Journal of Digital Terminology and Lexicography - ISSN:3103-3601
Vol. 2, No. 1 (2026) DOI: 10.25430/pupj.jdtl.1781816298

Defining and Relating Concepts in Medical Terminology: A Case Study on the Gut–Brain Interplay

Vanessa Bonato[1*], Laura Menotti[2], Federica Vezzani[1], Giorgio Maria Di Nunzio[2], Gianmaria Silvello[2]

[1*]Department of Linguistic and Literary Studies, University of Padua, Via E. Vendramini 13, Padova, 35137, Italy.
[2]Department of Information Engineering, University of Padua, Via Gradenigo 6a, Padova, 35131, Italy.

*Corresponding author(s). E-mail(s): vanessa.bonato@phd.unipd.it;
Contributing authors: laura.menotti@unipd.it; federica.vezzani@unipd.it; giorgiomaria.dinunzio@unipd.it; gianmaria.silvello@unipd.it;

Abstract

This contribution describes the terminology work carried out to analyze the conceptual dimension of terminology concerning the gut–brain interplay and associated neurodegenerative and mental health-related conditions. We present: 1) the methodology used to formulate intensional definitions of biomedical concepts, and 2) the approach used to identify associative relations between concepts in biomedical literature, along with a proposed typology of these relations. The biomedical concepts and the associative relations established between them are then compared with their respective representation in the Hereditary Ontology (HERO), which is the only formal resource specifically designed to model knowledge on the gut–brain interplay and associated medical conditions. The comparison between terminological data and data contained in the ontology sheds light on the distinct nature of terminological and ontological representations of specialized knowledge, highlighting both differences and points of intersection. This work constitutes a preliminary step toward the future development of a novel terminology resource, aimed at representing specialized knowledge on the gut–brain interplay and related conditions.

Keywords: Medical Terminology, Gut-Brain Interplay, Intensional Definitions, Concept Relations, HERO Ontology, Terminology-Ontology Alignment

1 Introduction

Analyzing specialized medical knowledge generally requires accounting for its inherent dynamism, which entails systematically tracing the evolution of concepts and the variation of terms used to designate them (Vezzani & Costa, 2024). Terminology work grounded in continuous updating therefore constitutes an essential prerequisite for the development of domain-specific terminology resources, particularly medical ones. In fact, these resources face the demanding task of keeping conceptualizations constantly aligned with their designations, by maintaining consistent and traceable records of concepts, concept relations, definitions, and term variants (Cabezas-García & León-Araúz, 2020; Vezzani, Di Nunzio, & Silecchia, 2022).

Terminology work informed by ongoing scientific advances is particularly important for representing specialized knowledge related to constantly evolving research areas in the medical domain, such as the gut–brain axis. Recent biomedical literature on the gut–brain axis – examining links between the gut microbiota and conditions affecting mental health, such as depression and anxiety (Fan, Li, & Chen, 2024; Lai & Xiong, 2025; Xiong et al., 2023) – has grown markedly, with PubMed publications more than doubled from 2020 to 2025 (Martinelli et al., 2026). Against this background, recently published biomedical documents and patient records represent primary sources for acquiring up-to-date medical knowledge, through the analysis of the terms currently used by domain experts to designate concepts.

Understanding and analyzing terminology within this interdisciplinary domain presents several challenges. Novel concepts are continuously emerging, thus requiring the timely drafting of concept definitions that are subject to frequent revision. Moreover, terminology exhibits considerable variability, contingent upon the research focus and clinical context. Therefore, in order to properly capture the domain-specific expertise in terminology resources, it is first necessary to carry out comprehensive terminology work that investigates both the linguistic and conceptual aspects of the terminology involved.

This article describes the work carried out to analyze the conceptual dimension of terminology related to the gut–brain axis. The present study is necessary to provide the basis for the future development of a terminology resource specific to this research domain. In particular, we: 1) describe the methodology for formulating intensional definitions of biomedical concepts, and 2) identify and analyze the typology of associative relations linking these concepts. Moreover, we assess differences in the level of granularity of terminology-derived concepts and concept relations compared to their representation in a formal domain-specific ontology.

The main contributions of this work are the defined biomedical concepts and the identified associative relations, specifically analyzed with respect to those represented in the Hereditary Ontology (HERO) (Silvello & Menotti, 2025).[1] HERO is the only existing formal resource modeling the gut–brain interplay, which supports the representation of phenoclinical and genomic data related to the gut—brain axis and its associated neurodegenerative and mental health-related conditions. Our analysis reveals that the terminology-derived concepts can either specialize the generic classes in HERO or serve as instances of broader HERO concepts. Similarly, the associative relations extracted through terminology work exhibit finer granularity compared to those encoded in HERO, reflecting the distinct scope underpinning each resource. We share the concepts and their intensional definitions, the concept relations with their typology, and the comparison with HERO at https://anonymous.4open.science/r/GutBrainAxisConcepts-4221/.

This work is structured as follows: Section 2 covers the theoretical background on intensional definitions and concept relations. Section 3 presents the terminology work conducted to analyze the conceptual dimension of terminology related to the gut–brain axis. In particular, we: 1) detail the methodology for formulating intensional definitions of biomedical concepts and present the definitions, and 2) analyze concept relations in biomedical texts. Section 4 provides a typology of associative relations. Section 5 compares identified concept relations with those in HERO. Section 6 presents conclusions and future work.

2 Theoretical Background

This section outlines the theoretical background on intensional definitions and concept relations. All definitions are drawn from the standards within ISO TC 37 about "Language and Terminology".[2]

2.1 Intensional Definitions

Investigating the conceptual dimension of terminology entails several different activities, among which identifying concepts, formulating their definitions, and organizing them into concept systems, in which concepts are represented and interrelated through different types of concept relations (Carvalho, 2018; Costa, 2013; Roche & Papadopoulou, 2024; Vezzani, 2022).

Different approaches to drafting definitions have been proposed in the literature. One of the key issues concerning terminological definitions is the identification and selection of the elements that are necessary to define a concept. In the work by Seppälä (2009), for example, a framework that brings into focus key aspects to be considered "when conceiving an evaluation scheme for definition extraction (DE) in terminology" (p. 47) is proposed. In particular, the raised and addressed question concerns which information is actually relevant and should be selected for inclusion in a terminographic definition. A related issue concerns the identification of the “relevance conditions” (p. 49), in terms of principles that should: 1) allow to

[1] https://hereditary.dei.unipd.it/ontology/. See also Faggioli et al. (2024) for more details about the phenoclinical part of HERO.
[2] https://www.iso.org/committee/48104.html

differentiate between defining elements and non-defining elements when dealing with "salient features" (p. 49), and 2) choose relevant features among many others that may be relevant for definitional purposes.

According to Temmerman (2000), the choice concerning how to define a unit of understanding depends on the unit being described. Specifically, if the unit of understanding "is part of a logically or ontologically structured frame or propositional ICM [Idealised Cognitive Model]", it can be defined by relying on a definition that contains the superordinate term as well as "the necessary and sufficient characteristics" that distinguish the defined unit from coordinate ones (p. 122). However, the adoption of the intensional definition is not always adequate. Indeed, when a unit of understanding exhibits a prototypical structure, it is necessary to rather define it by adopting a template-based approach.

In line with Temmerman (2000) and Seppälä (2015), San Martín (2022) argues that "necessary and sufficient characteristics should be replaced by relevant characteristics in the selection of the features in the definition" (p. 55). The author presents a corpus-based approach to the drafting of terminological definitions, which takes into account thematic variation. Specifically, the approach involves the drafting of multiple definitions for a concept, each representing knowledge about that concept in a determined specialized domain, to meet the specific information needs of users.

From a normative perspective, according to ISO 1087 (International Organization for Standardization, 2019), a concept is a "unit of knowledge created by a unique combination of characteristics". To define a concept unambiguously, it is necessary to pinpoint the essential characteristics, as well as the delimiting characteristics that set it apart from other concepts, which is also crucial for accurately conveying expert knowledge within a given domain. Although other types of definitions, such as extensional definitions or partitive definitions, could also be suitable for definitional purposes, in terminology work the intensional definition is regarded as the most informative type of definition and is therefore preferred (International Organization for Standardization, 2019, 2022; Löckinger, Kockaert, & Budin, 2015).

In particular, the intensional definition expresses the intension of a concept by stating two elements: the immediate generic concept, which serves as the broader category or superclass, and the delimiting characteristics, which are the essential characteristics that distinguish the concept from its coordinates (International Organization for Standardization, 2019). The generic concept indicates where the defined specific concept sits within a hierarchical structure of concepts, clarifying its relationship to more general categories. Meanwhile, the delimiting characteristics ensure that the concept is precisely distinguished from other concepts under the same generic concept, by highlighting its unique characteristics (International Organization for Standardization, 2022). Together, these two elements provide a comprehensive framework for representing concepts in a terminology resource, in a way that supports both clear communication and systematic organization of domain-specific knowledge.

