeSH is a controlled vocabulary primarily developed for indexing biomedical literature. Its practical, indexing-focused design is evident in the use of scope notes, which emphasise descriptive information rather than concise intensional definitions (ISO, 2022). Each entry includes a scope note outlining the clinical features associated with the condition. The label "dementia, Lewy body" appears alongside "Lewy body dementia" because MeSH prioritises controlled normalisation and hierarchical organisation over natural language usage. It applies syntactic inversion to group diseases under a broader category (e.g., dementia), improving indexing consistency and enabling more efficient retrieval across different terminological variants in biomedical literature.

While the term "Lewy body dementia" (associated in the corpus with the abbreviation LBD) appears in this MeSH entry, "dementia with Lewy bodies" (associated with the abbreviation DLB) is not included.

<Lewy body disease>[18] appears in multiple branches of the MeSH tree, including <Parkinsonian disorders>, <Dementia>, and <Synucleinopathies>, reflecting its organisation for efficient information retrieval.

6.1.2. NCIt

In NCIt, the preferred term "Lewy body dementia" (Code: C84826)[19] is associated in the corpus with the abbreviation LBD. It includes synonyms (term variants), such as "dementia with Lewy bodies" grouping corpus-attested variants under a single concept. Figure 2 shows the NCIt entry, including its term variants, superordinate concept and definition. Although presented as a definition, the formulation follows a genus–differentia structure but extends beyond the typical standard intensional definitions by incorporating additional descriptive and clinical details.

NCIt and MeSH adopt different preferred labels ("Lewy body dementia" and "Lewy body disease") for what appears to be the same concept, with only partial overlap in their term variants. This variation reflects differences in normalisation practices across resources.

[18] As a graphical convention, concepts are written with angular brackets, and terms in quotation marks.
[19] http://purl.obolibrary.org/obo/NCIT_C84826

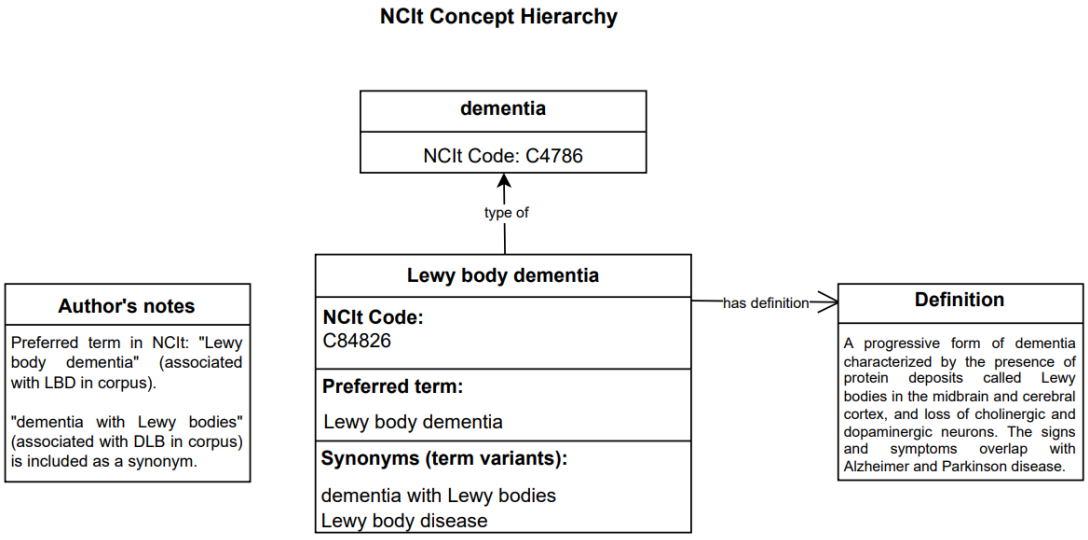

Fig. 2 <Lewy body dementia> in NCIt

6.1.3. SNOMED CT

In SNOMED CT, unlike in MeSH or NCIt, a distinction is made between two related concepts. The concept designated by the terms "diffuse Lewy body disease (disorder)" and "Lewy body disease" (SCTID: 80098002)[20] is distinguished from "dementia due to Lewy body disease (disorder)" (SCTID: 1363185006).[21] The latter is modelled as a subtype of both <Diffuse Lewy body disease (disorder)> and <Dementia (disorder)>,[22] reflecting SNOMED CT's polyhierarchical structure where a concept may have multiple superordinate concepts.

As illustrated in Figure 3, the abbreviation LBD, which in the corpus corresponds to "Lewy body dementia", is used in SNOMED CT as a variant of "Lewy body disease", thus referring to a different concept.

While MeSH treats "Lewy body dementia" and "diffuse Lewy body disease" as index terms for the same concept, and NCIt groups "Lewy body dementia", "Lewy body disease", and "dementia with Lewy bodies" as exact synonyms, SNOMED CT represents "diffuse Lewy body disease (disorder)" and "dementia due to Lewy body disease (disorder)" as distinct but related concepts.

These findings highlight differences in conceptual granularity across the analysed resources and have implications for how abbreviations should be interpreted. In particular, SNOMED CT records LBD as a variant of "Lewy body disease", whereas in the corpus LBD refers to "Lewy body dementia", thereby designating a different concept.

6.2. atypical antipsychotics (APP$^{(1)}$) and (β-)amyloid precursor protein (APP$^{(2)}$)

This subsection examines the concept designated by terms "atypical antipsychotics" and "β-amyloid precursor protein" (or "amyloid precursor protein"), both associated with the abbreviation APP in the HEREDITermCorpus_en, and analyses how they are represented across the selected resources.

6.2.1. atypical antipsychotics (APP$^{(1)}$)

i) MeSH

MeSH contains no concept corresponding directly to the term "atypical antipsychotics". Although attested in the HEREDITermCorpus_en with the abbreviation APP, the term does appear neither as a

[20] conceptId1=80098002
[21] conceptId1=1363185006
[22] conceptId1=52448006

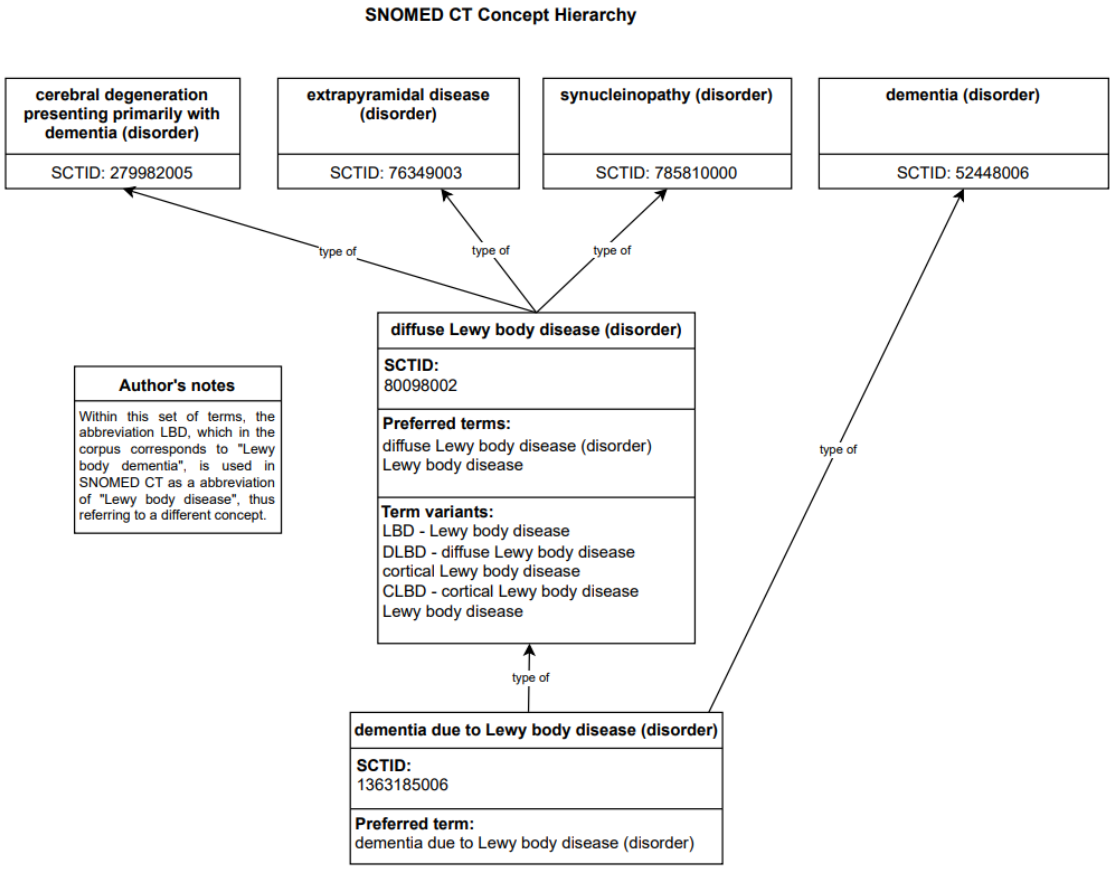

Fig. 3 <Diffuse Lewy body disease (disorder)> and <Dementia due to Lewy body disease (disorder)> in SNOMED CT

preferred term nor as an index term. Instead, MeSH organises this domain under broader categories such as <Antipsychotic agents> (Unique ID: D014150)[23] or through entries corresponding to individual substances, without explicitly encoding the distinction between typical and atypical antipsychotics. The concept <Antipsychotic agents> is accompanied by a scope note describing these agents as

> "Agents that control agitated psychotic behavior, alleviate acute psychotic states, reduce psychotic symptoms, and exert a quieting effect. They are used in SCHIZOPHRENIA; senile dementia; transient psychosis following surgery; or MYOCARDIAL INFARCTION; etc. These drugs are often referred to as neuroleptics alluding to the tendency to produce neurological side effects, but not all antipsychotics are likely to produce such effects. Many of these drugs may also be effective against nausea, emesis, and pruritus."

This scope note provides a general characterisation of antipsychotic substances rather than encoding the distinction between typical and atypical antipsychotics. As a description, it provides supplementary information about a concept without establishing strict conceptual boundaries. It may include contextual, pragmatic, or related information and is not required to differentiate the concept exhaustively from others. The absence of a distinct concept for "atypical antipsychotics" suggests that this classification, although attested in specialised discourse, is not systematically represented as a separate conceptual entity in MeSH. On the other hand, MeSH includes 16 indexed term variants for <Antipsychotic agents>, some of which overlap with the variants identified in NCIt and IOBC.

ii) NCIt

Similarly, NCIt contains no concept corresponding directly to the term "atypical antipsychotics". Instead,

[23] RDF Unique Identifier: https://id.nlm.nih.gov/mesh/D014150.html

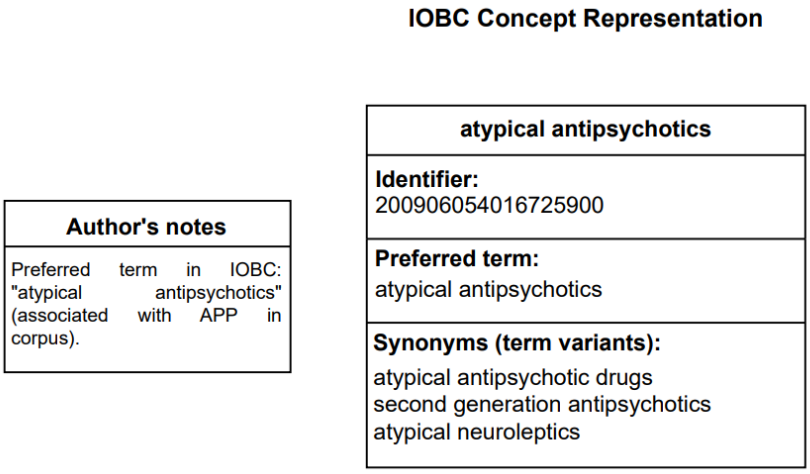

Fig. 4 <Atypical antipsychotics> in IOBC

the resource includes a broader concept[24] <Antipsychotic agent> (Code: C29710),[25] which encompasses this domain without explicitly distinguishing between typical and atypical antipsychotics. This concept is described as

> "Also known as neuroleptics, major tranquilizers, or antischizophrenics, natural or synthetic. Antipsychotic Agents relieve and control the symptoms of schizophrenic illness (hallucinations, delusions, dementia). Most antipsychotic agents interfere with various neurotransmitter functions, often blocking dopamine receptors, and induce diverse behavioral, endocrine, motor-kinetic effects."

