Defining Concepts, Generating Knowledge: A Structured Review of Terminological Definitions

Giorgio Maria Di Nunzio

University of Padua, Italy

Abstract

Terminological definitions play a central role in the representation, delimitation, and communication of specialized concepts. While they have traditionally been studied within terminology theory, specialized lexicography, and knowledge organization, they have recently acquired renewed relevance in computational settings, including automatic definition extraction, definition generation, ontology learning, and terminology-augmented generation. This paper presents a structured review of research on terminological definitions, tracing the transition from theoretical and methodological approaches to computational and generative perspectives. The review is based on a year-by-year screening of Google Scholar results from 2000 to 2026, complemented by historically relevant works from the 1990s. The paper identifies major research directions, including normative principles for definition writing, corpus-based and communicative approaches, knowledge-oriented modeling, and automatic definition extraction and generation. The review highlights how terminological definitions have progressively moved from manually authored components of terminological resources to objects of computational modeling, evaluation, and reuse.

Keywords

Terminological definitions, Computational terminology, Definition generation

1. Introduction

Terminological definitions occupy a central position in terminology work because they make explicit the conceptual content associated with specialized terms. In classical terminology, definitions are not merely explanatory paraphrases of words, but statements that delimit concepts within a domain-specific conceptual system. They support concept identification, term normalization, multilingual equivalence, knowledge organization, and the construction of terminological resources. For this reason, the study of terminological definitions has traditionally been connected to issues such as concept orientation, intensional definition, genus–differentia structures, conceptual relations, and the distinction between linguistic and conceptual levels of description.

The renewed interest in terminological definitions is also linked to recent developments in computational terminology, natural language processing, and generative artificial intelligence. Tasks such as automatic term extraction, definition extraction, definition generation, ontology learning, and terminology-augmented generation require models that can identify, represent, and produce definitional knowledge. In this context, terminological definitions become both an object of linguistic and conceptual analysis and a computational target. They are no longer used only as manually authored components of terminological databases, but also as data for training, evaluating, and controlling automatic systems.

This shift raises several research questions. First, it is necessary to clarify how terminological definitions have been understood across different research traditions, from normative and concept-oriented approaches to corpus-based, communicative, cognitive, and computational perspectives. Second, it is important to examine how the formal and functional properties of definitions have been operationalized in computational settings. Third, recent work on automatic definition extraction and generation requires a better connection with earlier terminology research, especially with respect to definition quality, conceptual adequacy, domain specificity, and expert validation.

International Workshop on Definition and Term Extraction Challenge (DETECH) 2026, June 24, 2026, Zadar, Croatia
giorgiomaria.dinunzio@unipd.it (G. Di Nunzio)
0000-0001-9709-6392 (G. Di Nunzio)

This paper presents a structured review of research on terminological definitions, with particular attention to the transition from theoretical and methodological studies to computational approaches. The review is motivated by the DETECH 2026 evaluation challenge,[1] where definition generation is one of the central tasks [1]. The aim is not to provide an exhaustive bibliometric mapping of all publications mentioning definitions in terminology, but rather to identify the main lines of development that are relevant for current work on definition-oriented terminology processing. The review is especially motivated by the growing importance of definition generation and evaluation in shared tasks and computational terminology scenarios, where the quality of generated definitions depends not only on linguistic fluency, but also on conceptual correctness, domain adequacy, and terminological precision.

The contribution of the paper is threefold. First, it reconstructs the main research directions that have shaped the study of terminological definitions from the 1990s to recent computational work. Second, it proposes a thematic organization of the literature, distinguishing between theoretical foundations, corpus-based and communicative approaches, knowledge-oriented modeling, and automatic definition extraction and generation. Third, it discusses open issues for future research, with particular emphasis on the evaluation of terminological definitions in computational settings.

2. Review Procedure

The review was conducted as a structured literature review aimed at identifying major research trends on terminological definitions from the early development of contemporary terminology theory to recent computational and generative approaches. The procedure was designed to be transparent and consistent, while acknowledging that the goal of the paper is not a fully reproducible bibliometric study, but a research-oriented synthesis of relevant literature across partially overlapping fields.

The search was carried out using Google Scholar, with the query string *"terminological definitions"* as the main entry point. Google Scholar was selected because the topic is distributed across several research communities, including terminology studies, specialized lexicography, translation studies, knowledge representation, ontology engineering, natural language processing, and generative artificial intelligence. This choice made it possible to retrieve publications from journals and conference proceedings that are not always indexed consistently in a single domain-specific bibliographic database.

The selection process followed two complementary steps. First, a small set of historically relevant works from the 1990s was retained in order to account for the theoretical foundations of terminological definition writing and for the transition from classical terminology approaches to more flexible, corpus-based, cognitive, and communicative models. These works were treated as background references rather than as part of the chronological screening.