2.2 Concept Relations

Alongside the identification and definition of concepts, terminology work also involves the individuation of relations among concepts. Concept relations, indeed, play a crucial role in the representation of specialized knowledge, as they reveal how concepts are interconnected within a domain. For example, concept relations are represented in different specialized resources that focus on the medical domain, including the UMLS Semantic Network [3] and SNOMED CT (Chang & Mostafa, 2021).[4]

Understanding how concepts are related to each other in resources and concept systems requires to consider the different types of concept relations that can be established between them. As stated by Nuopponen (2022), however, "[t]here are not only numerous types of concept relation, but also various ways to define and classify them" (p. 67). Over time, different typologies of concept relations have been proposed, for instance by Nuopponen (1994, 2005), Maroto García and Alcina (2009) and León-Araúz and Faber (2010).

In the work by Maroto García and Alcina (2009), for instance, a typology of concept relations is proposed, "in which each relationship is defined formally in terms of its properties and the nature of the conceptual classes involved" (p. 232). Each concept involved in a relation is attributed to a class, that can be entities, properties, or activities. The relations established between concepts can pertain to one of the following macro-categories of concept relations: 1) logical, 2) meronymic, 3) sequential, 4) argumental and circumstantial, and 5) other types of relations that cannot be mapped to these categories. The concept relations are formally defined in terms of properties, namely transitivity, symmetry, cardinality, and inverse

[3] https://www.nlm.nih.gov/research/umls/knowledge_sources/semantic_network/index.html
[4] https://www.nlm.nih.gov/healthit/snomedct/index.html

relationships. This description model was then used to populate a database containing concepts related to finished ceramic products, implemented using the ontology editor Protégé. [5]

Relations are also used to establish systematic links between specialized concepts in the EcoLexicon knowledge base, which collects knowledge about the environmental domain (León-Araúz & Faber, 2010). These relations, referred to as semantic relations, are also established between different kinds of concept types: entities, events, and properties. As stated by the authors, "concept nature alone determines the potential activation of certain semantic relations, but at the same time, semantic relations determine which kind of concepts can be part of the same conceptual proposition" (p. 14). However, taking into account the multidimensionality of concepts pertaining to the analyzed field, the field of environment is divided into different contextual domains, as "domain membership reconceptualizes versatile concepts, restricting their relational behaviour" (p. 16).

Concerning associative relations, as pointed out by Sambre and Wermuth (2015), they "have been analysed less by (prescriptive) terminology than hierarchical ones" (p. 101), and a complete and exhaustive list of this type of concept relations has not yet been finalized. In the mentioned study, which is grounded in frame semantics and construction grammar, associative relations in the medical domain are identified through the analysis of titles of medical papers published in English-language journals. Specifically, the paper focuses on the interconnection of associative relations of instrument, time and cause.

In this work, we describe the typologies of concept relations respectively provided by ISO 1087 (International Organization for Standardization, 2019) and Nuopponen (2022), focusing then on Nuopponen's typology as the theoretical reference of the study.

According to ISO 1087 (International Organization for Standardization, 2019), concept relations fall into two distinct types: hierarchical relations and associative relations. Hierarchical relations can be either generic or partitive. For instance, the generic relation (or "`is_a`" relation) links <Human> to <Organism>, to indicate that humans are organisms.[6] The partitive relation (or "`is_part_of`" relation) can be used to link <Bacteria> to <Microbiome>, indicating that bacteria are part of the microbiome. All concept relations other than the hierarchical ones are classified as associative relations. Among these, the sequential relation specifically constitutes an "associative relation by which concepts can be ordered by a relevant ordering criterion". Concepts, indeed, can be interrelated through three types of sequential relations: spatial, temporal, and causal.

Nuopponen (2022), instead, proposes a typology of concept relations that comprises seven distinct macro-categories, each encompassing several relation types:

1. Generic relations: link superordinate concepts to subordinates, coordinate concepts to each other, as well as "other pairs of concepts on different abstraction levels in the same concept system" (p. 72).
2. Contiguity relations: cover part-whole (partitive), material–component, property, locative, enhancement, ownership, rank, and temporal relations.
3. Activity relations: connect an activity concept with, for instance, its agent, object, tools, materials, methods, manner or activity time.
4. Origination relations: link objects, that can be either concrete or abstract, to their origin.
5. Developmental relations: describe the phases or stages through which an object evolves.
6. Interactional relations: arise from an interplay between referenced objects, and include a causal element.
7. Causal relations: associate causes or circumstances with their effects or consequences and incorporate counteraction relations, which link preventive or curative measures to the causes, effects, symptoms, or patients they address.

[5] https://protege.stanford.edu

[6] In this paper, we use the following graphic notation: <Concepts> are capitalized and enclosed in single chevrons, 'terms' are not capitalized and enclosed in single quotes, "`concept relations`" are typeset in monospace font and enclosed in double quotes, *definitions of concepts from sources* and *example sentences* containing relation markers are in italics, "intensional definitions" are between double quotes, "*Classes*" in the HERO ontology are in italics and between double quotes.

3 The Conceptual Dimension of the Gut-Brain Axis Terminology

To analyze the conceptual dimension of terminology related to the gut–brain axis, we focused on the definitions of concepts. We then extracted term occurrences in biomedical literature as evidence for underlying concepts, and individuated the concept relations established between them. In this contribution, we present the definitions of key concepts that are of specific interest to the gut–brain axis, namely <Disease>, <Disorder>, <Finding>, <Microbiome>, <Microbiota>, and <Bacteria>. These foundational concepts occur extensively in the scientific literature related to the gut–brain axis and in biomedical sources, thereby underscoring their relevance within this specialized domain and the necessity of their definition.

The definitions of concepts comprised in the future terminology resource will be drafted as intensional definitions, following the formulation guidelines outlined in ISO 704 (International Organization for Standardization, 2022). The intended audience is physicians, language professionals, and non-experts of the domain, enabling the dissemination of specialized knowledge. The target category of users additionally include patients affected by different neurodegenerative diseases. Considering this, in the terminology resource, intensional definitions will be also offered to non-experts to provide them access to specialized domain knowledge. Definitions will, however, be complemented by patient-centered explanations, which will serve as simplified versions of the intensional definitions in plain language (Bonato, Vezzani, & Di Nunzio, 2026).

As for concept relations, the work focuses on associative relations. In particular, we will extract relation markers from biomedical literature to analyze the types of associative relations that link the microbiome, microbiota and bacteria to diseases, disorders, and findings. The relations will be examined in light of the typology of concept relations developed by Nuopponen (2022), a key reference typology in terminology science. We will then compare associative relations against those represented in the HERO Ontology.

3.1 Intensional Definitions: Methodology and Examples

As mentioned, intensional definitions are terminological definitions that aim to define concepts unambiguously, thereby supporting a clear and systematic representation of specialized knowledge of a domain within a given conceptual framework. In what follows, we present the methodology adopted for drafting intensional definitions of biomedical concepts. We then provide the definitions of the concepts analyzed in this study.

3.1.1 Methodology

The methodological approach used to formulate the intensional definitions of biomedical concepts consists of two steps: 1) the recollection of definitions of concepts from specialized biomedical sources; and 2) the formulation of the intensional definition drawing on knowledge derived from source definitions. In particular, the presented methodology was used to draft more than 1,200 definitions of biomedical concepts related to the gut–brain axis and associated medical conditions.

The main specialized medical sources consulted are SNOMED CT, MeSH Terms,[7] and the ICD-11.[8] We also used Ontobee (Ong et al., 2017) as a reference search engine to find relevant ontologies or vocabularies,[9] with NCIT (Golbeck et al., 2003; Sioutos et al., 2007) emerging as a key source.[10] In some cases, the intensional definitions were derived from multiple definitions retrieved from NCIT, where alternative definitions for the same concept can be present. The specialized sources cited were used as the primary reference for drafting the definitions. This consultation was indeed instrumental in capturing the specialized knowledge shared by experts in the biomedical domain. In some cases, however, scientific articles mainly retrieved from PubMed[11] were additionally consulted to support the drafting process. This step proved essential for retrieving definitions when these were not available in the specialized resources consulted. This was the case for newly emerging concepts related to the gut–brain axis, such as ‘psychobiotic’ or ‘paraprobiotic’, which were defined in scientific papers representing the most up-to-date knowledge in the

[7] https://www.ncbi.nlm.nih.gov/mesh/
[8] https://icd.who.int/en/
[9] https://ontobee.org
[10] http://purl.obolibrary.org/obo/ncit.owl
[11] https://pubmed.ncbi.nlm.nih.gov

domain. In addition to this circumstance, it was also important to draw on scientific papers in which knowledge about concepts already defined in specialized resources is represented, to derive updated information reflecting the conceptual evolution driven by the continuous advancement of scientific research.