As in MeSH, the absence of a distinct concept suggests that this classification, although attested in specialised discourse, is not systematically represented as a separate conceptual entity. Furthermore, NCIt includes 7 indexed term variants for <Antipsychotic agent>, several of which overlap with those in MeSH.

iii) SNOMED CT

Like MeSH and NCIt, SNOMED CT contains no concept corresponding directly to the term "atypical antipsychotics". Instead, the domain is represented through broader concepts such as <Anti-psychotic agent (substance)> (SCTID: 372482001) and <Medicinal product acting as antipsychotic agent (product)> (SCTID: 10784006), reflecting the distinction between agents and medicinal products.

As in the other resources, distinctions such as that between typical and atypical antipsychotics are not represented as separate concepts, although attested in specialised discourse.

iv) Interlinking Ontology for Biological Concepts (IOBC)

A search conducted in BioPortal indicates that the concept <Atypical antipsychotics> is represented in the Interlinking Ontology for Biological Concepts (IOBC)[26] (Identifier: 200906054016725900)[27] (see Figure 4).

The comparison shows that, although the term "atypical antipsychotics" is attested in the HEREDITermCorpus_en, it is not represented as a distinct concept in all the analysed resources. It is only represented in IOBC, where no abbreviation is associated with the concept. These differences may reflect the distinct purposes, scopes, and conceptual modelling principles of the resources.

[24] In NCIt, C29710 is labelled as "Class: Antipsychotic Agent". From our viewpoint, a class represents a concept, in the framework of OWL / RDFS Classes: "A common knowledge representation requirement is to describe concepts or "kinds of things". RDFS classes are sets used to describe concepts as categories of resources, typically for a particular domain." (Lacy, 2005, p. 113).
[25] NCIT_C29710
[26] Interlinking Ontology for Biological Concepts
[27] id/200906054016725900

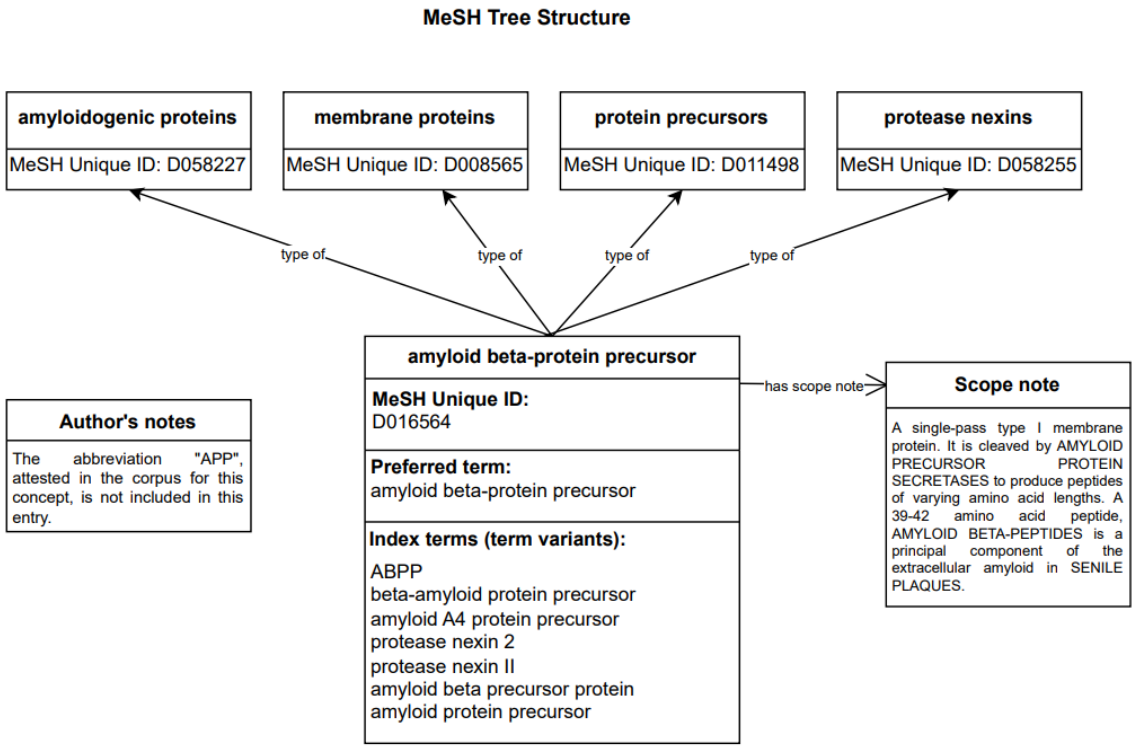

Fig. 5 <Amyloid beta-protein precursor> in MeSH

This discrepancy between resources motivated a deeper analysis of IOBC, revealing that <Atypical antipsychotics> is modelled as a subtype of <Antipsychotic>. The superordinate concept includes nine indexed term variants, several similar to those identified in MeSH and NCIt for <Antipsychotic agents> and <Antipsychotic agent> respectively, suggesting semantic overlap among the corresponding concepts. These findings are further discussed in Section 7 from a knowledge organisation perspective.

6.2.2. (β-)amyloid precursor protein (APP[(2)])

i) MeSH

This concept is represented in MeSH by "amyloid beta-protein precursor" (Unique ID: D016564),[28] which corresponds to "amyloid precursor protein" and "β-amyloid precursor protein" attested in the HEREDITermCorpus_en.

This entry includes a scope note describing the concept and several term variants.

In MeSH, ABPP is associated with the concept <Amyloid beta-protein precursor>, whereas APP, attested in the corpus, does not appear in this entry (see Figure 5).

Unlike the case of <Atypical antipsychotics>, this concept is clearly represented as a distinct conceptual entity in MeSH.

ii) NCIt

In NCIt, this concept is represented by the preferred term "amyloid-beta precursor protein" (Code C16285),[29] which corresponds to the terms "amyloid precursor protein" and "β-amyloid precursor protein" attested in the HEREDITermCorpus_en.

The entry includes a definition and numerous term variants, including several full forms and abbreviations most notably APP, for a total of 32 variants (see Figure 6).

[28] mesh/68016564
[29] ncit/C16285

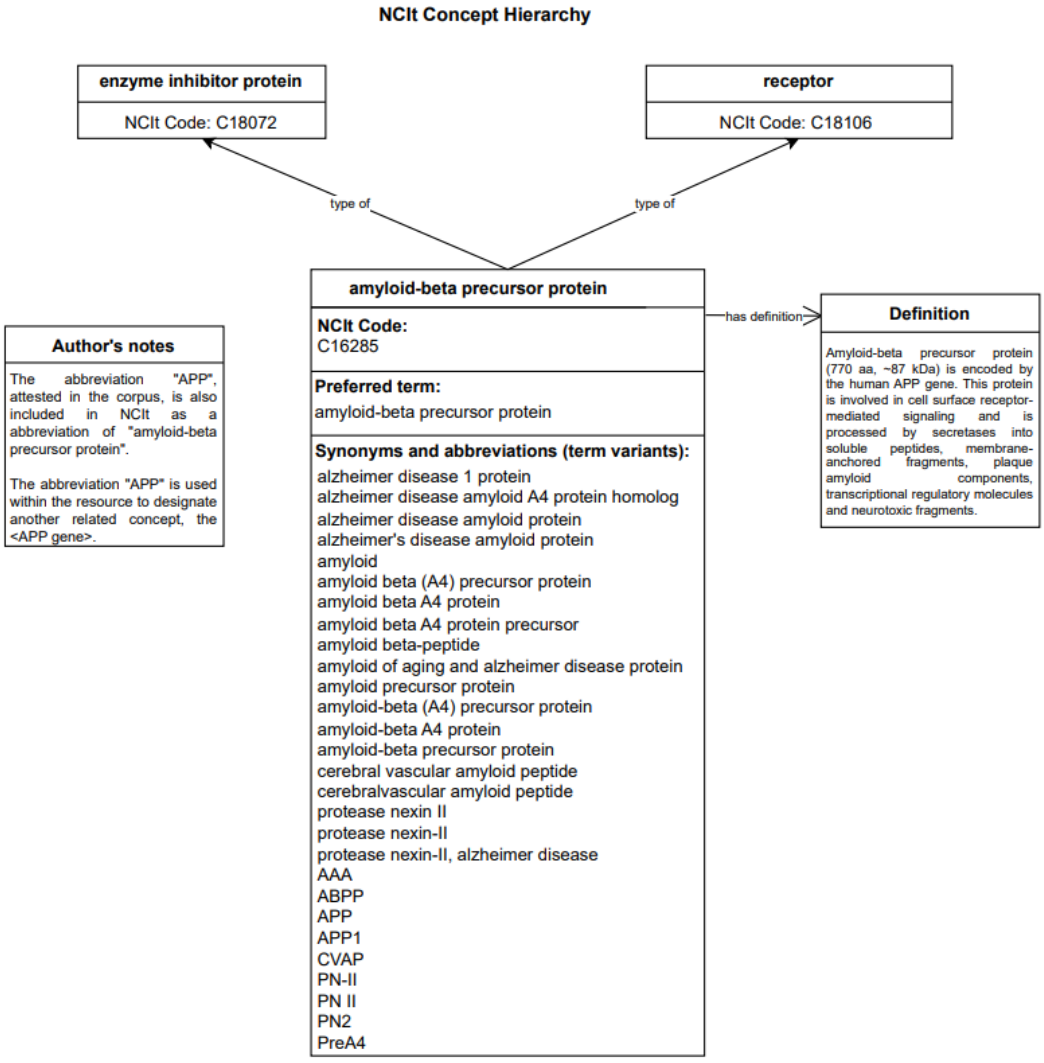

Fig. 6 <Amyloid-beta precursor protein> in NCIt

NCIt also includes a distinct concept <APP gene> (Code: C28537),[30] represented by the preferred term "APP gene" and the term variant "amyloid beta (A4) precursor protein gene", both of which incorporate the abbreviation APP. This shows that the abbreviation APP is used within the resource to designate both the protein and the gene, which are modelled as separate but related concepts.

This highlights a potential source of ambiguity, as the same abbreviation is used to designate closely related but distinct concepts within the resource.

iii) SNOMED CT

In SNOMED CT, no concept corresponding to "amyloid precursor protein" could be identified. Searches for related terms yield concepts such as <Autosomal dominant Alzheimer disease due to mutation of amyloid precursor protein (disorder)>, which presupposes the existence of the protein without modelling it as a distinct conceptual entity.

Accordingly, APP, attested in the corpus in association with "amyloid precursor protein", is not represented in this resource in relation to this concept.

6.3. Primary progressive aphasia (PPA)

This subsection examines the concept designated by "primary progressive aphasia", associated with the

[30] ncit/C28537

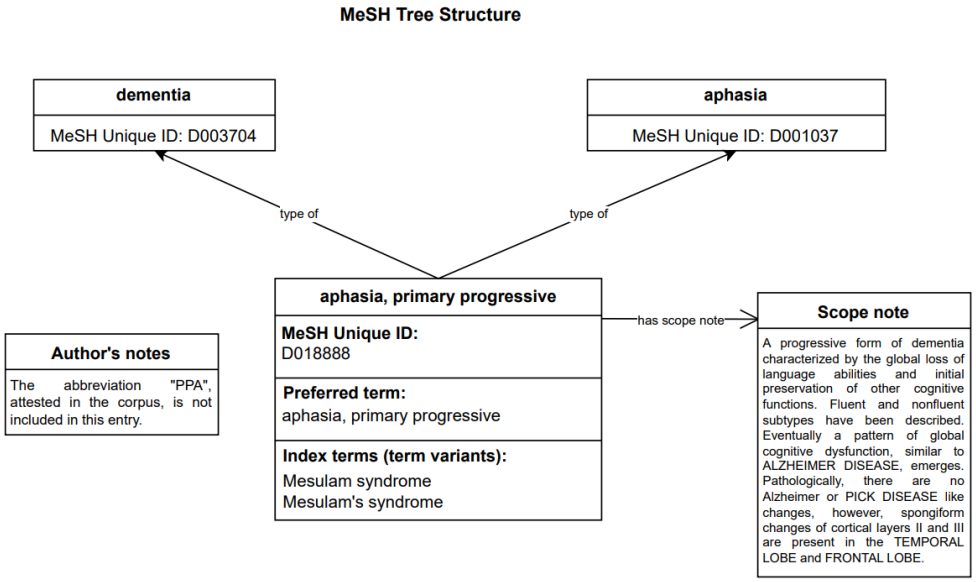

Fig. 7 <Aphasia, primary progressive> in MeSH

PPA in the HEREDITermCorpus_en. It analyses how this concept is represented across the selected resources.