Second, the literature from 2000 to 2026 was screened year by year, moving backwards from 2026 to 2000. For each year, the Google Scholar results were manually inspected according to relevance ranking. Approximately the first fifty records were examined for each year. Each record was manually assessed on the basis of title, abstract, venue, and, when necessary, full text.

Only journal articles and conference papers were retained. Book chapters, edited-volume contributions, manuals, and handbook-style texts were generally excluded. This exclusion was not based on their lack of relevance, since several book chapters have played an important role in terminology studies, but on the focus of the review on primary research contributions published in journals and conference proceedings. Exceptions were considered only when a work had clear historical or theoretical relevance for the development of the field.

The inclusion criteria were the following: (i) the publication explicitly discusses terminological definitions, definition writing, definition modeling, or definition extraction and generation in relation to specialized concepts; (ii) the contribution is situated within terminology, specialized lexicography, knowledge representation, computational terminology, natural language processing, or closely related fields; and (iii) the publication proposes, evaluates, compares, or critically discusses methods, models, principles, or applications concerning definitions of terms.

[1] https://detech2026.dei.unipd.it/

Publications were excluded when definitions were mentioned only incidentally, when the focus was exclusively on general lexicographic definitions without a terminological dimension, or when the work addressed terminology only as a secondary component of a broader unrelated task. Publications were also excluded when they did not provide sufficient methodological, theoretical, or empirical contribution to the study of terminological definitions.

The final set of papers was organized thematically rather than strictly chronologically. This organization makes it possible to identify continuities and shifts across the literature, including the movement from normative principles for definition writing, to corpus-based and cognitive approaches, to ontological and knowledge-based modeling, and finally to computational methods for automatic definition extraction and generation. The resulting review therefore emphasizes how terminological definitions have progressively moved from being treated primarily as manually written entries in terminological resources to becoming objects of computational modeling, extraction, generation, validation, and reuse.

A limitation of this procedure is that Google Scholar rankings are dynamic and depend on factors that are not fully transparent. For this reason, the review should be understood as a structured and transparent survey of the literature rather than as a fully reproducible systematic review. To mitigate this limitation, the screening was conducted systematically across publication years, using the same query, the same relevance-based inspection depth, and the same inclusion and exclusion criteria throughout the process.

3. Before the 2000s: Theoretical Foundations of Terminological Definitions

The pre-2000 literature on terminological definitions is characterized by a strong concern with the conceptual status, internal structure, and linguistic formulation of definitions in terminological resources. In this period, the definition is primarily understood as a device for delimiting a specialized concept within a conceptual system, rather than as a paraphrase of a lexical item. This distinction is crucial because it separates terminological definition from both lexicographic definition and encyclopedic description. While lexicography proceeds from words to meanings, terminography is traditionally described as onomasiological: it starts from concepts and then associates them with terms. From this perspective, a terminological definition is expected to identify the relevant concept, locate it within a domain, and distinguish it from neighbouring concepts [2, 3].

A first major line of work concerns the inheritance of the classical analytic model of definition. The genus–differentia structure remains the dominant reference model: the concept to be defined is related to a superordinate concept and then distinguished from co-hyponyms through one or more specific characteristics. Ndi-Kimbi's analysis of definitions in BS/ISO 2382 shows that terminological definitions continue to rely heavily on traditional rules derived from logic and philosophy, including requirements of equivalence between definiendum and definiens, avoidance of circularity, preference for positive formulation, and clarity of expression [4]. At the same time, this work also shows that actual terminological definitions frequently deviate from such rules, revealing a gap between normative principles and definitional practice.

Sager and L'Homme develop this issue by proposing a more controlled model for analytical definitions in terminological databases [5]. Their starting point is that conventional free-text definitions are poorly suited to computational storage, retrieval, and reuse. Although many technical definitions share recurrent underlying structures, their surface realization is often too variable to support efficient processing. Their proposal therefore aims to regularize the analytical definition by making explicit the categories of information that constitute the definiens. This is one of the clearest pre-2000 attempts to connect terminological definition writing with database design and computational control. The definition is no longer treated only as a textual field in a term record, but as a structured representation whose internal components can support authoring, validation, retrieval, and knowledge extraction.

A closely related contribution is the distinction between the conceptual structure of definitions and their linguistic realization. Sager and Ndi-Kimbi argue that many difficulties in definition writing

arise because guidelines often confuse conceptual organization with surface linguistic form [6]. Their analysis separates the conceptual relations involved in definition from the linguistic means used to express them. The analytical definition is thus interpreted as the linguistic realization of a conceptual operation: locating a concept through a superordinate category and specifying the characteristics that delimit it within the relevant conceptual field. This distinction is important for later computational approaches because it anticipates the need to separate definitional content from definitional wording.