Accordingly, for the recollection of definitions of concepts analyzed in the present work, in addition to resources such as NCIT and MeSH, scientific publications representing specialized knowledge in the biomedical domain were also consulted. A relevant aspect is that the set of definitions collected for the concepts <Microbiome>, <Microbiota>, and <Disease> also includes the corresponding definitions provided by the domain-specific ontology OHMI (Ontology of Host–Microbiome Interactions).[12]

To formulate intensional definitions of concepts, two components must always be included: 1) the immediate generic concept, and 2) the delimiting characteristics.

Concerning the identification of the immediate generic concept, it was fundamental to compare the various definitions collected for each concept (Bonato et al., 2026) given that, within specialized medical sources, concepts are often not defined in the same way and may respectively include terms that designate different generic concepts. In this respect, ISO 1087 (International Organization for Standardization, 2019) and ISO 704 (International Organization for Standardization, 2022) establish that the first element of an intensional definition is the most immediate generic concept. Considering this, it is necessary to compare the generic concepts used in different definitions, to determine which one is the immediate generic concept of the analyzed concept. An effective approach to do this is to represent the generic concepts included in the definitions in a concept system, which also needs to include the concept being defined. By establishing hierarchical relations among generic concepts, it becomes possible to identify the most immediate generic one.

By way of example, two distinct generic concepts are associated with <Dementia> in definitions contained in specialized medical resources. While in NCIT <Dementia> is defined as a condition,[13] MeSH Terms defines it as a mental disorder.[14] <Condition> is a superordinate concept with respect to <Mental disorder>, which, when comparing the two concepts under analysis, is in turn the immediate generic concept of <Dementia>. In the concept system, <Mental disorder> is related to <Condition> via a generic relation, and <Dementia> is linked to <Mental disorder> through another generic relation.

This comparison work also proved useful for identifying the essential characteristics of the defined concepts. For instance, the recurrence of certain characteristics across different definitions of a single concept may indicate that they are regarded as essential characteristics that make up that concept.

Moreover, a contrastive approach is particularly relevant for identifying the delimiting characteristics of a concept with respect to its coordinate concepts. This highlights the importance of collecting multiple definitions for the same concept. The adopted approach, indeed, not only enables a more comprehensive acquisition of specialized knowledge on a concept than reliance on a single definition, but also facilitates the identification of differences among the analyzed concepts from a conceptual viewpoint.

As an example of an intensional definition drawn from the corpus of definitions, we define the concept <Anxiety disorder> as a “mental disorder characterized by anxious feeling or fear”. The intensional definition of the concept <Mood disorder>, on the other hand, is “mental disorder characterized by disturbance in mood as the predominant feature”. The intensional definitions share the immediate generic concept <Mental disorder>, that indicates that the concepts <Anxiety disorder> and <Mood disorder> are coordinate concepts within the concept system.

By comparing the intensional definitions, it can be observed that the delimiting characteristics are the only components that distinguish one coordinate concept from the other. It should be noted, however, that we consider the collection of delimiting characteristics exhaustive when they allow the distinction of concepts that share the same immediate generic concept within the concept system being developed. That is to say, if another concept system is adopted, a different immediate generic concept may be associated, and consequently, the delimiting characteristics required to differentiate the subordinate concepts, in general, differ from one concept system to another. Within the developed concept system, <Anxiety disorder> and <Mood disorder> respectively constitute the immediate generic concepts for their subordinate concepts. Such subordinate concepts inherit the characteristics of the immediate generic concept, to which their own delimiting characteristics are then added. Based on this, for instance, the definition of each concept

[12] http://purl.obolibrary.org/obo/ohmi.owl
[13] http://purl.obolibrary.org/obo/NCIT_C4786
[14] https://www.ncbi.nlm.nih.gov/mesh/68003704

that shares the immediate generic concept <Anxiety disorder> will include ‘anxiety disorder’ as its first element, followed by the respective delimiting characteristics of the subordinate concept being defined.

3.1.2 Intensional Definitions of Analyzed Concepts

Based on this methodology, we formulated the intensional definitions of the concepts analyzed in this work.[15] The intensional definition of <Microbiome> is “biome that consists of the microorganisms, their genomes (i.e., genes) and the surrounding environment where the microorganisms reside". <Microbiota> is defined as “biota composed of microorganisms, including bacteria, viruses, archaea and fungi, that exist in an organism". A <Bacterium>, instead, is a “microorganism that is unicellular, prokaryotic, and reproduces by cell division". <Disease> can be defined as “pathologic process with a characteristic set of signs and symptoms", while <Disorder> refers to a “condition that is abnormal, that affects body function". To conclude, the intensional definition of <Finding> is “evidence, such as the clinical, laboratory or molecular evidence, or absence of evidence of disease, as well as the interpretation of data or observations resulting from planned evaluations".

In particular, by examining the definitions of the concepts <Disease> and <Disorder>, it can be observed that they do not share the same generic concept nor the same delimiting characteristics. Nevertheless, the distinction between these concepts is not evident in specialized resources. In fact, in SNOMED CT, the fully specified name “Disease (disorder)” is associated with multiple terms that can be used to designate the concept: the preferred term is ‘disease’; however, ‘disorder’ is considered an acceptable synonym.[16] In the National Cancer Institute Thesaurus (NCIT), instead, “Disease or Disorder” is represented as a single class.[17] The same applies to the class “Disease, Disorder, or Finding”, to which one definition is associated.[18] Although in ontology design a class can cover multiple concepts for practical reasons, from a terminological perspective three distinct concepts are grouped within the same class. However, distinguishing findings from diseases and disorders is crucial in the biomedical domain. For example, a bacterium could not directly influence a disease or disorder. Instead, it could have an influence on a finding associated with a disease or disorder. An example of this can be derived from a study that demonstrated that the L. reuteri bacterium improves social behavior in children affected by Autism Spectrum Disorder (Mazzone et al., 2024). Social behavior is a finding, as it is evidence linked to a determined disorder, rather than a disease or disorder. These considerations highlight that a thorough terminological analysis is necessary to distinguish between concepts in the biomedical domain, whose accurate definition is fundamental for the representation of specialized knowledge.

An additional consideration concerns the intensional definition of <Finding>. According to NCIT, [19] a finding may refer to the presence of clinical, laboratory, or molecular evidence. However, as also specified in the same definition, the absence of evidence of a disease is likewise considered a finding in itself. These circumstances can be regarded as mutually exclusive, since evidence may either be present or absent. For this reason, a disjunctive structure was retained in the intensional definition. In the intensional definition, indeed, the use of the conjunction “or” does not indicate that an alternative or a non-essential characteristic is included, but is maintained to ensure alignment with clinical practice.

3.2 Concept Relations in Biomedical Literature

In this work, we focus on the concept relations between: 1) the microbiome (and microbiota) and diseases, disorders, and findings, and 2) bacteria and diseases, disorders, and findings.

The identification of concept relations was carried out on a dataset of abstracts and titles of scientific papers retrieved from PubMed.[20] These texts are authored by experts in the biomedical domain; hence, they represent a valuable source for the verbalization of the specialized knowledge on the gut–brain axis. The dataset comprises over 1,000 abstracts relevant for the domain of study, retrieved through two distinct queries: 1) “gut microbiota” AND “mental health”, and 2) “gut microbiota” AND “Parkinson”. In total, the number of words amounts to more than 265,000. The papers from which the titles and abstracts were

[15] The list of biomedical concepts and their intensional definitions is available at: https://anonymous.4open.science/r/GutBrainAxisConcepts-4221/.
[16] http://purl.bioontology.org/ontology/SNOMEDCT/64572001
[17] http://purl.obolibrary.org/obo/NCIT_C2991
[18] http://purl.obolibrary.org/obo/NCIT_C7057
[19] http://purl.obolibrary.org/obo/NCIT_C3367
[20] In this preliminary study, the dataset consists of titles and abstracts of scientific papers. In the future, the dataset will include the full texts of scientific papers.

extracted cover the period from 2013 to 2023, thus reflecting the evolution of medical knowledge on the gut–brain interplay and its respective relation with mental health conditions and Parkinson's disease.

Given the sample of the corpus considered for this preliminary study, one annotator with prior annotation experience in the biomedical domain, as well as in terminology, extracted the terminological data analyzed in this paper. Indeed, the main goal was to analyze concept relations in biomedical literature and to propose an initial domain-specific typology of such relations, thereby laying the groundwork for the future development of a fully validated annotation scheme.