6.3.1. MeSH

In MeSH, the relevant concept is represented by the preferred label "Aphasia, primary progressive" (Unique ID: D018888),[31] which corresponds to the indexing conventions of the resource.

As illustrated in Figure 7, PPA, although attested in the HEREDITermCorpus_en, is not indexed in this entry.

6.3.2. NCIt

In NCIt, the relevant concept is represented by the preferred term "primary progressive aphasia" (Code: C85024)[32] (see Figure 8), which corresponds to the term attested in the HEREDITermCorpus_en.

The definition of the concept in NCIt follows a concise genus–differentia structure and therefore aligns more closely with an intensional definition, in contrast with the more descriptive formulations observed in this resource for the other analysed concepts.

The entry does not include any term variants or abbreviations. This contrasts with corpus usage, where the abbreviation PPA is attested, and with other concepts analysed in this section, which include multiple variants. PPA is, however, attested in NCIt as a term variant for several unrelated concepts, including (1) <Acetaminophen-cysteine protein adduct>, (2) <Phenylpropanolamine measurement>, (3) <Phenylpyruvate measurement>, and (4) <Phosphonoacetic acid>. This reveals a discrepancy between corpus usage and the representation of abbreviations in the resource: although PPA is used in NCIt to designate unrelated concepts, it is not associated with <Primary progressive aphasia>.

6.3.3. SNOMED CT

In SNOMED CT, "primary progressive aphasia" is a term variant designating the concept whose preferred term is "primary progressive apraxia of speech (disorder)" (SCTID: 723124007).[33] As illustrated in Figure 9, this concept is integrated into a polyhierarchical structure and differs from the representations observed in MeSH and NCIt and its preferred term. Whereas those resources use a designation corresponding directly to "primary progressive aphasia", SNOMED CT selects a different preferred term,

[31] mesh/?term=D018888
[32] ncit/C85024
[33] conceptId1=723124007

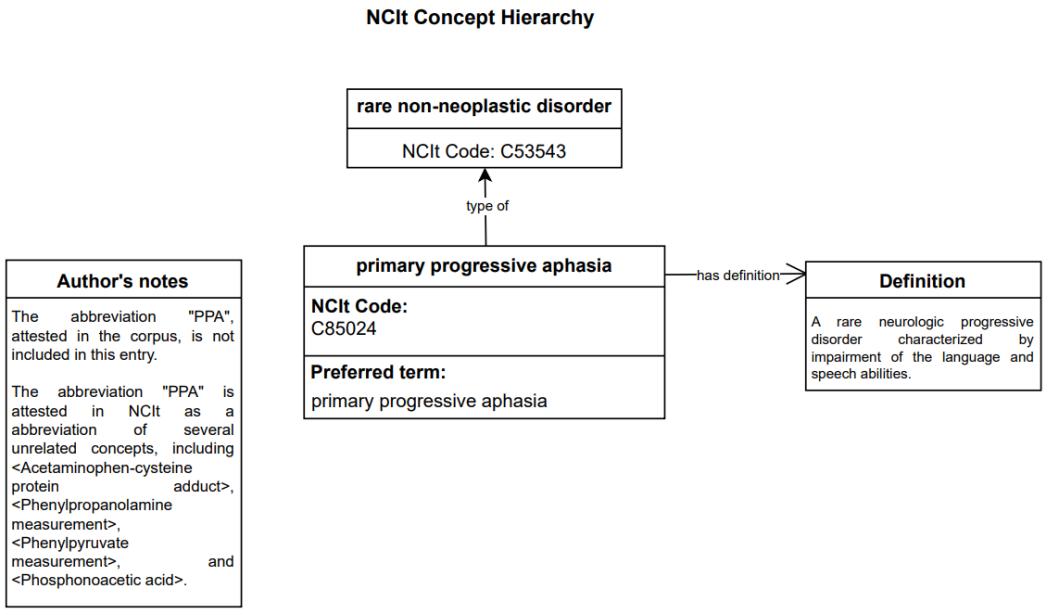

Fig. 8 <Primary progressive aphasia> in NCIt

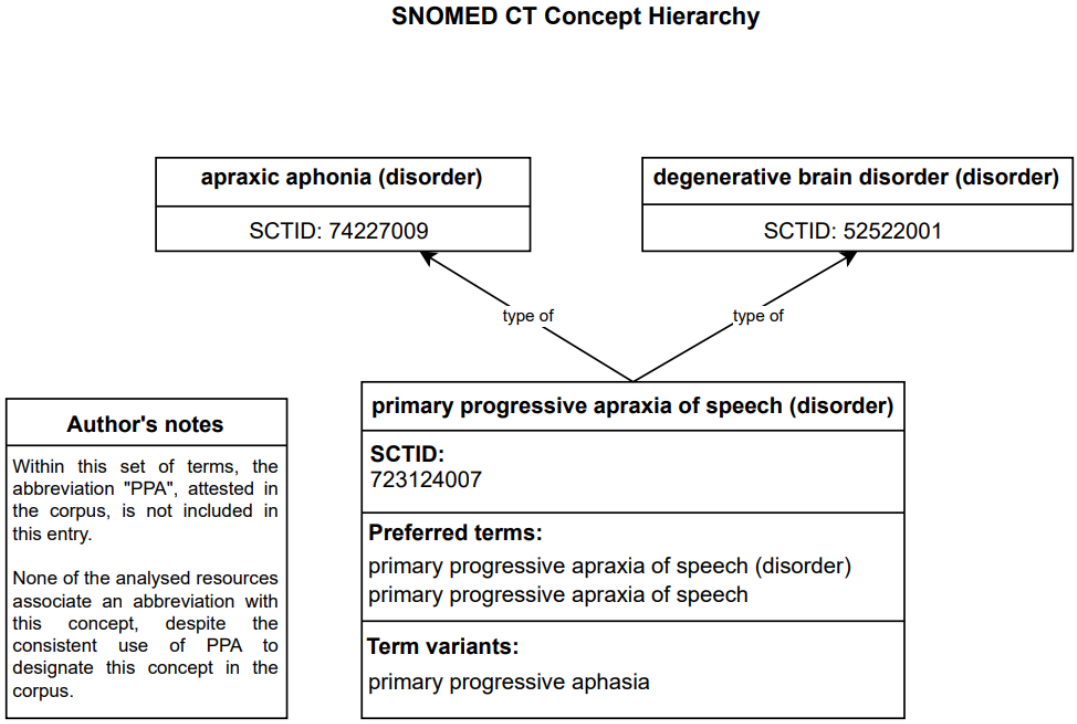

Fig. 9 <Primary progressive apraxia of speech (disorder)> in SNOMED CT

namely “Primary progressive apraxia of speech (disorder)”. It further includes “Primary progressive apraxia of speech” as a term variant and “primary progressive aphasia” as an acceptable term variant.

Notably, none of the analysed resources associate an abbreviation with this concept, despite the consistent use of PPA in the HEREDITermCorpus_en (cf. Section 4).

Overall, combining the linguistic analysis of corpus-attested abbreviations and their full forms with the conceptual analysis of their representations in biomedical resources revealed significant lexical variation, multilingual gaps, conceptual ambiguities, and disparate levels of granularity across platforms.

To address these systemic issues, the next sections introduce the development of an integrated terminological resource motivated by the research conducted within the HEREDITARY project. Combining an ontology and a multilingual dictionary, this resource aims to integrate conceptual and linguistic dimensions through a multilingual approach aligned, wherever possible, with pre-existing resources.

7. OntoAbbMed

This section presents the systematisation of the information associated with the concepts designated by the linguistic data extracted from the corpus (cf. Section 4) in an OWL[34] ontology, for which we used the Protégé ontology editor.[35]

At this initial stage, the ontology is openly available on GitHub[36] and comprises a total of 32 classes – corresponding, in our perspective, to concepts – whose definitions were formally axiomatized by means of 59 logical axioms and 155 assertions, and linguistically complemented with 298 annotation assertions. Among the 32 classes, 7 are equivalent classes (cf. Appendix C, Table 6).

In Appendix C (see Table 6), we highlight in bold the metrics associated with the assertions and annotations required for the systematisation of conceptual and linguistic information, i.e., for axiomatizing the hierarchical relations – although not exclusively – between the five concepts designated by the abbreviations and corresponding full forms analysed in Sections 5 and 6. The remaining values result from the importation of RDF vocabularies, namely Dublin Core Elements[37] and SKOS OWL1 DL[38], which we use to annotate linguistic and/or conceptual information (e.g., `dc:creator`[39]; `skos:related`:[40]).

The concepts included in the ontology were systematised according to the analysis presented in Section 6 and aligned with the consulted resources, enabling the integration of conceptual and linguistic information, particularly where equivalence between concepts is observed. The identification of equivalence results from the comparative analysis of the concepts and their respective shared hierarchical positions. The concept under analysis shares at least one superordinate concept, in cases of polyhierarchy.

In other cases, conceptual relations were inferred through the interpretation of natural language definitions provided in the biomedical resources – a strategy that could not be applied to resources that do not describe the characteristics of the concept. By way of example, the concept <APP Gene> results from the interpretation of a natural language definition where the associative relation established with one of the concepts focused on in this study, namely, <Amyloid beta-protein precursor>, can be identified, as represented below in Figure 10. From the five initial concepts, we identified a total of 30 hierarchically organised concepts by means of the subsumption relation, following the Aristotelian *proximum genus–species* perspective. We note, however, that not all concepts establish a direct subsumption relation, although the sharing of more distant superordinate concepts can still be observed due to polyhierarchy (i.e., a concept is subsumed by two or more superordinate concepts).

To all concepts, we associated conceptual and linguistic information identified in the biomedical resources (cf. Section 6), namely natural language definitions and the linguistic forms used as concept labels. Among these, one is commonly regarded as the preferred term, while the remaining ones correspond to term variants.[41]

7.1. Conceptual Organisation

For this study, we focused on the concepts designated by the terms associated with the abbreviation APP, which points both to polyreferentiality and monoreferentiality in multilingual and monolingual contexts

[34] https://www.w3.org/TR/owl-features/
[35] https://protege.stanford.edu/
[36] OntoAbbMed
[37] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[38] https://www.w3.org/2009/08/skos-reference/skos-owl1-dl.rdf
[39] http://purl.org/dc/elements/1.1/creator
[40] https://www.w3.org/TR/skos-reference/#semantic-relations
[41] Although some biomedical resources use the label *synonym*, we opted to adopt the terminology of our theoretical framework, namely *term variant*.

Fig. 10 Polyhierarchy of <Amyloid beta-protein precursor>

(cf. Section 5), as well as to usages attested in texts produced by domain experts but not indexed in biomedical resources (cf. Section 6).

In the example illustrated in Figure 10, we focus on the concept <Amyloid beta-protein precursor> (D016564; MeSH/NCBI),[42] which establishes a polyhierarchical relation with six concepts, i.e., it is a specific concept (i.e., a subtype) subsumed by <Receptor>, <Enzyme inhibitor protein>, <Amyloidogenic proteins>, <Enzyme precursors>, <Membrane proteins>, and <Protease nexins> – as semi-formally represented in Section 6.2.2.