The linguistic realization of definitions is further investigated by Ndi-Kimbi through the role of verbs in terminological definitions [7]. Whereas the genus in analytical definitions is generally realized by nominal expressions, verbs play a central role in expressing relations, states, functions, activities, and processes associated with the defined concept. This work broadens the analysis of definitions beyond noun phrase structures and shows that definitional meaning is also organized through predicative patterns. For computational terminology, this observation is relevant because verbs and verbal constructions often encode the semantic relations that connect the definiendum to other concepts in the domain.

Another important strand of the pre-2000 literature connects terminological definitions with knowledge representation and concept-oriented systems. Eck and Meyer explicitly frame definition construction as part of concept analysis in a terminological knowledge base [8]. Their work brings the Aristotelian tradition into a computational environment by treating intensional terminological definitions as products of a controlled procedure supported by a knowledge base. The objective is not merely to write definitions, but to guide the terminologist in constructing definitions that are consistent in content and form and that clearly indicate the place of concepts within a concept system. This approach anticipates later concerns with definition authoring tools, terminological knowledge bases, and semi-automatic support for terminography.

Heinsohn's work on terminological logics is not directly concerned with terminological definition writing in the terminographic sense, but it is relevant to the broader pre-2000 context because it shows how terminological knowledge representation was already being connected to formal inference and uncertainty modeling [9]. In terminological logics, definitions organize concepts in hierarchies and support reasoning over concept descriptions. By introducing probabilistic extensions, Heinsohn addresses the limits of purely categorical definitions in real-world domains, where concept dependencies may be uncertain or non-definitional. This line of work points to a broader issue that remains relevant for contemporary computational terminology: the relation between strict definitional knowledge, domain regularities, and uncertain or defeasible conceptual relations.

Taken together, these works define the main theoretical coordinates of terminological definitions before the 2000s. First, terminological definitions are concept-oriented: they delimit specialized concepts rather than simply explain words. Second, their preferred model is analytical and intensional, based on a superordinate concept and distinguishing characteristics. Third, their quality depends on the relation between conceptual structure and linguistic realization. Fourth, their increasing integration into term banks, terminological databases, and knowledge bases already raises questions of formalization, standardization, computational control, and reuse. These issues provide the background for the developments after 2000, when terminological definitions become increasingly connected to corpora, knowledge organization, ontology engineering, and automatic processing.

4. From 2000 to 2015: Corpus-Based, Ontological, and Knowledge-Oriented Approaches

The period between 2000 and 2015 can be understood as a phase of consolidation and diversification in the study of terminological definitions. The pre-2000 literature had already established the definition as a concept-oriented device, normally grounded in intensional and analytical principles. In the following fifteen years, this theoretical basis was not abandoned, but it was progressively reinterpreted in relation to new problems: the status of concepts in different terminological traditions, the tension between conceptual stability and contextual variation, the integration of definitions into ontologies and

terminological knowledge bases, and the emergence of computational methods for definition extraction and evaluation.

A first line of development concerns the clarification of basic theoretical categories. Leitchik and Shelov revisit the distinction between *notion* and *concept*, showing how different terminological traditions have used these categories in partially divergent ways [10]. Their discussion is relevant because terminological definitions depend on the status assigned to the unit being defined: if the object of definition is a specialized concept, then definition writing cannot be reduced to the explanation of word meaning. This issue is also central in ten Hacken's analysis of the tension between definition and reality [11]. Against approaches that reject classical terminological definitions on the basis of prototype theory or sociocognitive terminology, ten Hacken argues that classical definitions remain necessary for certain types of terms, especially in domains where clear-cut classification is required, such as scientific and legal communication. The point is not that all specialized vocabulary can be defined by necessary and sufficient conditions, but that some terms require such definitions precisely because they function as instruments of classification.

The intensional definition remains the main reference model in this period, but it is increasingly treated as a practical and operational tool rather than as a purely normative ideal. Löckinger, Kockaert, and Budin provide a systematic account of intensional definitions from the perspective of terminology science and terminology management [12]. Their contribution is particularly important because it connects theoretical principles with definition writing and assessment practices. Intensional definitions are presented as a way to represent concepts and concept systems consistently in natural language, while also opening the possibility of semi-automatic exploitation, for instance by deriving concept models from definitions or generating definitions from concept models. This explicitly anticipates the later computational turn, where definitions become not only terminographic products but also structured data for automatic processing.

A second line of development concerns the evaluation and modeling of definitional content. Seppälä's work is central in this respect. In her 2009 proposal, she argues that the evaluation of definition extraction cannot be limited to detecting whether a text span is definition-like; it must also assess whether the extracted information is relevant for a terminographic definition of a given concept in a given expert domain [13]. This shifts attention from the surface identification of definitional contexts to the problem of feature relevance. Not all information about a concept is definitionally relevant, and relevance depends on the domain, the type of concept, and the purpose of the definition. This line of argument is further developed in Seppälä's ontological framework for modeling the contents of definitions [14]. Here, the goal is to provide a domain- and language-independent framework for analyzing and constraining definition contents, with the broader aim of supporting definition-authoring tools and improving the integration of textual definitions into ontologies.