The process of identification of concept relations in the dataset was performed manually, and it followed two phases. In the first phase, the terms that designate the concepts analyzed were identified in the dataset. To identify terms, the annotator read all the abstracts and titles comprised in the dataset. Specifically, terms that designate the microbiome, microbiota, bacteria, diseases, disorders and findings were identified. Accordingly, the annotation process was guided by the criterion that terms designating, for instance, chemicals, drugs, or dietary supplements were excluded from selection, despite their presence in the texts. Considering diseases, disorders and findings as generic concepts, for the purposes of the analysis, also terms in the texts referring to mental health-related states and neurodegenerative diseases and disorders were identified, which, through a process of abstraction, were associated with these generic concepts. Both single-word and multi-word terms were identified. Within the corpus, terms denoting analyzed concepts are predominantly recurrent. For instance, the term 'microbiome', either as single-word term or constituting multi-word terms, counts more than 1,000 occurrences. A similar pattern is observed for the term 'microbiota', whose frequency in the corpus, however, exceeds 3,000 occurrences. With regard to diseases, disorders, or findings, their occurrences are also notably high in the corpus. By way of example, the term 'Parkinson's disease' appears more than 600 times, whereas the term 'disorder', either as a single-word term or as part of multi-word terms, occurs over 1,200 times. These frequencies, which refer to the analyzed corpus, reflect the query design and should therefore not be interpreted as an indication of unbiased domain coverage.

In the second phase, we examined sentences in which terms designating different types of microbiome, microbiota, and bacteria co-occurred with the terms that designate diseases, disorders, or findings, to observe the relations that link them. Specifically, once the terms were identified, we examined which relation markers were used to express a relation between concepts. In particular, relation identification was restricted to the sentence level, which opens up the study of document-level domain-specific concept relations as a promising direction for future terminology work. Among the relation markers identified, we focus in this paper on those regarded as especially relevant to the domain, which served as a basis for developing a preliminary typology of associative relations specific to the gut–brain axis domain. An additional criterion that guided the annotation was the consideration of verb phrases and nouns, in addition to verbs, as domain-specific relation markers, insofar as an explicit concept relation was verbalized and assessed between the respective pairs of analyzed concepts.

The approach used to identify relation markers that verbalize associative relations between concepts is therefore situated within the knowledge-rich context extraction framework (Bowker, 2003; Condamines, 2022; Marshman, 2022; Meyer, 2001). In particular, knowledge-rich context can be defined "as naturally occurring utterances that explicitly describe attributes of domain-specific concepts or semantic relations holding between them at a certain point in time [...]" (Meyer, 2001, p. 23).

In the following, we present the results of this analysis, citing the sentences that include relation markers expressing associative relations between the concepts under study. Within the context of the present work, however, an important clarification concerns an exclusion criterion adopted for the identification of associative relations. The aim of the study is to identify those associative relations that express, with a sufficient degree of specificity, a given type of association between the concepts under analysis. This implies that highly general relations expressed by predicates such as "`is associated with`" have not been included in the proposed typology. This decision is motivated by the need to ensure that the annotation captures relations that are highly informative and analytically relevant, rather than underspecified kinds of concept relations. We illustrate this choice by means of the following example, extracted from the corpus:

E. coli is associated with obesity and metabolic syndrome.

Generic Concepts
Microbiome or Microbiota | Bacteria | Disease, Disorder, or Finding

(a). Sentences expressing associative relations between the microbiome or microbiota and diseases, disorders, or findings.

(1). The contribution of the microbiota to **induce** gastrointestinal inflammation is hypothesized to be a key component of alpha-synuclein (aSyn) aggregation within the gastrointestinal (GI) tract in the pathological progression of Parkinson's disease (PD). [PMID: 36278360]

(2). The gut microbiota **regulates** autism-like behavior by mediating vitamin B(6) homeostasis in EphB6-deficient mice. [PMID: 32819434]

(3). Accumulating evidence demonstrated that the gut microbiome is emerging as a critical **regulator** of various neurodegenerative diseases (e.g., Parkinson's disease, Alzheimer's disease), suggesting that the crosstalk between the gut microbiome and neuroinflammation may contribute to the development of these diseases, for example, gut dysbiosis and microbiota-derived metabolites can trigger inflammation in the brain. [PMID: 38262196]

(4). However, it remains unclear how gut dysbiosis contributes to mental dysfunction, and how correction of the gut microbiota **alleviates** neuropsychiatric disorders. [PMID: 37927130]

(5). Our findings reveal for the first time, that Alzheimer's symptoms can be transferred to a healthy young organism via the gut microbiota, confirming a **causal role** of gut microbiota in Alzheimer's disease [...] [PMID: 37849234]

(6). There is a large and growing body of evidence that the gut microbiome **alters susceptibility to** great number of human diseases, including nervous system function. [PMID: 35868581]

(b). Sentences expressing associative relations between the bacteria and diseases, disorders, or findings.

(1). According to the infection hypothesis in relation to neurodegeneration, Helicobacter pylori as non-commensal gut microbiome has been advocated as **trigger** and/or **mediator** of neurodegenerative diseases, such as the development of Alzheimer's disease. [PMID: 36984505]

(2). Several studies concluded that Lactobacillus plantarum is a GABA-producing bacteria and can **alleviate** depression. [PMID: 36900437]

(3). Several studies showed an **increase** of Lactobacillus, Bifidobacterium, Verrucomicrobiaceae and Akkermansia in PD. A **decrease** of Faecalibacterium spp., Coprococcus spp., Blautia spp., Prevotella spp., and Prevotellaceae was observed in PD. [PMID: 29857583]

Fig. 1 Sentences expressing associative relations between the microbiome or microbiota and diseases, disorders, or findings (a) and between bacteria and disease, disorders, or findings (b). Terms that designate microbiome or microbiota are highlighted in pink. Terms that designate bacteria are reported in green. Terms that designate diseases, disorders, or findings are highlighted in blue. Associative relations are displayed in bold and underlined.

In this sentence, it is clearly stated that a bacterium, Escherichia coli (E. coli), is associated with two different conditions, namely obesity and metabolic syndrome. While this statement is relevant for establishing that a link exists between the specific bacterium and the two medical conditions, it does not specify the exact role of E. coli with respect to these conditions, nor the precise nature of the link between them. Within the framework of terminology work and of the present study, such a relation can therefore be considered insufficiently specific in terms of information conveyed. Given the generality of this relation and the limited degree of informativeness it provides, predicates of this type (i.e., "`is associated with`" or "`is linked to`") were not included in the extraction process. Consequently, such relations are not covered by the present analysis and typology of concept relations tailored to the gut–brain axis domain.

Moreover, it is relevant to note that another exclusion criterion should concern cases in which a predicate is present in a given sentence, but does not actually signal an associative relation. In light of this, a distinction should be made between false positives and relation markers that genuinely indicate the presence of an associative relation between concepts.

Figure 1 reports the sentences that express associative relations between the microbiome or microbiota and diseases, disorders, or findings (a), and sentences that express associative relations between bacteria and diseases, disorders, or findings (b). With particular reference to the last sentence (3b), 'PD' is a term variant of 'Parkinson's disease'.

The analysis of biomedical literature focused on the gut–brain axis and related medical conditions led to the identification of five associative relations linking microbiome and microbiota to diseases, disorders, or findings: 1) "`induce`", 2) "`regulate`", 3) "`alleviate`", 4) "`has causal role in`", and 5) "`alter susceptibility to`". [21] The following five associative relations, on the other hand, link bacteria to diseases, disorders, or findings: 1) "`trigger`", 2) "`mediate`", 3) "`alleviate`", 4) "`increase in`", and 5) "`decrease in`". As can be observed, the relation marker "`alleviate`" constitutes a relation marker signaling an associative relation holding between both microbiome or microbiota and diseases, disorders, or finding, and between bacteria and diseases, disorders, or findings. On the contrary, other relation markers

[21] The list of the associative relations and their typology is available at: https://anonymous.4open.science/r/GutBrainAxisConcepts-4221/

do not recur in both concept pairs. This observation, however, is limited to the corpus analyzed in this preliminary study. Further investigation is needed to determine whether other relation markers identified in this work may also function as predicates of associative relations involving both analyzed concept pairs.

Statistics on the occurrence of relation markers in the corpus can also be reported, specifically for those markers that verbalize associative relations between the concepts under investigation. For instance, concerning associative relations between the microbiome or microbiota and diseases, disorders, or findings, the relation marker "`induce`" occurs 4 times, the marker "`regulate`" 4 times, the marker "`alleviate`" 1 time, the marker "`has causal role in`" 1 time, and the marker "`alter susceptibility to`" 7 times. Concerning the relation markers that link bacteria to diseases, disorders, or findings, the relation marker "`trigger`" occurs 2 times, the marker "`mediate`" 1 times, the marker "`alleviate`" 5 times, the marker "`increase in`" 7 times, and the marker "`decrease in`" 4 times.