Figure 11 illustrates the Protégé interface, particularly the section where logical restrictions can be axiomatized in Manchester OWL DL Syntax[43] in order to describe the concept of interest. The polyhierarchy of the concept <Amyloid beta-protein precursor> (D016564; MeSH) was axiomatized (SubClassOf) through the intersection of four superordinate concepts, namely <Amyloidogenic proteins>, <Enzyme precursors>, <Membrane proteins>, and <Protease nexins>. By declaring the intersection between these four concepts, HermiT OWL reasoner[44] infers four superordinate concepts, a Protégé editor plugin that enables the visualisation of additional information both in the concept description[45] interface (Figure

[42] mesh/68016564
[43] https://www.w3.org/TR/owl2-manchester-syntax/
[44] http://www.hermit-reasoner.com/
[45] The label *description* used in this section aligns with the terminology adopted in the Protégé interface, as illustrated in Figure 11, where the formal

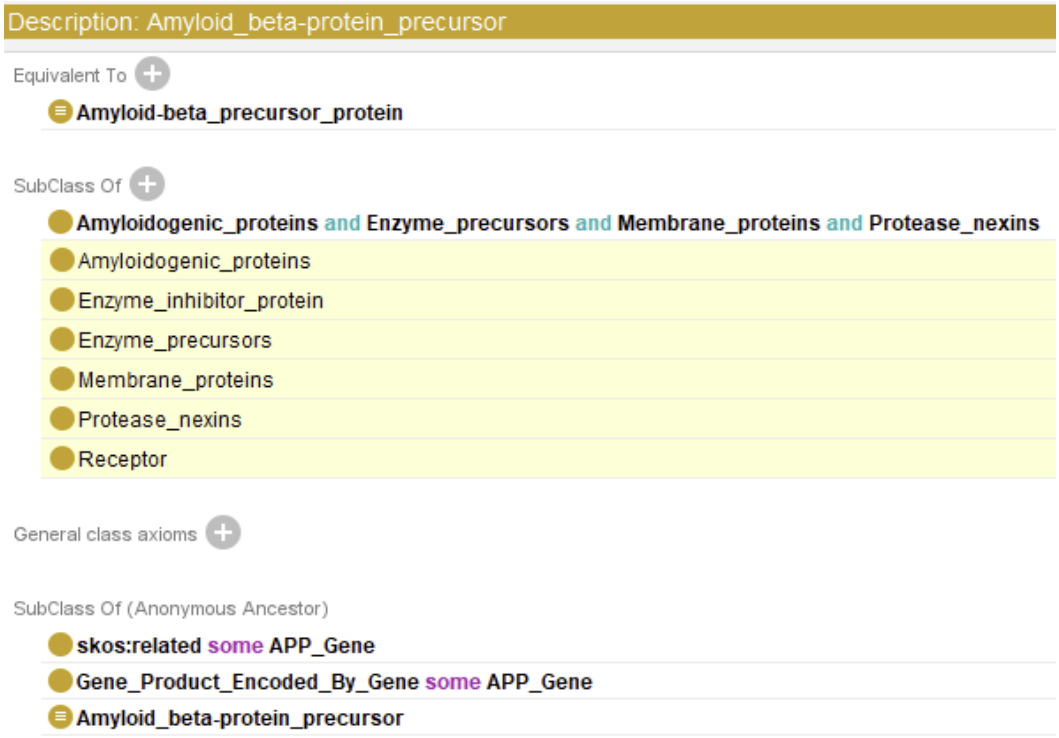

Fig. 11 Formal definition of <Amyloid beta-protein precursor> and HermiT inferences

11) (the inferred concepts are highlighted in yellow) and in the form of an OWLViz graph[46] (cf. Figure 14).

Through the reasoner inference process, additional information is obtained indicating that this concept is also a subtype of <Enzyme inhibitor protein> and <Receptor>. In other words, it is polyhierarchically subsumed by six concepts. This additional information results from the axiomatisation of the equivalence relation (Equivalent To) with the concept <Amyloid-beta precursor protein> (C18265; NCIt EVS Explore[47]). The equivalence relation stems from the observation, in the source resources, that both concepts share the same term variants and exhibit semantic similarity in their natural language descriptions (see Figure 5 and Figure 6, in Section 6). Semantic similarity refers to how closely two pieces of information share the same meaning, even when they use different words or structures. In this context, similarity is understood in terms of shared concepts and characteristics present in specialised definitions or broader superordinate concepts. Importantly, in this paper, similarity is not established through computational measurements or numerical vector representations, but rather through conceptual relations reflected in authoritative terminological resources.

Finally, because of the equivalence relation, we obtain the inference (Anonymous Ancestor) that this concept establishes an associative relation with <APP Gene>, of type Gene_product_Encoded_By_Gene. The descriptor of the associative relation was modelled on MeSH, where "APP gene" is indexed to <Amyloid beta-protein precursor> as a "related term".

With Figure 12, we illustrate the source of the conceptual information "inherited" by the concept <Amyloid beta-protein precursor> (D016564) discussed thus far. As can be observed, the axiomatized declarations used to define the concept <Amyloid-beta precursor protein> (C18265) are at the origin of the complementary information logically inferred and added to concept D016564. Likewise, although C18265 is defined as a concept subsumed (SubClassOf) by two concepts, namely <Enzyme inhibitor protein> and

definition is axiomatized.
[46] https://protegewiki.stanford.edu/wiki/OWLViz
[47] NCIt EVS Explore

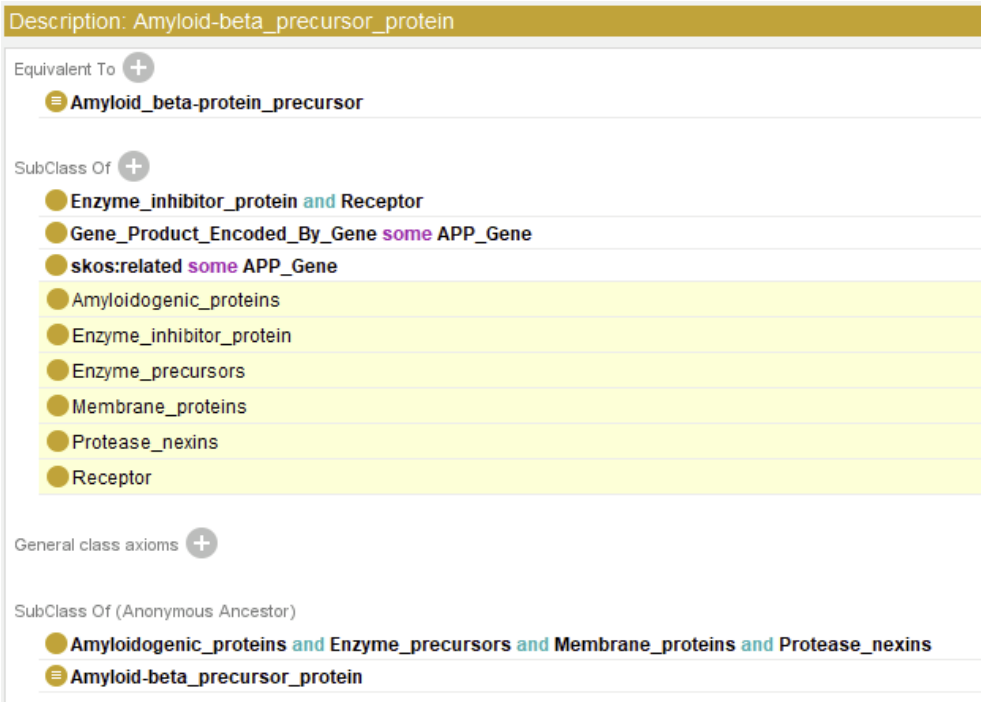

Fig. 12 Formal definition of <Amyloid-beta protein precursor> and HermiT inferences

<Receptor>, through the equivalence relation it inherits four superordinate concepts from <Amyloid beta-protein precursor> D016564, thereby also establishing a polyhierarchical subsumption relation involving six concepts.

As mentioned, the decisions and declarations used to axiomatize the conceptual relations that delimit the position of the concept by means of the relations it establishes with neighbouring concepts derive from the analysis of the systematisation – whether hierarchical or associative – observed in the authoritative resources, in this case MeSH/NCBI and NCIt EVS Explore.

To provide a clearer visualisation of the additional information obtained through HermiT reasoner inferences, we present two representations of the resulting polyhierarchy derived from the assertions defining the concept <Amyloid-beta precursor protein> (C18265), first axiomatized with only two superordinate concepts (Figure 13) – without inference – and subsequently with six superordinate concepts, thus highlighting the reasoner inference based on logical axioms (Figure 14).

Figure 13 illustrates an OWLViz graph where asserted equivalence is represented: equivalent concepts are shown as orange nodes, while the equivalence relation between them is represented by the green arc (and its inverse by the purple arc). The two green arcs represent the polyhierarchical subsumption relation inferred from the logical declaration: <Amyloid-beta precursor protein> is subsumed by the intersection of two concepts. The grey arcs represent hierarchical relations: (1) explicitly declared, without logical axioms, as in the case of <Receptor> and <Enzyme inhibitor protein>, and (2) inherited through the equivalence relation, as in the case of <Amyloid beta-protein precursor>.

As shown in Figure 14, the concept <Amyloid-beta precursor protein> (C16285), highlighted with a rectangle, exhibits a polyhierarchical subsumption involving six superordinate concepts. Although it is explicitly defined as subsumed by (SubClassOf) only two concepts (see Figure 12 and Figure 13), the HermiT reasoner infers four additional superordinate concepts through its equivalence with <Amyloid beta-protein precursor> (D016564).

In summary, the decision to bring together information available across several resources and grounded in different systematisation philosophies enables us to represent knowledge within a broader spectrum. To account for this breadth, the equivalence established between concepts – where such equivalence is observed – highlights the enrichment of conceptual information which, from a methodological perspective, facilitates the formal organisation of knowledge, as further discussed below.

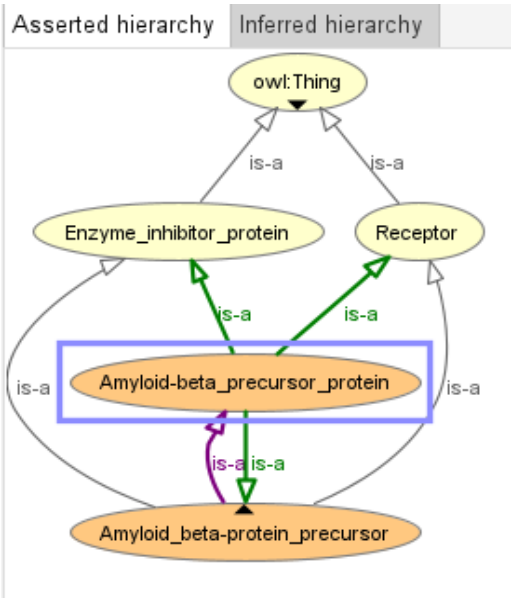

Fig. 13 Visualisation of the hierarchy obtained from the declarations, without inference

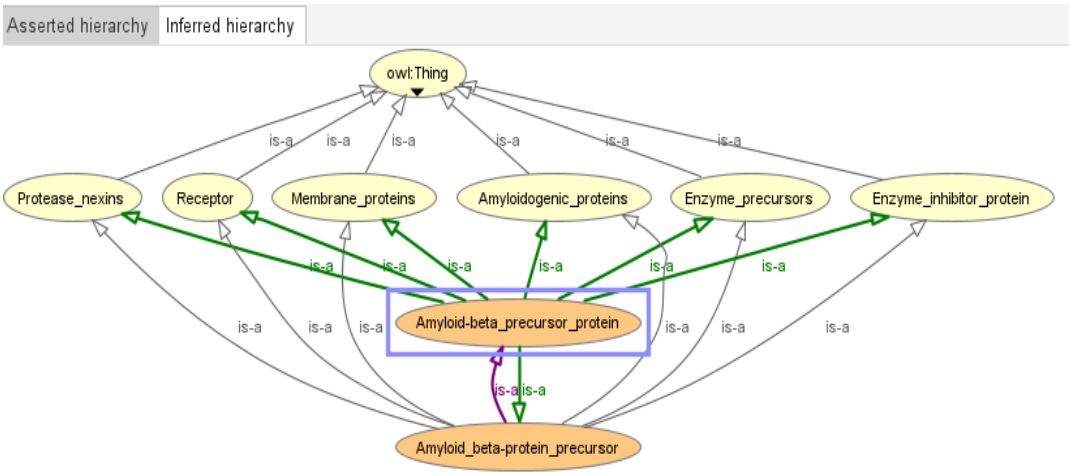

Fig. 14 Visualisation of the hierarchy inferred through the equivalence relation between D016564 and C18265

7.2. Linguistic Information: Term (skos:prefLabel) and Denominative Variants (skos:altLabel)

For the ontology, we focused on two types of linguistic expressions, namely those regarded by the biomedical resources as preferred terms and admissible terms – the latter, in our perspective, a variant of the preferred term – to account for the cases of ambiguity observed in specialised texts, resulting from:

- the use of terms sharing the same abbreviation;
- the occurrence of denominative variation in the full form of the abbreviation;
- terms and/or their respective abbreviations identified in the corpus whose linguistic forms are not indexed in the biomedical thesauri.

By annotating two distinct yet complementary levels of information associated with each concept, namely linguistic and conceptual information, we highlight the expressiveness of the double dimension of Terminology (cf. Section 2).

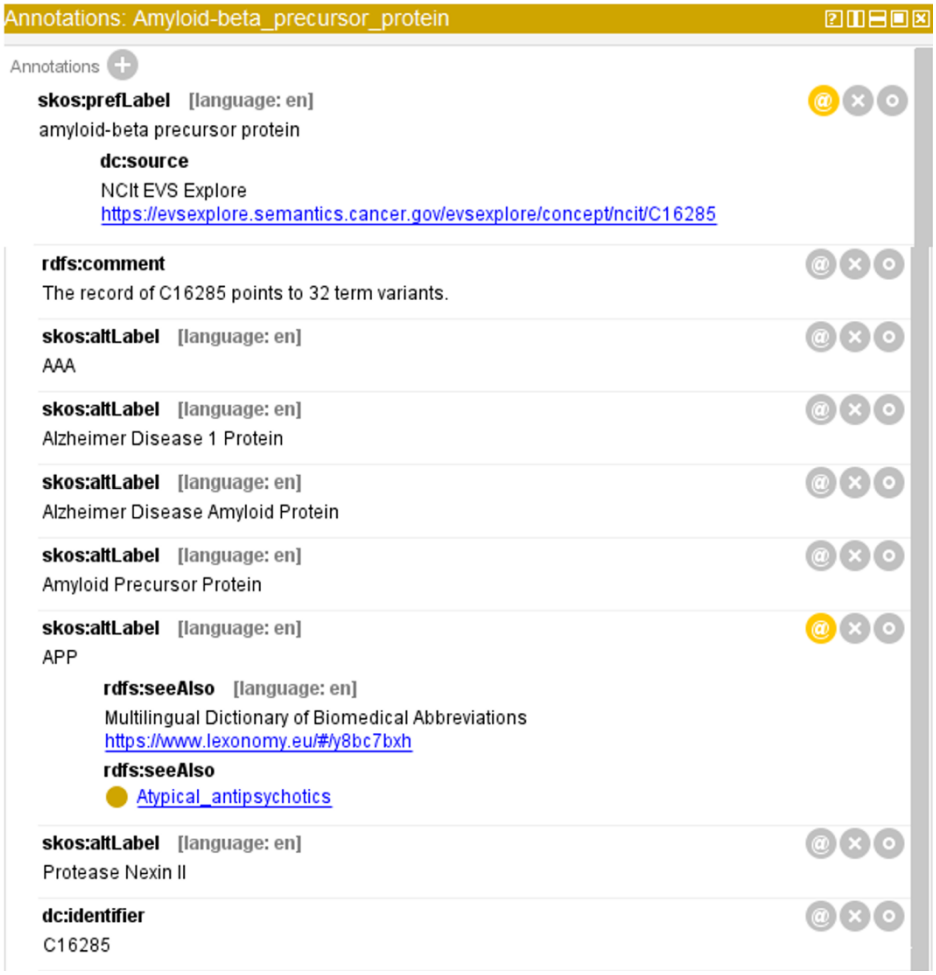

Fig. 15 Linguistic information and metadata associated with C16285 (NCIt EVS Explore)

Figure 15 represents the linguistic information annotated for one of the equivalent concepts discussed thus far, namely <Amyloid-beta precursor protein> (NCIt EVS Explore) with the `dc:identifier` C16285. As mentioned, we used the RDF vocabularies, namely DCMI Metadata Terms[48] and SKOS (W3C), in addition to the RDFS[49] labels incorporated by default in Protégé.

As illustrated in Figure 15, the first annotation (`skos:prefLabel`) indicates the preferred term, namely the linguistic form "amyloid-beta precursor protein", complemented with a sub-annotation indicating the source (`dc:source`). The term variants (`skos:altLabel`) are likewise annotated, e.g., "APP"; "Alzheimer disease 1 protein", among others. As indicated in the comment (`rdfs:comment`), "The record of C16285 points to 32 term variants" in the resource.

We draw attention to the sub-annotation `rdfs:seeAlso` associated with the `skos:altLabel` APP, which serves as an internal disambiguation mechanism by cross-referring to <Atypical antipsychotics>, a concept in OntoAbbMed.

The decision to refer to the concept <Atypical antipsychotics> (IOBC) is related to the use of the abbreviation APP, both in English and Portuguese, for two terms whose full forms differ and designate different concepts (cf. Section 5.1.2). As mentioned in Section 6.2.2, the only resource indexing the term "atypical antipsychotics" does not consider the abbreviation APP as a variant of the preferred term. This observation is annotated as `skos:hiddenLabel` APP, thereby signalling that the abbreviation is not

[48] https://www.dublincore.org/specifications/dublin-core/dcmi-terms/
[49] https://www.w3.org/TR/rdf-schema/

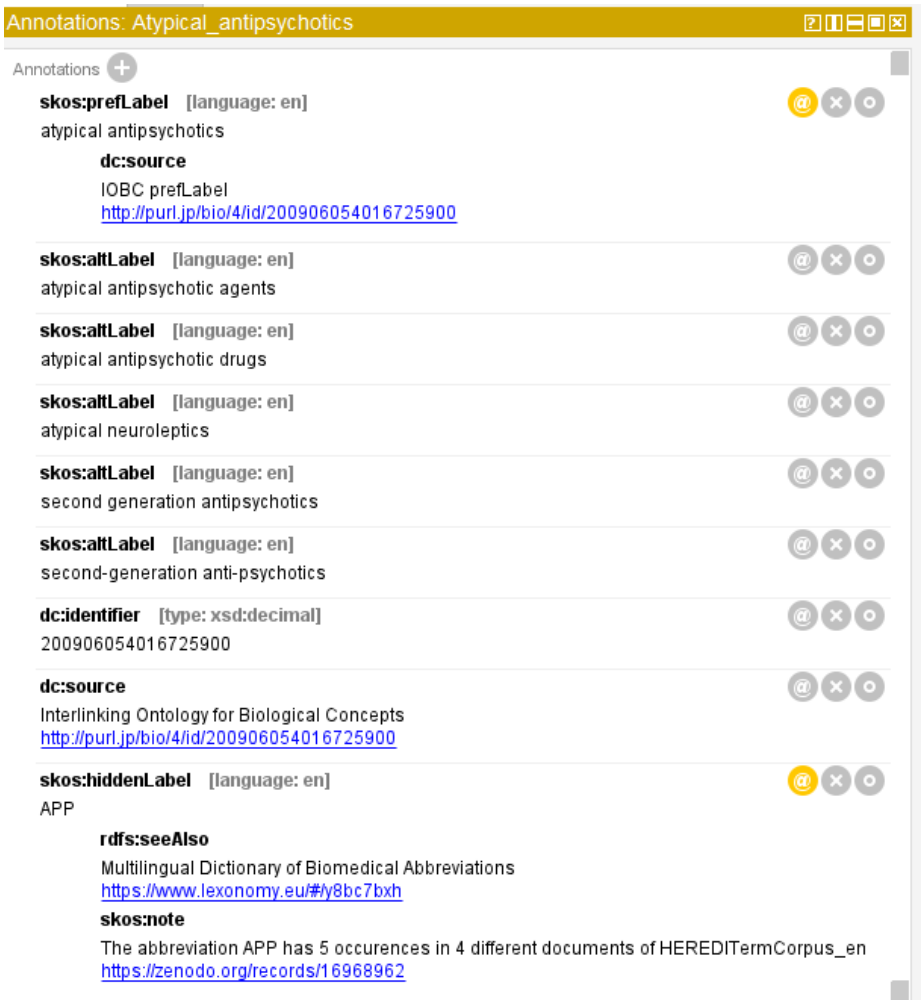

Fig. 16 Annotations associated with the concept <Atypical antipsychotics> (IOBC)

indexed to this concept, and sub-annotated (`skos:note`) with information concerning the use of the abbreviation in five corpus texts and its corresponding external source, as illustrated in Figure 16.

Finally, given the absence of a natural language definition associated with this concept (ID 200906054016725900), we refer to an external terminological resource – (`rdfs:seeAlso`) *Multilingual Dictionary of Abbreviations*, developed for this study – where the abbreviations extracted from the multilingual corpora, together with their conceptual and linguistic information, are systematised (cf. Section 8).

Regarding the concept <Amyloid beta-protein precursor> (MeSH/NCBI), `dc:identifier` D016564 (see. Figure 17), the preferred term (`skos:prefLabel`) is "amyloid beta-protein precursor" – where variation in hyphenation can be observed – accompanied by the source (`dc:source`) and its corresponding unique identifier: T049550. In the source resource, the abbreviation APP is not considered an admissible variant of the term, although a total of 11 variants is indexed to the preferred term. Among these variants, only seven were recorded in the ontology, including the abbreviation (`skos:altLabel`) ABPP. To each `skos:altLabel`, we added a sub-annotation to specify the corresponding source and URI (`dc:source`).

Finally, following the methodology described thus far, we annotated concept D016564 with information intended to address ambiguity issues arising from the use of the abbreviation APP in specialised texts. As illustrated in Figure 18, the abbreviation APP is annotated as `skos:hiddenLabel`