These frequencies may be partly attributable to the size of the corpus analyzed, but they may also reflect a second circumstance. Indeed, the biomedical literature more frequently reports associative relations involving gut–brain axis concepts in a highly general way, typically through predicates such as "`is associated with`" or “`is linked to`”. By contrast, associative relations verbalized through relation markers that convey a more specific type of link between concepts tend to occur less frequently and may therefore be more difficult to identify. At the same time, however, this further motivates the need to identify and analyze associative relations that convey more specific information, as they enable the representation of domain knowledge at an increasingly fine-grained level.

From a grammatical viewpoint, it can be observed that the different relation markers that verbalize associative relations in sentences are not only verbalized through verbs and verbal phrases. In fact, nouns that convey concept relations have also been considered.

For example, sentence 3a mentions the role of the gut microbiome as a regulator of various neurodegenerative diseases. In this sentence, the term ‘regulator’ grammatically functions as a noun. The use of the noun in the sentence indicates the presence of a concept relation between the gut microbiome and neurodegenerative diseases, implying the action of regulating. Therefore, the verb ‘regulate’ serves as the predicate of the concept relation. The same reasoning is applied to the term ‘mediator’, used in sentence 1b to express the mediating action performed by the bacterium Helicobacter pylori with respect to neurodegenerative diseases. In this case, the predicate used to express the concept relation is "`mediate`".

4 Typology of Biomedical Concept Relations

As mentioned above, different types of associative relation can be established between concepts. One of the objectives of this study is to investigate at a fine-grained level the associative relations that link the concepts analyzed by determining the exact nature of these connections. As an illustrative example, assessing the nature of the connection between the gut microbiota and a specific disease requires understanding whether the concepts are related through a unidirectional or bidirectional relation, and whether this relation is causal or not.

To identify the various types of concept relations at the terminological level, we refer to the classification of concept relations proposed by Nuopponen (2022). This typology provides a theoretical domain-agnostic framework that can be therefore applied in any specialized domain of activity. In this contribution, we adopt this typology as a framework for application to the gut–brain axis, which allows us to identify similarities in categorization and to underscore the need for new typologies specifically adapted to the representation of domain knowledge.

4.1 Mapped Concept Relations

The hierarchical relations identified in the dataset can be mapped to the generic super/subordination relations, which form part of the generic concept relations. The partitive relations, instead, can be mapped to the compound relations, included among the partitive relations, themselves categorized under the broader macro-category named contiguity relations.

Concerning associative relations, the relations found in the biomedical literature can, in some cases, be mapped to the types of concept relations. Relation markers such as "`induce`", "`has causal role in`", and "`trigger`", indeed, all correspond to causal concept relations. Specifically, they express a causator relation, defined as the connection between a causing agent and its effect. Consider the sentence:

*The contribution of the microbiota to **induce** gastrointestinal inflammation is hypothesized to be a key component of alpha-synuclein (aSyn) aggregation.*

In this instance, a causator relation is established, linking the causing agent (the microbiota) to the effect (gastrointestinal inflammation), verbalized by the marker "`induce`".

The marker "`alleviate`" also expresses a relation that can be mapped to the causal concept relations; however, it indicates a counteraction relation between two concepts. To illustrate this, we take the following sentence as an example:

*Several studies concluded that Lactobacillus plantarum is a GABA-producing bacteria and can **alleviate** depression.*

In the sentence, the Lactobacillus plantarum bacterium constitutes a preventive or curing measure, while depression is the alleviated health condition.

4.2 Typology of Gut-Brain Axis Concept Relations

The analysis of associative relations revealed that different relations can be considered specific to the gut–brain field of study. The reference is to relations respectively expressed by the predicates "`regulate`", "`mediate`", "`alter susceptibility to`", "`increase in`" and "`decrease in`". Considering this, we propose a typology of associative relations that link the microbiome, microbiota and bacteria to diseases, disorders or findings. These associative relations will be featured in the future terminology resource, to precisely express the relations holding between concepts related to the gut–brain interplay.

In particular, the identification of associative relations specifying the function exerted by the microbiota or microbiome on diseases, disorders, or findings, as well as by bacteria on diseases, disorders, or findings, is considered important in this study in light of the current state of knowledge regarding the link between the microbiome (and its components) and diseases or disorders of different origins, including neurodegenerative and mental health-related conditions. The investigation of these connections, indeed, is central to the research concerning the gut–brain axis, as currently the specification of the nature of the exact link between the microbiota and these conditions has not yet been fully elucidated. For this reason, it can be considered relevant to further investigate this issue by developing a domain-specific typology that can also be diachronically updated in accordance with scientific advances in the field.

This is also the reason underlying the fact that some of the associative relations identified in the analyzed corpus are not classified as causal relations in the present analysis. This decision stems from the fact that the precise nature of the relationship between the gut and the brain remains a matter of medical investigation. The same consideration applies to the relations between the gut–brain axis and the diseases or disorders currently linked to it. Although a substantial body of literature has documented a link between the gut microbiota and microbiome and different medical conditions, the directionality, mechanisms, and extent of this connection have not yet been fully established. Consequently, many causal claims found in the literature remain provisional, underlying probabilistic or indirect causation, and are often framed in terms of associations or hypotheses. From a methodological perspective, treating these relations as fully established causality could risk introducing a level of certainty that is not consistently supported by the current state of shared expert knowledge. For this reason, one of the broader objectives of the study, particularly when considered from a diachronic perspective, is precisely to investigate how scientific discourse evolves in the gut–brain axis domain, to ensure an exact representation of domain knowledge from a terminological perspective.

4.2.1 Functional Relations

Concerning domain-specific associative relations, the relation marker "`regulate`" is found in the following sentence:

*The gut microbiota **regulates** autism-like behavior.*

To understand the relation, we searched for the definition of <Regulate>. Initially, we found the following definition in the OBO Relations Ontology for the object property "regulates": *p regulates q iff p is causally upstream of q, the execution of p is not constant and varies according to specific conditions, and p influences the rate or magnitude of execution of q due to an effect either on some enabler of q or some enabler* *of a part of q.*[22] In the sentence where the relation marker is present, however, the verb describes a function that the gut microbiota exerts on a finding (which, in this case, is autism-like behavior), and a causal relation is not explicitly established between the concepts. We also consulted the definition of the class "regulation of cell adhesion" in Gene Ontology, to observe how <Regulation> is defined: *[a]ny process that modulates the frequency, rate or extent of attachment of a cell to another cell or to the extracellular matrix.*[23] From this definition, it can be derived that <Regulation> is considered as a process that modulates cell adhesion, without references to causal relations. The definition of <Modulation>, indeed, is *[t]o adjust, or change.*[24]

The relation cannot be classified within the interactional relations proposed by Nuopponen (2022), because these presuppose an interplay between the concepts involved in the relation. For instance, among the various types of interactional relations proposed by the author, two types of associative relations are included: 1) dependency relation, established between "someone or something that is determined or conditioned by another and this other" (p. 82), and 2) correlation relation, that link "two phenomena, when one changes, the other also changes" (p. 82). Dependency relations presuppose that the entities that take part in the relation are "mutually dependent on each other" (p. 82). The sentence mentioned above, however, represents a one-way relation. The gut microbiota is the agent performing the regulating action, while autism-like behavior is the target of that action (see sentence 6a, Figure 1). Therefore, the relationship is unidirectional: the gut microbiota regulates that specific behavior. It does not imply that autism-like behavior regulates or affects the gut microbiota, nor that there is a mutual dependency between the two. This relation cannot even be classified as a correlation relation, since a correlation relation is "a reciprocal one between two entities, which means that when one changes, the other also changes" (Nuopponen, 2022, p. 83). In this case, however, only autism-like behavior is regulated, whereas the gut microbiota is not described as being affected by changes in behavior. This case therefore underscores the need to propose a new typology of concept relations that is specifically suited to represent knowledge related to the specific domain under analysis.

Based on this reasoning, we propose to create a macro-category of associative relations that groups the functions performed by the microbiota, the microbiome, or bacteria on diseases, disorders or findings. This macro-category is called functional relations. The relation verbalized by the relation marker "`regulate`" can be classified as a functional relation, and within the category of functional relations it constitutes a regulation relation, in which the regulating agent exerts a regulatory function on a disease, disorder or finding. In the cited sentence, the regulating agent (gut microbiota) exerts a regulating function on a finding (autism-like behavior).

The functional relations macro-category can also be used to classify two other associative relations identified in the analyzed biomedical literature, respectively verbalized through the relation markers "`mediate`" and "`alter susceptibility to`".

Concerning the relation marker "`mediate`", the following definition of <Mediate> is in the Merriam Webster Medical Dictionary: *to transmit or carry (as a physical process or effect) as an intermediate mechanism or agency.*[25] In NCIT, the concept <Mediator> is defined as *[a]n agent that acts as a link between parties, objects, or actions.*[26] Also in this case, "`mediate`" is a relation marker that conveys a functional relation. Specifically, the relation can be expressed as a mediation relation. Referring to the following sentence, in which the relation marker appears, it can thus be noted that the mediator (Helicobacter pylori) exerts a mediating function with respect to a group of diseases (neurodegenerative diseases).

> *Helicobacter pylori as non-commensal gut microbiome has been advocated as trigger and/or* ***mediator*** *of neurodegenerative diseases*

Functional relations can also encompass the relation verbalized by the relation marker "`alter susceptibility to`". The relation marker is present in the following sentence:

> *There is a large and growing body of evidence that the gut microbiome* ***alters susceptibility to*** *great number of human diseases*

[22] http://purl.obolibrary.org/obo/RO_0002211
[23] http://purl.obolibrary.org/obo/GO_0030155
[24] http://purl.obolibrary.org/obo/NCIT_C90063
[25] https://www.merriam-webster.com/dictionary/mediate#medicalDictionary
[26] http://purl.obolibrary.org/obo/NCIT_C61426

In this sentence, it is clear that the gut microbiome performs the function of altering susceptibility to certain diseases. <Susceptibility> is defined as *the constitutional or inborn state disposing to a disease, group of diseases, or metabolic or structural anomaly.* [27] A <Disposition> is *[t]he tendency of something to act in a certain manner under given circumstances resulting from natural constitution; nature; quality; orderly arrangement.* [28] Considering these definitions, it can be observed that the alteration of susceptibility does not imply a causal influence on the onset of diseases, disorders, or findings. In other words, the fact that the gut microbiome may, for example, increase the probability of developing a disease by altering susceptibility does not imply that the disease will necessarily occur in an individual. For this reason, a causal relationship is not claimed in the context of the sentence. The relation verbalized by "`alter susceptibility to`", therefore, can be considered a susceptibility relation in which a susceptibility factor (the gut microbiome) alters the susceptibility to diseases affecting humans.

Functional relations are thus framed in this work as associative relations that capture the specific function exerted by the microbiota or microbiome on diseases, disorders, and related findings. A larger corpus is nevertheless required in order to determine whether the proposed relation is case-sensitive or whether it holds consistently across all occurrences in papers related to the gut–brain axis, with specific reference to relations between the microbiome or microbiota and diseases, disorders, or findings. This is particularly necessary in order to verify whether some relations may involve causality. In particular, as noted above, it is also essential to approach this analysis from a diachronic perspective in order to trace the future evolution of scientific knowledge surrounding the gut–brain axis domain.

4.2.2 Quantitative Relations

Two additional relations derived from the literature can be considered domain-specific. These refer to the associative connections indicated by the phrases "`increase in`" and "`decrease in`". The markers "`increase in`" and "`decrease in`" denote a quantitative change in a biological variable. In the analyzed texts, these markers do not establish an interactional relation. Indeed, the context of these markers suggests a rise or fall in specific bacteria associated with Parkinson's disease; however, there is no mention of a dynamic interaction between bacteria and Parkinson's disease.

Thus, we categorize this as a quantitative relation, which links a measured variable (bacteria) to a disease, disorder, or finding. This type of associative relation, specifically, can be considered important for the representation of knowledge related to the gut–brain axis. Indeed, for example, a change in the abundance of bacteria (such as an increase or decrease) observed in relation to the onset of a disease may represent highly relevant information, as it may signal clinically meaningful quantitative changes from a biological or pathological perspective.

The quantitative relation can be viewed as a broad category that encompasses various specific relationships denoted by different quantitative indicators, such as relative abundance or biomarker levels, thus providing a more precise specification of the relations between relevant concepts.

5 Comparison with HERO

Among the goals of HERO is the representation of information automatically extracted from biomedical texts concerning the gut–brain interplay. In particular, HERO has served as the reference schema for representing gut–brain axis entities and relations extracted from thousands of PubMed abstracts, and was later used in an evaluation challenge to test automatic information extraction systems in this domain (Martinelli et al., 2025; Nentidis et al., 2026); all the data can be openly downloaded and explored online (Irrera, Martinelli, Piron, & Silvello, 2026).

This section aims to assess how corpus-derived knowledge about concepts relevant to the gut–brain axis domain aligns with an existing reference ontology modeling the same domain, as well as at identifying potential gaps between a terminology-driven perspective and an ontological representation. This also highlights differences in the level of granularity of domain-specific information extracted from our corpus-derived evidence compared to a curated ontology. These considerations are regarded as a preliminary analysis that will be further developed in future work, based on ongoing corpus analysis including a diachronic perspective, and in light of further developments of the HERO ontology.

[27] http://purl.obolibrary.org/obo/NCIT_C16505
[28] http://purl.obolibrary.org/obo/NCIT_C41205

In particular, the comparison is currently carried out between HERO and the terminological data being collected to populate a dedicated terminology resource for the gut–brain axis domain. The terminology resource is intended to provide a structured representation of domain knowledge, including concepts, terms, definitions, as well as hierarchical and associative concept relations. For this reason, the comparison with HERO focuses not only on the concepts identified through terminology work, but also on the level of granularity at which domain knowledge can be represented through associative relations.

The differences between terminology-derived concepts and concept relations and HERO offer a case-specific example of how established distinctions between terminological and ontological representations occur in the gut–brain axis domain. While a terminology resource can represent the conceptual and linguistic dimensions of specialized knowledge, ontologies seek to provide an explicit and formal specification of conceptualization shared in a given domain. Thus, we expect differences in the level of granularity of domain-specific information extracted from our corpus-derived evidence from a terminological viewpoint compared to a curated ontology.

In particular, by analyzing the concepts defined at the terminological level and those in the ontology it can be observed that, in some cases, concepts in terminology can be considered as specific instances with respect to the classes used in the ontology to represent them.[29] For example, in terminology, we distinguish between <Disease>, <Disorder> and <Finding>, as they are three distinct concepts linguistically verbalized by three different terms. In the ontology, this distinction is not made, as the class "*Disease, Disorder, or Finding*" groups together the three concepts, following NCIT.[30] The mismatch arises because, from a computer science perspective, the information extraction task was not tailored to extract such fine-grained concepts, but the goal was to extract the interaction between the microbiome and its effect on symptoms or diseases indistinctively.

The same reasoning applies to the associative relation "`is linked to`", which links the class "*Microbiome*" to the class "*Disease, Disorder or Finding*" within the ontology.[31] The relation indicates that there is a connection between the microbiome and diseases, disorders or findings, without specifying the exact nature of the link. The associative relations identified through terminology work, namely "`induce`", "`regulate`", "`alleviate`", "`has causal role in`" and "`alter susceptibility to`", express the nature of this connection in a more specific way. However, for the task of information extraction, distinguishing between more specific relations would yield too few instances to build a dataset suitable for training automatic information extraction systems.

Further considerations concern the concept <Bacterium>, defined at the terminological level as a "microorganism that is unicellular, prokaryotic, and reproduces by cell division" (Cfr. Section 3.1). In information extraction, it has been observed that within scientific papers reference is made to bacterial species or families. For this reason, in HERO, bacteria are not represented through a class named "*Bacterium*". Instead, the concept <Bacterium> defined at the terminological level is mapped in Hereditary Ontology (HERO) to the NCIT taxonomy of bacteria, distinguishing between "*Species*", "*Genus*", "*Family*", "*Order*", "*Class*", "*Phylum*", "*Super Kingdom*", and "*Domain*". [32] In this case, HERO presents a level of granularity in terms of taxonomic depth that is higher with respect to the level of analysis presented from a terminological perspective in the current state of the work.

In the ontology, three relations link bacteria to diseases, disorders, or findings: 1) the relations "`influence`" [33] and "`interact`" [34], that link bacteria to diseases, disorders and findings, and 2) the relation "`change abundance`", that links diseases, disorders and findings to bacteria. [35] In HERO, the relation "`influence`" implies a causal relation, but a different level of specificity can be noted with respect to the associative relation "`trigger`". The associative relation "`interact`", instead, that does not indicate causality, and can be considered as more generic with respect to the associative relations "`mediate`" and "`alleviate`" (Cfr. Section 3.2).