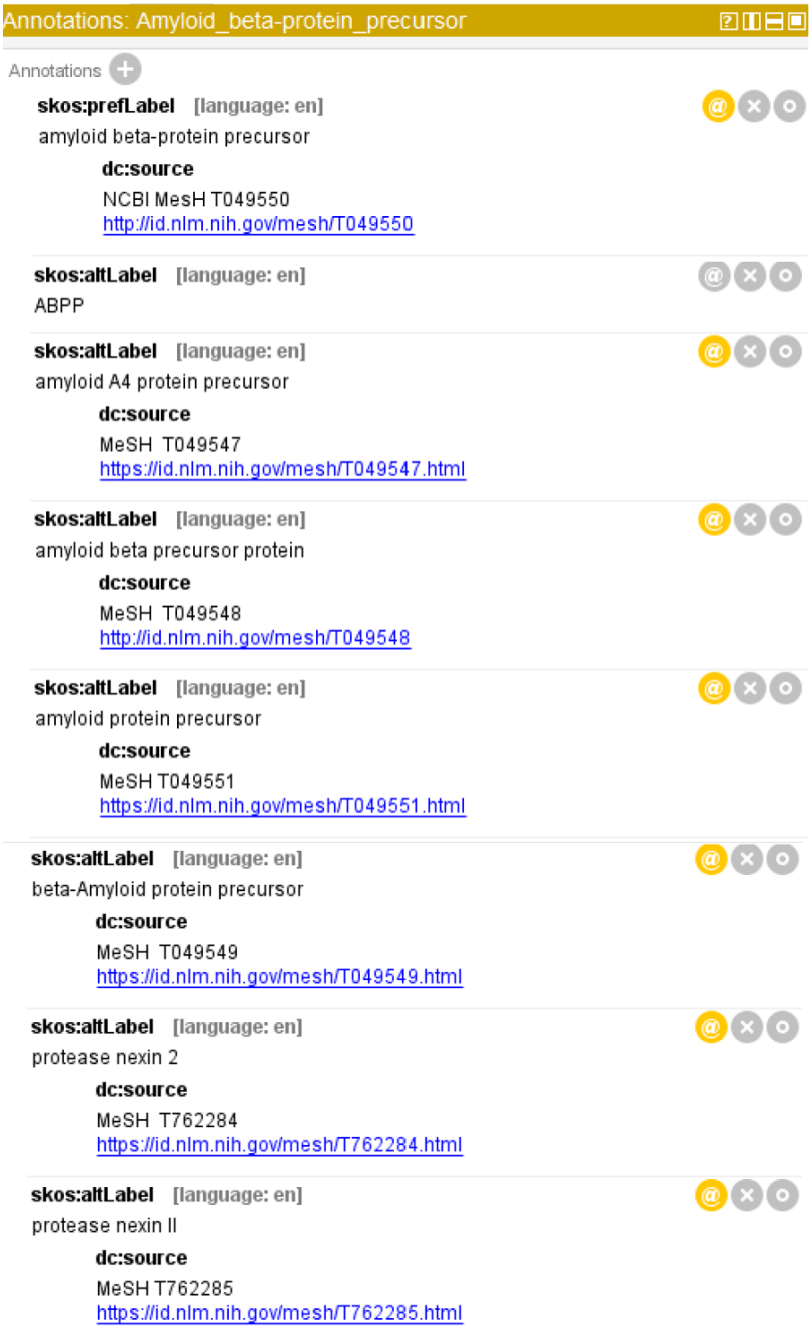

Fig. 17 skos:prefLabel and skos:altLabel associated with <Amyloid beta-protein precursor> (MeSH/NCBI)

in order to indicate that it is not indexed to concept D016564 as a term variant, although it is associated with this concept by domain experts.

In parallel, we added two sub-annotations:

(1) `rdfs:seeAlso` *Multilingual Dictionary of Biomedical Abbreviations*, suggesting an external resource supporting the disambiguation and/or confirmation of the use of the abbreviation; and

(2) a complementary note (`skos:note`):

> "1. The abbreviation APP, as attested in the HEREDITermCorpus_en, is not indexed in MeSH;
>
> 2. The prefLabel corresponds to the terms 'amyloid precursor protein' and 'β-amyloid precursor protein' attested in the HEREDITermCorpus_en.".

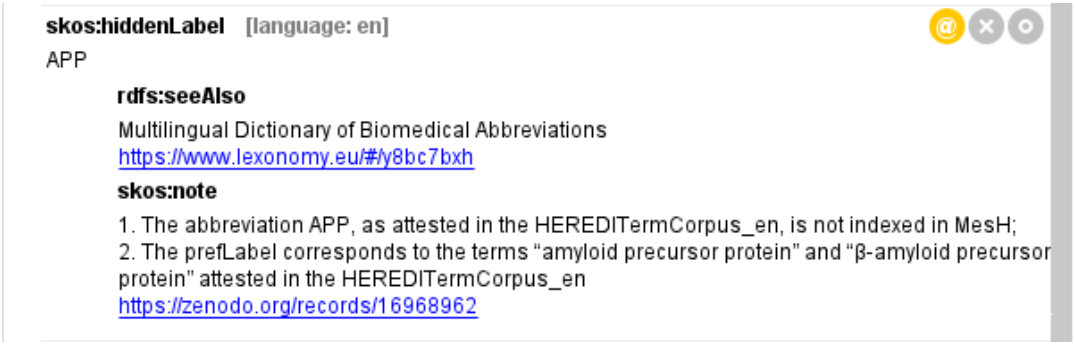

Fig. 18 Information supporting the disambiguation of the abbreviation APP

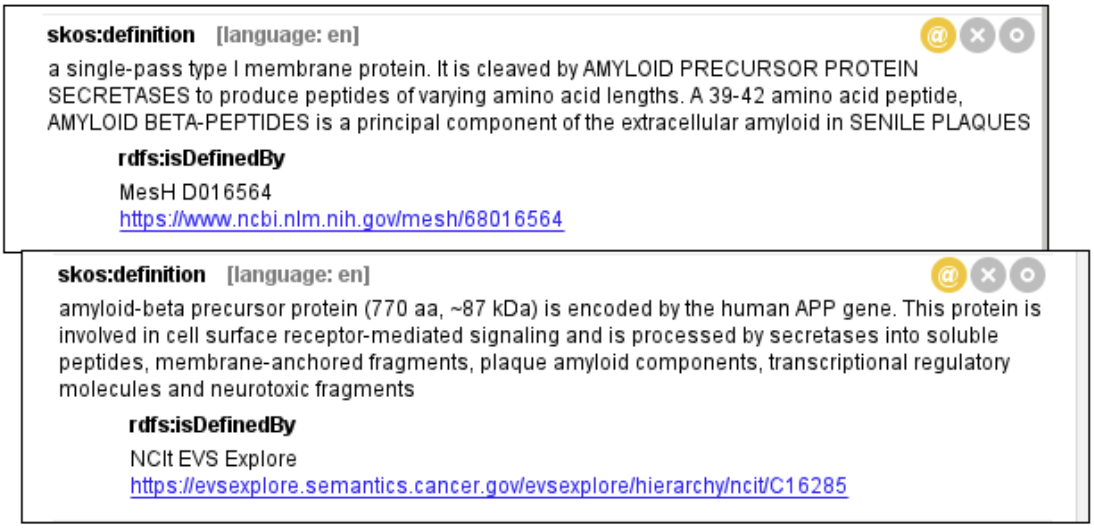

Fig. 19 Definitions of equivalent concepts D016564 (MeSH) and C16285 (NCIt)

7.3. Natural Language Definitions Aligned with Biomedical Resources

As with the annotation of preferred terms and their respective indexed variants, we used – whenever available – natural language definitions from the consulted biomedical resources to support the systematisation of the concepts to be integrated into the ontology. Concept definitions are annotated using the `skos:definition` label and complemented with a sub-annotation specified by the `rdfs:isDefinedBy` label, where we explicitly indicate the source resource and its corresponding URI.

Figure 19 illustrates two definitions extracted from the MeSH-NCBI and NCIt EVS Explore resources and annotated (`skos:definition`) to concepts D016564 (MeSH) and C16285 (NCIt). As mentioned, these are equivalent concepts whose conceptual and linguistic information complement one another. For example, the definition of concept C16285 refers to the concept <APP Gene>, reflecting a pragmatic[50] – associative – dependency between this concept and C16285. Although representing a distinct concept, <APP Gene> shares the abbreviation APP within the structure of the preferred term (skos:prefLabel). We opted to include it in the ontology due to the observed semantic proximity.

As a final note, we emphasise that, although the knowledge organisation presented here is grounded in specialised resources, it still requires validation by domain experts — a fundamental stage of terminological work. Appendix D (see Figure 21) illustrates the hierarchy of the 32 concepts currently systematised in OntoAbbMed, whose polyhierarchical relations result from HermiT inferences.

[50] According to ISO 1087:2000 (ISO, 2000), an *associative relation* or *pragmatic relation* is a "relation between two concepts (3.2.1) having a nonhierarchical thematic connection by virtue of experience".

The limited conceptual information available for <Atypical Antipsychotic> in the consulted resources did not allow us to determine whether it establishes relations with <Antipsychotic agent(s)>, although the term variants suggest semantic proximity. We leave this analysis for future work.

Competency questions, commonly used to assess the scope and intended application of an ontology, were not elaborated at this stage. They will be further specified and documented in subsequent work alongside the ontology's accompanying documentation. As the ontology is an ongoing effort, competency questions will be developed in parallel with the expansion of its conceptual coverage. Nevertheless, as a preliminary validation step, the inferences generated by the HermiT reasoner align with the conceptual representations presented in Section 6, providing evidence of the ontology's consistency, coherence, and adequacy in representing the concepts of interest.

8. Multilingual Dictionary of Biomedical Abbreviations

This section presents the systematisation, in the form of a multilingual domain dictionary, of: (i) conceptual information and (ii) linguistic information – i.e., the concept definition and the preferred term in English, respectively – observed in the biomedical resources (cf. Section 6), as well as (iii) linguistic information extracted from the corpus (cf. Section 4).

The *Multilingual Dictionary of Biomedical Abbreviations*[51] was developed using the Lexonomy editor.[52] The decision to use this editor is related to interoperability objectives between resources, namely the association of metadata with the sources of linguistic data in languages other than English – the corpora – and with OntoAbbMed. We consider these resources complement one another in light of the double dimension of Terminology. Whereas the corpora allow us to provide usage examples and statistical evidence of linguistic forms, OntoAbbMed aims to support abbreviation disambiguation by integrating complementary information derived from multiple authoritative resources.

Among the linguistic data, we highlight the full forms and corresponding abbreviations identified as term variants, as well as contexts of use for the linguistic forms under analysis. English abbreviations and their corresponding full forms are aligned with the preferred terms identified in MeSH, NCIt, SNOMED CT, and IOBC. Whenever denominative variation observed in the corpus is absent from the biomedical resources, it is presented in the dictionary entries as a disambiguation mechanism for users of both the linguistic and ontological resources.

In summary, although the dictionary organises language both inter- and intra-linguistically, it relies on conceptual organisation to reduce the ambiguity observed in specialised texts.

8.1. Microstructure of the Terminological Record

Appendix E (see Figure 22) presents the microstructure of the terminological entry for ABPP in English and its equivalents in Portuguese and Italian. The English abbreviation and its corresponding full form are aligned with *Source MeSH T049550*, as specified in the metadata. Immediately below, an external link labelled *OntoAbbMed* points to the ontology OWL file (cf. Section 7) – a resource integrating information from several biomedical resources – available on GitHub. The Portuguese and Italian abbreviations and full forms are linked to their respective corpora, likewise referenced through the metadata, pointing to the corresponding repository.

Aligned with the ontology's `skos:seeAlso` annotations, we associate the label "See also" with the linguistic form, whenever necessary, as a disambiguation mechanism. In the example illustrated in Appendix E (see Figure 22), the Portuguese abbreviation APP is used for two distinct full forms (cf. Section 5).

We emphasise that, although the present resource is aligned with the analysed biomedical resources, its terms are written in lower case, following the conventions of *ISO/IEC Directives, Part 2*,[53] the primary reference for drafting ISO standards.[54]

[51] https://www.lexonomy.eu/#/y8bc7bxh
[52] https://www.lexonomy.eu/
[53] https://www.iso.org/sites/directives/current/part2/index.xhtml#_idTextAnchor020
[54] https://www.iso.org/home.html

Similarly, the definition is aligned with *Source MeSH D016564*, as indicated in the metadata. Although terminological conventions prescribe otherwise, it begins with a capital letter and retains the final full stop, to remain faithful to the cited source.

This is followed by a context illustrating the use of the full form. Metadata indicate the source, here *PD-Rv2023-MDPI(6)*. As mentioned in Section 3, the descriptor of this text shows it is a scientific review paper on Parkinson disease, published in 2023 and collected from the MDPI repository. The digit 6 indicates five additional texts with this classification in the corpus. Following the adopted methodology, the source corpus is likewise referenced through a small arrow and a link icon.

A final terminological field provides additional notes to complement linguistic or conceptual information, although not exclusively. For the ABPP@en entry, the following information is added:

1. *The abbreviation APP, although attested in the HEREDITermCorpus_en, is not indexed in NCBI MeSH D016564*;
2. *The full form of the term in English corresponds to the terms "amyloid precursor protein" and "β-amyloid precursor protein" attested in the HEREDITermCorpus_en".*

We consider this information useful for resolving ambiguity in specialised texts, although such forms are not regarded as variants of a preferred term in the resources analysed in this study. This information is also annotated in OntoAbbMed.

Finally, in accordance with good practices in terminological database management, each entry concludes with administrative metadata: author | date | update date.