This latter consideration is based on the observation of the specific function exerted by bacteria, particularly considering whether it is beneficial or detrimental, and whether a definite role is performed by

[29] A table comparing all concepts and relations defined at the terminological level and in HERO is available at: https://anonymous.4open.science/r/GutBrainAxisConcepts-4221/.
[30] https://w3id.org/brainteaser/ontology/schema/DiseaseDisorderOrFinding
[31] https://w3id.org/hereditary/ontology/gutbrain/schema/isLinkedTo
[32] https://w3id.org/hereditary/ontology/gutbrain/schema/Domain
[33] https://w3id.org/hereditary/ontology/gutbrain/schema/influence
[34] https://w3id.org/hereditary/ontology/gutbrain/schema/interact
[35] https://w3id.org/hereditary/ontology/gutbrain/schema/changeAbundance

Table 1 Mapping between terminology-derived associative relations and HERO relations. We report the possible triples we can extract from the sentences in Figure 1. The first column reports the identifier of the sentence in the figure.

SID	Concept (Head)	Terminology-derived predicate	HERO-mapping	Concept (Tail)
1a	<Microbiota>	induce	is linked to	<Gastrointestinal inflammation>
2a	<Gut microbiota>	regulate	is linked to	<Autism-like behavior>
3a	<Gut microbiome>	regulate	is linked to	<Neurodegenerative disease>
3a	<Gut microbiome>	regulate	is linked to	<Parkinson's disease>
3a	<Gut microbiome>	regulate	is linked to	<Alzheimer's disease>
4a	<Gut microbiota>	alleviate	interact	<Neuropsychiatric disorder>
5a	<Gut microbiota>	has causal role in	is linked to	<Alzheimer's disease>
6a	<Gut microbiome>	alter susceptibility to	is linked to	<Human disease>
1b	<Helicobacter pylori>	trigger	influence	<Neurodegenerative disease>
1b	<Helicobacter pylori>	trigger	influence	<Alzheimer's disease>
1b	<Helicobacter pylori>	mediate	interact	<Neurodegenerative disease>
1b	<Helicobacter pylori>	mediate	interact	<Alzheimer's disease>
2b	<Lactobacillus plantarum>	alleviate	interact	<Depression>
3b	<Lactobacillus>	increase in	- - -	<Parkinson's disease>
3b	<Bifidobacterium>	increase in	- - -	<Parkinson's disease>
3b	<Verrucomicrobiaceae>	increase in	- - -	<Parkinson's disease>
3b	<Akkermansia>	increase in	- - -	<Parkinson's disease>
3b	<Faecalibacterium spp.>	decrease in	- - -	<Parkinson's disease>
3b	<Coprococcus spp.>	decrease in	- - -	<Parkinson's disease>
3b	<Blautia spp.>	decrease in	- - -	<Parkinson's disease>
3b	<Prevotella spp.>	decrease in	- - -	<Parkinson's disease>
3b	<Prevotellaceae>	decrease in	- - -	<Parkinson's disease>

bacteria with respect to diseases, disorders, or findings. Following this line of reasoning, while "`interact`" may refer to various types of interaction and, in this case, implies a mode that is not explicitly specified, the act of mediating presupposes a determined functional role, namely that of an intermediary. The act of alleviating presupposes a beneficial function, in this case exerted by Lactobacillus plantarum on depression (see sentence 2b, Figure 1). From this perspective, the positive or negative function exerted by the microbiota, the microbiome, or bacteria on a disease, disorder, or finding may constitute a relevant analytical criterion for the study of associative relations in the gut–brain axis domain.

We can additionally observe that the relation "`change abundance`", used in HERO ontology, encodes a causal interpretation that goes beyond the purely quantitative relation expressed by the predicates "`increase in`" and "`decrease in`" identified through terminology work. In particular, the relation in HERO implies a causal relation between disease, disorder, or findings and bacteria but the associative relations simply record the presence of a change in quantity, without implying a cause-effect relation.

Table 1 presents the concept–relation–concept triplets that can be extracted from the sentences in Figure 1. We also report the mapping between the relations identified in the corpus and the corresponding relation markers adopted in HERO.

6 Conclusions and Future Work

In this work, we presented the terminology work carried out to analyze the conceptual dimension of terminology related to the gut–brain interplay.

We focused on the intensional definitions and associative relations linking the concepts <Microbiome>, <Microbiota>, <Bacterium>, <Disease>, <Disorder>, and <Finding>. To identify the types of associative relations found in the biomedical literature, we mapped them to the typology proposed by Nuopponen (2022). Our analysis suggests that some associative relations were not straightforwardly captured without further disambiguation, as they are specific to the gut–brain field of study. Indeed, associative relations may require domain-specific refinement. Building on this, we proposed two additional macro-categories of associative relations, specific to the gut–brain axis domain: functional relations and quantitative relations.

We then analyzed the defined concepts and examined the associative relations compared to those represented in HERO, to bridge them and identify potential differences in the representation of specialized medical knowledge. The comparison with HERO highlights differences in granularity and representation between terminology-based and ontology-based modeling of biomedical knowledge. These differences reflect the distinct nature of terminological and ontological representation of specialized knowledge concerning the analyzed domain: while in terminology specialized knowledge representation follows the principles that underlie terminology work, an ontology provides a formal and explicit representation of how automatically extracted information is structured in scientific texts.

As future work, we plan to further explore concept relations related to the gut–brain interplay, expanding the analysis. Indeed, this preliminary work provided a basis for extending the analysis to a broader set of relation markers and for refining a domain-specific typology of associative relations. In particular, the aim is to develop a comprehensive domain-specific typology of concept relations. Moreover, a diachronic analysis of associative relations will be carried out on an ongoing basis, drawing on continuously updated scientific literature to accurately trace the evolution of specialized knowledge on the gut–brain interplay.

A further study will focus on the relations established between the concepts analyzed in the paper that are expressed in biomedical literature through the relation markers "`is associated with`" and "`is linked to`", which are highly general and lack informational specificity, thus requiring in-depth analysis.

Another aspect to consider is that the present study relied on a single annotator; therefore, inter-annotator agreement could not be assessed. Future research could extend the study by involving multiple expert annotators and reporting inter-annotator agreement metrics on a stratified sample. This would also improve the reproducibility of the annotation process.

We also aim to further explore the interconnections between the representation of specialized knowledge from a strictly terminological perspective and in the HERO ontology.

This work, therefore, establishes the basis for an extended analysis of terminology related to the gut–brain interplay and associated medical conditions. The methodology for the formulation of intensional definitions and the typology of concept relations, indeed, will be applied to examine the conceptual dimension of this terminology within a future domain-specific terminology resource.

Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT-5 in order to: Grammar and spelling check, Paraphrase and reword. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

Acknowledgements

This work has been partially supported by the HEREDITARY Project, as part of the European Union's Horizon Europe research and innovation programme under grant agreement No GA 101137074, and it has been part of the initiatives of the Center for Studies in Computational Terminology (CENTRICO) of the University of Padua and in the research directions of the Italian Common Language Resources and Technology Infrastructure CLARIN-IT.

References

Bonato, V., Vezzani, F., Di Nunzio, G.M. (2026). From complexity to inclusivity: A methodology for drafting patient-centered explanations of gut-brain axis concepts. G.M. Di Nunzio, F. Vezzani, L. Ermakova, H. Azarbonyad, & J. Kamps (Eds.), *Proceedings of the 2nd Workshop on Evaluating Text Difficulty in a Multilingual Context (DeTermIt! 2026)* (p. 33-41). ELRA Language Resources Association. Retrieved from http://lrec-conf.org/proceedings/lrec2026/workshops/determit/2026.determit-1.0.pdf

Bowker, L. (2003). Lexical knowledge patterns, semantic relations, and language varieties: Exploring the possibilities for refining information retrieval in an international context. *Cataloging & Classification Quarterly*, *37*(1-2), 153–171, https://doi.org/10.1300/J104v37n01_11

Cabezas-García, M., & León-Araúz, P. (2020). Term variation in terminographic resources: A review and a proposal. Z. Gavriilidou, M. Mitsiaki, & A. Fliatouras (Eds.), *Lexicography for Inclusion: Proceedings* *of the 19th EURALEX International Congress, 7-11 September 2021, Alexandroupolis, Vol. 1* (pp. 405–414). Democritus University of Thrace. Retrieved from https://www.euralex.org/elx_proceedings/Euralex2020-2021/EURALEX2020-2021_Vol1-p405-414.pdf

Carvalho, S. (2018). *A terminological approach to knowledge organization within the scope of endometriosis: the EndoTerm project* (PhD thesis, Universidade NOVA de Lisboa). Retrieved from http://hdl.handle.net/10362/49745

Chang, E., & Mostafa, J. (2021). The use of SNOMED CT, 2013-2020: A literature review. *Journal of the American Medical Informatics Association*, *28*(9), 2017-2026, https://doi.org/10.1093/jamia/ocab084

Condamines, A. (2022). How the notion of "Knowledge Rich Context" can be characterized today. *Frontiers in Communication*, *7*, Article 824711, https://doi.org/10.3389/fcomm.2022.824711