9. Conclusion

This study examined abbreviation ambiguity in multilingual biomedical texts from a terminological and knowledge-organisation perspective. Through the analysis of multilingual comparable corpora in English, Portuguese, and Italian, we identified multiple cases of intra- and interlingual denominative variation involving abbreviated and full forms, including mono- and polyreferential ambiguity. The systematisation of the results enabled the establishment of a typology of ambiguity within the domain, showing that abbreviation usage in specialised biomedical texts is highly variable and often diverges from the representations provided in authoritative biomedical resources.

Comparative analysis of MeSH, NCIt, SNOMED CT, and IOBC revealed important differences in conceptual granularity, preferred labels, indexed term variants, and abbreviation representation. In several cases, corpus-attested abbreviations were absent from the consulted resources; in others the same abbreviation was associated with different concepts depending on the resource. These observations highlight the limitations of considering abbreviations exclusively from a linguistic or corpus-based perspective, reinforcing the need for approaches that integrate conceptual organisation and multilingual terminological variation.

To address these challenges, this work proposed OntoAbbMed, an ontology designed to systematise conceptual and linguistic information associated with ambiguity arising from biomedical abbreviation use. By integrating corpus-derived information with data aligned with biomedical resources, the ontology formally represents conceptual relations, equivalence relations, polyhierarchies, and disambiguation mechanisms. The *Multilingual Dictionary of Biomedical Abbreviations* complements the ontology by organising linguistic evidence, multilingual equivalents, corpus attestations, and contextual usage examples.

Methodologically, this study demonstrates the relevance of combining corpus linguistics, terminology, biomedical ontologies, and knowledge engineering to support specialised knowledge organisation. The articulation between linguistic and conceptual dimensions proved particularly useful for identifying discrepancies between expert discourse and biomedical resources, as well as representing ambiguity in a structured and interoperable manner.

Nevertheless, this work constitutes an initial stage of a broader research agenda. The ontology currently requires validation by biomedical domain experts, particularly regarding conceptual relations that could not be fully inferred from the consulted resources. Future work will extend the multilingual

coverage of the corpora and ontology, expand the number of represented concepts and abbreviations, refine disambiguation mechanisms, and explore interoperability with additional biomedical ontologies and NLP applications.

Overall, this study contributes to ongoing discussions concerning biomedical terminology management, semantic interoperability, and multilingual knowledge representation, while providing practical resources to support the interpretation and disambiguation of abbreviations in specialised biomedical communication.

Declaration on Generative AI

During the preparation of this work, the authors used Perplexity to check grammar and spelling and to assist with paraphrasing and rewording the text. After employing this tool, the authors reviewed and edited the content as necessary and accepted full responsibility for the publication's content. The final English version of the chapter was proofread by Maria João Ferro, whom the authors thank for her valuable feedback.

Acknowledgements

This research was partly supported by Portuguese national funding through the FCT – Portuguese Foundation for Science and Technology, I.P., as part of the project UID/03213/2025 (https://doi.org/10.54499/UID/03213/2025), and by the European Union NextGenerationEU, as part of the project UID/PRR/03213/2025 (https://doi.org/10.54499/UID/PRR/03213/2025) and the HEREDITARY Project, funded by the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101137074.

References

Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press. https://doi.org/10.1515/9780748626908

Berger, S., Grzonka, P., Hunziker, S., Frei, A. I., & Sutter, R. (2026). When shortcuts fall short: The hidden danger of abbreviations in critical care. *Journal of Critical Care*, *91*, Article 155236. https://doi.org/10.1016/j.jcrc.2025.155236

Bowker, L., & Hawkins, S. (2006). Variation in the organization of medical terms: Exploring some motivations for term choice. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *12*(1), 79-110. https://doi.org/10.1075/term.12.1.05bow

Brunetti, L., Santell, J. P., & Hicks, R. W. (2007). The impact of abbreviations on patient safety. *The Joint Commission Journal on Quality and Patient Safety*, *33*(9), 576-583. https://doi.org/10.1016/s1553-7250(07)33062-6

Cabré Castellví, M. T. (1999). *Terminology: Theory, methods and applications*. John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.1

Costa, R. (2013). Terminology and specialised lexicography: Two complementary domains. *Lexicographica*, *29*, 29-42. https://doi.org/10.1515/lexi-2013-0004

Daille, B. (2017). *Term variation in specialised corpora: Characterisation, automatic discovery and applications*. John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.19

Grossman, Liu, L., Grossman, R. H., Mitchell, E. G., Weng, C., Natarajan, K., Hripcsak, G., & Vawdrey, D. K. (2021). A deep database of medical abbreviations and acronyms for natural language processing. *Scientific Data*, *8*(1), Article 149. https://doi.org/10.1038/s41597-021-00929-4

International Organization for Standardization. (2000). *Terminology work — Vocabulary. Part 1: Theory and application* (ISO Standard No. 1087-1:2000). https://www.iso.org/standard/20057.html

International Organization for Standardization. (2019). *Terminology work and terminology science — Vocabulary* (ISO Standard No. 1087:2019). https://www.iso.org/standard/62330.html

International Organization for Standardization. (2022). *Terminology work — Principles and methods* (ISO Standard No. 704:2022). https://www.iso.org/standard/79077.html

International Organization for Standardization. (2025). *Management of terminology resources — Terminology extraction* (ISO Standard No. 5078:2025). https://www.iso.org/standard/81917.html

Jaber, A. & Martínez, P. (2021). Disambiguating clinical abbreviations using pre-trained word embeddings. In C. Pesquita, A. Fred, & H. Gamboa (Eds.), *Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2021)* - HEALTHINF (pp. 501-508). SciTePress. https://dx.doi.org/10.5220/0010256105010508

Jayatilake, D. C., & Oyibo, S. O. (2023). Interpretation and misinterpretation of medical abbreviations found in patient medical records: A cross-sectional survey. *Cureus*, *15*(9), Article e44735. https://doi.org/10.7759/cureus.44735

Lacy, L. W. (2005). *OWL: Representing information using the Web Ontology Language*. Trafford Publishing (UK) Ltd. ISBN 1-4120-3448-5

Liu, H., Lussier, Y., & Friedman, C. (2001). A study of abbreviations in UMLS. *Proceedings of the AMIA Symposium 2001* (pp. 393-397). PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC2243414/

Liu, H., Aronson, A. R., & Friedman, C. (2002). A study of abbreviations in MEDLINE abstracts. In I. S. Kohane (Ed.), *Proceedings of the AMIA Symposium 2002* (pp. 464-468). PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC2244212/

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In D. Bourigault, C. Jacquemin, & M.-C. L'Homme (Eds.), *Recent advances in computational terminology* (pp. 279-302). John Benjamins Publishing Company. https://doi.org/10.1075/nlp.2.15mey

Pakhomov, S., Pedersen, T., & Chute, C. G. (2002). Abbreviation and acronym disambiguation in clinical discourse. *AMIA Annual Symposium Proceedings 2002* (pp. 589-593). PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC1560669/

Ramos, M., & Costa, R. (2024). Applying text mining methods to construct a domain ontology from definitions. *Umanistica Digitale*, *8*(18), 103–143. https://doi.org/10.6092/issn.2532-8816/19284

Santos, C., & Costa, R. (2015). Domain specificity: Semasiological and onomasiological knowledge representation. In H. J. Kockaert & F. Steurs (Eds.), *Handbook of terminology: Volume 1* (pp. 153–

179). John Benjamins. https://doi.org/10.1075/hot.1.dom1

Schwartz, A. S., & Hearst, M. A. (2003). A simple algorithm for identifying abbreviation definitions in biomedical text. In R. B. Altman, A. K. Dunker, L. Hunter, T. A. Jung, & T. E. Klein (Eds.), *Biocomputing 2003: Proceedings of the Pacific Symposium* (pp. 451–462). World Scientific. https://doi.org/10.1142/9789812776303_0042

Sinclair, J. (1996). *Preliminary recommendations on corpus typology* (EAGLES Document eag-tcwg-ctyp/p). Expert Advisory Group on Language Engineering Standards. https://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html

The Joint Commission. (2003, December). National patient safety goal on abbreviations clarified, implementation revised. *Joint Commission Perspectives*, *23*(12), 14-15.

Wermuth, M.-C., & Verplaetse, H. (2019). Medical terminology in the Western world: Current situation. In A. Alsulaiman & A. Allaithy (Eds.), *Handbook of Terminology, Volume 2: Terminology in the Arab world* (pp. 83-108). John Benjamins Publishing Company. https://doi.org/10.1075/hot.2.05wer

Xu, H., Stetson, P. D., & Friedman, C. (2007). A study of abbreviations in clinical notes. *AMIA Annual Symposium Proceedings 2007* (pp. 821-825). PubMed Central. https://pmc.ncbi.nlm.nih.gov/articles/PMC2655910/

Appendix A The First 26 Lines of the Concordance returned by CQL (1).

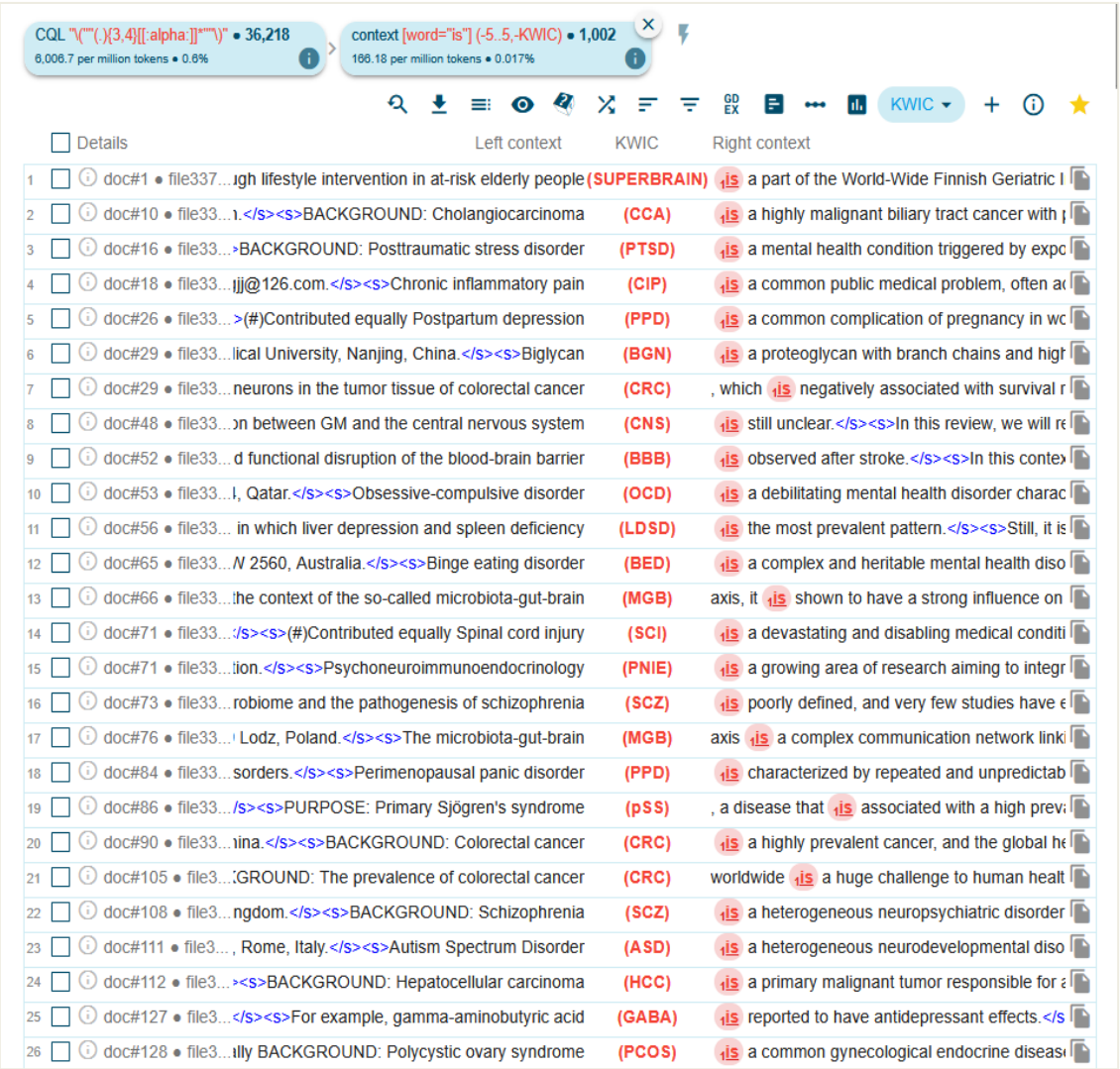

Fig. 20 The first 26 lines of the concordance returned by CQL (1)

Appendix B Typology of Intra- and Interlingual Denominative Variation of Full Forms and Corresponding Abbreviations

Table 5 Typology of intra- and interlingual denominative variation of full forms and corresponding abbreviations

pt	Abbreviation	Full form	C=	C=	C=	C≠	C≠	C≠
	DCL	demência com corpos de Lewy (1)	DLB @pt	LBD @en	BDL @it			
		demência dos corpos de Lewy (2)	DLB @pt	LBD @en	BDL @it			
	DLB	demência de corpos de Lewy	DCL @pt	LBD @en	BDL @it			
	APP (1)	afasia primária progressiva	-	PPA @en	APPL @it	APP[2] @pt	APP[1] @en	
	APP (2)	proteína precursora amilóide	-	APP[2] @en	APP @it	APP[1] @pt		
en	DLB	dementia with Lewy bodies	LBD @en	DCL @pt	BDL @it			
	LBD	Lewy body dementia	DLB @en	DCL @pt	BDL @it			
	APP (1)	atypical antipsychotics	-	-	-	APP[2] @en	APP[1] @pt	APP @it
	APP (2)	β-amyloid precursor protein (1)	-	APP[2] @pt	APP @it	APP[1] @en		
		amyloid precursor protein (2)	-	APP[2] @pt	APP @it	APP[1] @en		
	PPA	primary progressive aphasia	-	APP[1] @pt	APPL @it			
it	DLB	demenza a corpi di Lewy (1)	LBD @it	BDL @it	DCL @pt			
		demenza con corpi di Lewy (2)	LBD @it	BDL @it	DCL @pt			
	LBD	demenza a corpi de Lewy	BDL @it	DLB @it	DCL @pt			
	BDL	demenza a corpi de Lewy	LBD @it	DLB @it	DCL @pt			
	APP	proteina percursore dell'amiloide (1)	-	APP[2] @en	APP[2] @pt	APP[1] @en	APP[1] @pt	
		percursore della proteina beta-amiloide (2)	-	APP[2] @en	APP[2] @pt	APP[1] @en	APP[1] @pt	
		precursore di Aβ (3)	-	APP[2] @en	APP[2] @pt	APP[1] @en	APP[1] @pt	
	APPL	sindrome dell'afaisa primaria progressiva	-	PPA @en	APP[1] @pt			

In Table 5, three of the English abbreviations, namely DLB, LBD and APP, are highlighted in bold to illustrate their intralingual usage in both Portuguese and Italian. The columns “C=” and “C≠” are used to identify the common usage of a given abbreviation, either in the same language (intralingual) or across languages (interlingual), as well as the concepts they designate. The mono-referentiality is marked in column “C=” (same concept), whereas the polyreferentiality is marked in column “C≠” (different concept).

Appendix C OntoAbbMed Metrics

Table 6 OntoAbbMed metrics

Metrics	**#**
Axiom	**607**
Logical axiom count	**59**
Declaration axioms count	**155**
Class count	**32**
Object property count	18
Data property count	14
Individual count	0
Annotation Property count	**83**
Class axioms	
SubClassOf	**15**
EquivalentClasses	**7**
Hidden GCI Count	9
Object Property axioms	
SubObjectPropertyOf	15
InverseObjectProperties	4
FunctionalObjectProperty	1
TransitiveObjectProperty	3
SymmetricObjectProperty	4
Data property axioms	
SubDataPropertyOf	10
Annotations axioms	
AnnotationAssertion	**305**
AnnotationPropertyDomain	3
AnnotationPropertyRangeOf	13

Appendix D Inferred Hierarchy of the 32 Concepts Currently Systematised in OntoAbbMed, in OWLViz

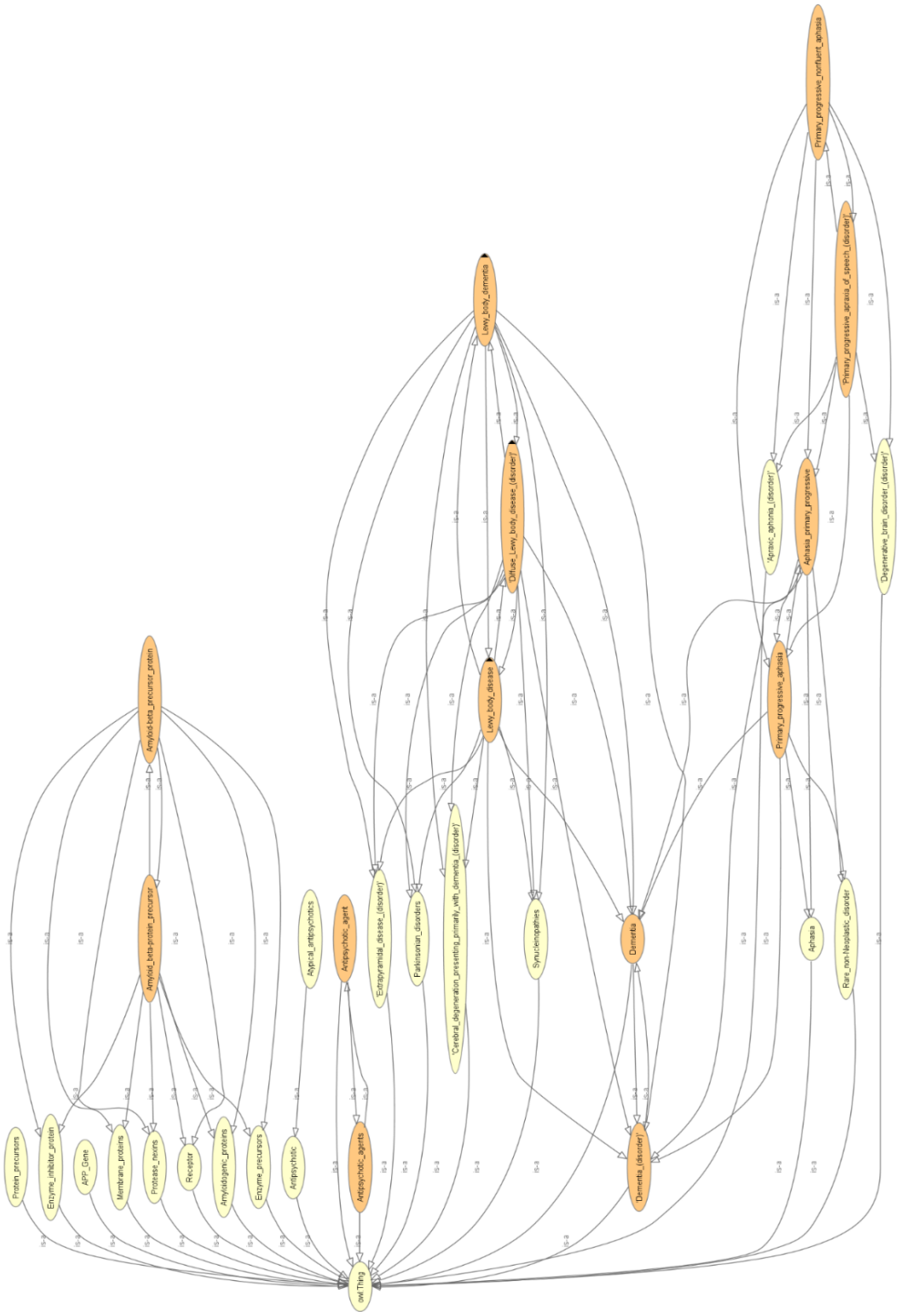

Fig. 21 Inferred hierarchy of the 32 concepts currently systematised in OntoAbbMed, in OWLViz

Appendix E Microstructure of ABPP@en and the Equivalents APP@pt and APP@it

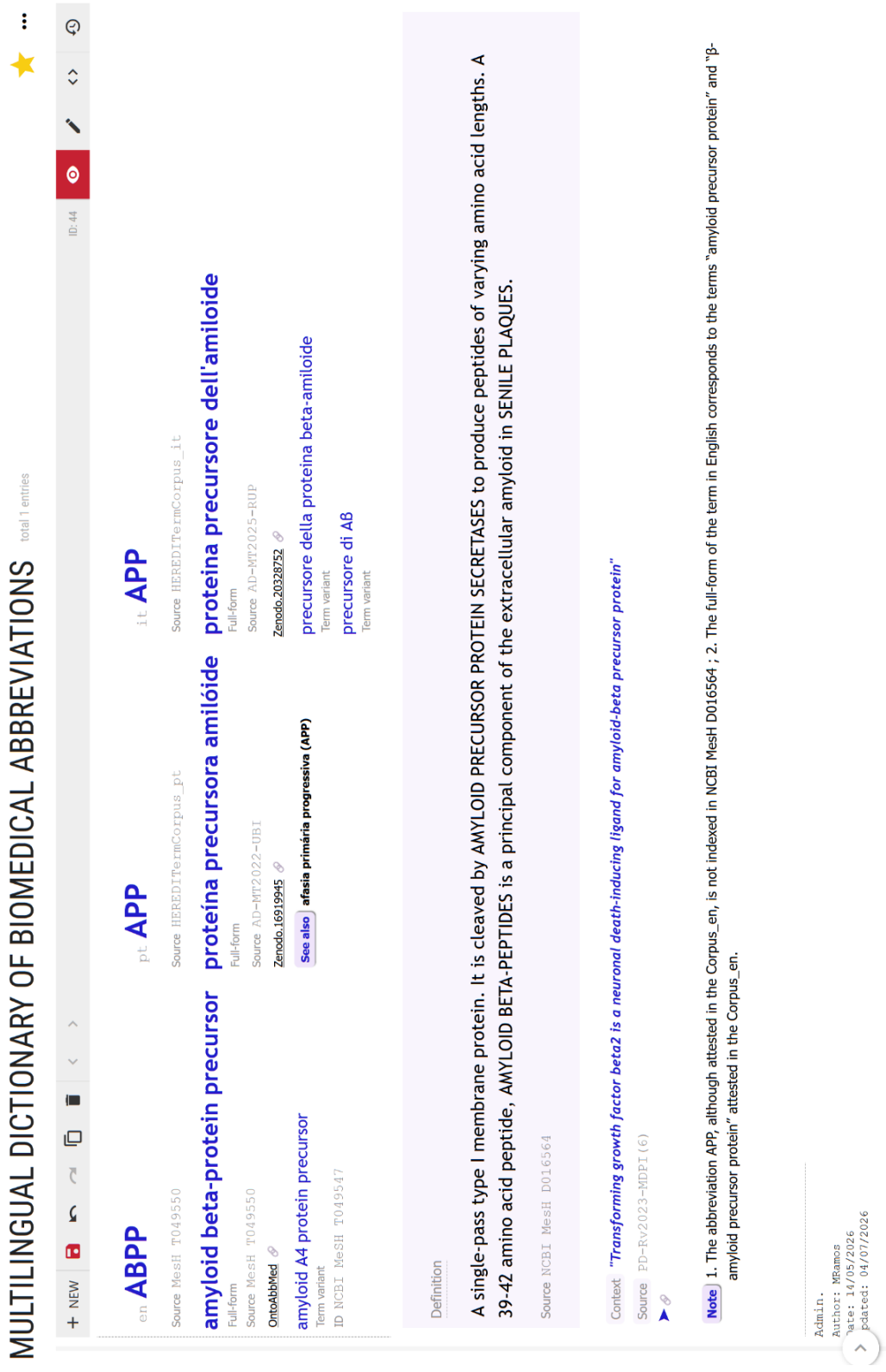

Fig. 22 Microstructure of ABPP@en and the equivalents APP@pt and APP@it