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

Faggioli, G., Menotti, L., Marchesin, S., Chiò, A., Dagliati, A., de Carvalho, M., … Ferro, N. (2024). An extensible and unifying approach to retrospective clinical data modeling: the BrainTeaser Ontology. *Journal of Biomedical Semantics*, *15*(1), Article 16, https://doi.org/10.1186/s13326-024-00317-y

Fan, T., Li, L., Chen, Y. (2024). Causal relationships between gut microbiota and depression/anxiety disorders: A 2-sample mendelian randomization study. *Medicine*, *103*(36), Article e39543, https://doi.org/10.1097/MD.0000000000039543

Golbeck, J., Fragoso, G., Hartel, F., Hendler, J., Oberthaler, J., Parsia, B. (2003). The National Cancer Institute's Thésaurus and Ontology. *Journal of Web Semantics*, *1*(1), 75-80, https://doi.org/10.1016/j.websem.2003.07.007

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

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

Irrera, O., Martinelli, M., Piron, S., Silvello, G. (2026). GutBrainKB: Exploring the gut-brain interaction through a reliable biomedical KB. R. Campos et al. (Eds.), *Advances in Information Retrieval: 48th European Conference on Information Retrieval, ECIR 2026, Delft, The Netherlands, March 29 - April 2, 2026, Proceedings, Part IV* (pp. 136–141). Springer. https://doi.org/10.1007/978-3-032-21321-1_19

Lai, Y., & Xiong, P. (2025). Analysis of gut microbiota and depression and anxiety: Mendelian randomization from three datasets. *General Hospital Psychiatry*, *94*, 206-218, https://doi.org/10.1016/j.genhosppsych.2025.03.012

León-Araúz, P., & Faber, P. (2010). Natural and contextual constraints for domain-specific relations. V.B. Mititelu, V. Pekar, & E. Barbu (Eds.), *Proceedings of the Workshop Semantic Relations. Theory and Applications* (p. 12-17). Retrieved from http://www.lrec-conf.org/proceedings/lrec2010/workshops/W9.pdf

Löckinger, G., Kockaert, H.J., Budin, G. (2015). Intensional definitions. In H.J. Kockaert & F. Steurs (Eds.), *Handbook of Terminology* (Vol. 1, pp. 60–81). John Benjamins Publishing Company. https://doi.org/10.1075/hot.1.int1

Maroto García, N., & Alcina, A. (2009). Formal description of conceptual relationships with a view to implementing them in the ontology editor Protégé. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *15*(2), 232-257, https://doi.org/10.1075/term.15.2.04mar

Marshman, E. (2022). Knowledge patterns in corpora. P. Faber & M.-C. L'Homme (Eds.), *Theoretical Perspectives on Terminology: Explaining Terms, Concepts and Specialized Knowledge* (p. 291-310). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.13mar

Martinelli, M., Marchesin, S., Bonato, V., Di Nunzio, G.M., Ferro, N., Irrera, O., … Silvello, G. (2026). A domain-specific curated benchmark for entity and document-level relation extraction. V. Demberg, K. Inui, & L. Marquez (Eds.), *Findings of the Association for Computational Linguistics: EACL 2026* (pp. 5693–5711). Association for Computational Linguistics. Retrieved from https://aclanthology.org/2026.findings-eacl.301/

Martinelli, M., Silvello, G., Bonato, V., Di Nunzio, G.M., Ferro, N., Irrera, O., … Vezzani, F. (2025). Overview of GutBrainIE@CLEF 2025: Gut-brain interplay information extraction. G. Faggioli, N. Ferro, P. Rosso, & D. Spina (Eds.), *Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2025* (pp. 65–98). CEUR-WS. Retrieved from https://ceur-ws.org/Vol-4038/paper_5.pdf

Mazzone, L., Dooling, S.W., Volpe, E., Uljarević, M., Waters, J.L., Sabatini, A., … Hardan, A.Y. (2024). Precision microbial intervention improves social behavior but not autism severity: A pilot double-blind randomized placebo-controlled trial. *Cell Host & Microbe*, *32*(1), 106-116.e6, https://doi.org/10.1016/j.chom.2023.11.021

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* (p. 279-302). John Benjamins Publishing Company. https://doi.org/10.1075/nlp.2.15mey

Nentidis, A., Katsimpras, G., Krithara, A., Krallinger, M., Rodríguez-Ortega, M., Rodriguez-López, E., … Paliouras, G. (2026). BioASQ at CLEF2026: the fourteenth edition of the large-scale biomedical semantic indexing and question answering challenge. R. Campos et al. (Eds.), *Advances in Information Retrieval - 48th European Conference on Information Retrieval, ECIR 2026, Delft, The Netherlands, March 29 - April 2, 2026, Proceedings, Part IV* (pp. 315–324). Springer. https://doi.org/10.1007/978-3-032-21321-1_42

Nuopponen, A. (1994). *Begreppssystem för terminologisk analys* (PhD thesis). University of Vaasa.

Nuopponen, A. (2005). Concept relations: An update of a concept relation classification. B.N. Madsen & H.E. Thomsen (Eds.), *Terminology and Content Development - TKE 2005, 7th International Conference on Terminology and Knowledge Engineering* (Vol. 5, pp. 127–138). Copenhagen: Litera.

Nuopponen, A. (2022). Conceptual relations: From the General Theory of Terminology to knowledge bases. P. Faber & M.-C. L'Homme (Eds.), *Theoretical Perspectives on Terminology: Explaining Terms, Concepts and Specialized Knowledge* (Vol. 23, pp. 63–86). John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.23.03nuo

Ong, E., Xiang, Z., Zhao, B., Liu, Y., Lin, Y., Zheng, J., … He, Y. (2017). Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration. *Nucleic Acids Research*, *45*(D1), D347–D352, https://doi.org/10.1093/nar/gkw918

Roche, C., & Papadopoulou, M. (2024). Terminology evaluation: The contribution of formal approaches. F. Vezzani et al. (Eds.), *Proceedings of the 3rd International Conference on Multilingual Digital Terminology Today (MDTT 2024), Granada, Spain, June 27-28, 2024.* CEUR-WS. Retrieved from https://ceur-ws.org/Vol-3703/paper3.pdf

Sambre, P., & Wermuth, C. (2015). Associative relations and instrumentality in causality. In H.J. Kockaert & F. Steurs (Eds.), *Handbook of Terminology* (Vol. 1, p. 101-127). John Benjamins Publishing Company. https://doi.org/10.1075/hot.1.ass1

San Martín, A. (2022, 03). A flexible approach to terminological definitions: Representing thematic variation. *International Journal of Lexicography*, *35*(1), 53-74, https://doi.org/10.1093/ijl/ecab013

Seppälä, S. (2009). A proposal for a framework to evaluate feature relevance for terminographic definitions. G. Sierra, M. Pozzi, & J.-M. Torres (Eds.), *Proceedings of the 1st Workshop on Definition Extraction* (pp. 47–53). USA: Association for Computational Linguistics. Retrieved from https://aclanthology.org/W09-4408/

Seppälä, S. (2015). An ontological framework for modeling the contents of definitions. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *21*(1), 23-50, https://doi.org/10.1075/term.21.1.02sep

Silvello, G., & Menotti, L. (2025). *The Hereditary Ontology for Phenoclinical Data (HERO-Phenoclinical).* Zenodo. Retrieved from https://doi.org/10.5281/zenodo.15519053

Sioutos, N., de Coronado, S., Haber, M.W., Hartel, F.W., Shaiu, W.-L., Wright, L.W. (2007). NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information. *Journal of Biomedical Informatics*, *40*(1), 30-43, https://doi.org/10.1016/j.jbi.2006.02.013

Temmerman, R. (2000). *Towards new ways of terminology description: The sociocognitive approach.* John Benjamins Publishing Company. https://doi.org/10.1075/tlrp.3

Vezzani, F. (2022). *Terminologie numérique: Conception, Représentation et Gestion.* Peter Lang Verlag. https://doi.org/10.3726/b19407

Vezzani, F., & Costa, R. (2024). Variation in psychopathological terminology: A case study on Body Dysmorphic Disorder. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *30*(1), 81-106, https://doi.org/10.1075/term.00078.vez

Vezzani, F., Di Nunzio, G.M., Silecchia, S. (2022). The phraseology of international disarmament treaties: the DITTO terminology resource. *Umanistica Digitale*, *6*(14), 91–117, https://doi.org/10.6092/issn.2532-8816/14796

Xiong, R.-G., Li, J., Cheng, J., Zhou, D.-D., Wu, S.-X., Huang, S.-Y., … Li, H.-B. (2023). The role of gut microbiota in anxiety, depression, and other mental disorders as well as the protective effects of dietary components. *Nutrients*, *15*(14), 3258, https://doi.org/10.3390/nu15143258