The same period also sees a stronger connection between terminological definitions and ontological modeling. Ceusters and Smith propose a unified framework for biomedical terminologies and ontologies, motivated by the need for interoperable, non-redundant annotation of biomedical data and scientific texts [15]. Their work is not primarily about definition writing, but it is highly relevant for the present review because it shows how definitions, terms, and concepts become part of a broader problem of ontological realism, semantic interoperability, and reliable mapping across resources. In this perspective, definitions are no longer only explanatory texts for human users; they contribute to the precise identification of entities, categories, and relations in computationally usable knowledge structures.

Frame-based terminology and multidimensional categorization provide another important development. Faber, Mairal, and Magaña describe the linking of EcoLexicon, a domain-specific environmental resource, to a general ontology [16]. This work shows how specialized knowledge can be represented as an extension of general knowledge, combining multilingual terminological information with conceptual and relational structures. Araúz and San Martín further develop this perspective by focusing on multidimensional categorization in terminological definitions [17]. In EcoLexicon, environmental concepts may be categorized from different perspectives, which can lead to multiple inheritance and information overload. Their solution consists in recontextualizing conceptual networks according to discipline-based domains and applying the same contextual constraints to definitional propositions.

The resulting idea of flexible terminological definitions is significant because it weakens the assumption that one concept must always be represented by one stable definition independently of context, while preserving the concept-oriented basis of terminological work.

A related concern appears in studies dealing with terminology management, specialized neology, and domain-specific resources. Roldán-Vendrell and Fernández-Domínguez analyze emergent neologisms and lexical gaps in specialized languages, showing how new terms may arise not only to name new realities but also to fill gaps within existing conceptual and denominative systems [18]. Their work is relevant for terminological definitions because the interpretation of neologisms depends on their position in a conceptual field and on the definitions that make their distinctive value explicit. Cenac discusses Romanian technico-scientific terminology with attention to the term–concept relation, intension and extension, and the strategies used for defining scientific terms and concepts [19]. Ciobanu, in turn, emphasizes the role of terminology centres in managing computer terminology, where systematic organization and definition of concepts are presented as necessary for clarity, precision, consistency, and professional communication [20]. These contributions show that, alongside theoretical and ontological developments, terminological definition remains a practical instrument for terminology management in dynamic specialized domains.

The relation between terminological and lexical resources is also reconsidered. Dobrowolska and Szpakowicz examine the treatment of terminology in WordNet and plWordNet, focusing on the boundary between lexical and terminological information [21]. Their analysis is relevant because wordnets are organized primarily through semantic relations, whereas terminological resources require attention to specialized concepts and their domain-specific definitions. The problem is therefore not only how to add terminological units to lexical resources, but how to represent the difference between general-language meanings and terminological concepts. This issue becomes increasingly important for computational terminology, where lexical databases, ontologies, and terminological resources are often combined.

A further strand of work links terminological definitions to glossary construction and domain normalization. Suárez-Figueroa, Aguado de Cea, and Gómez-Pérez present a methodology for creating a glossary in ontology engineering, motivated by the lack of an explicitly agreed terminology for processes and activities involved in ontology development [22]. Their work illustrates the role of definitions in stabilizing terminology within an emerging technical field. The glossary is not simply a list of terms, but a resource built through methodological steps, expert consensus, and explicit relations such as subtype, composition, and synonymy. In this sense, definitions contribute to the normalization of a domain whose terminology is still unstable.

Finally, this period marks the emergence of computational work on definitions. Saggion and Gaizauskas propose a method for mining online sources for definition knowledge [23]. Their approach addresses the difficulty of finding definition-bearing passages in large text collections, especially when the definiendum alone is insufficient to retrieve relevant contexts. By gathering secondary terms from external sources such as encyclopedias, lexical databases, and the Web, their system improves the retrieval and extraction of definition strings. This work represents an early and important step toward automatic definition extraction. In comparison with later neural and generative approaches, it remains strongly pattern- and retrieval-oriented, but it already frames definitions as extractable knowledge objects that can support automatic glossary construction and question answering.

Overall, the years 2000–2015 show a transition from classical definition principles to a more diversified ecology of definition-related research. The analytical and intensional model remains central, but it is increasingly complemented by feature-relevance evaluation, ontological analysis, frame-based and multidimensional representation, terminology management, glossary construction, and early definition extraction. This period therefore prepares the later GenAI transition: before definitions could be generated by neural or large language models, they first had to be reconceptualized as structured, evaluable, context-sensitive, and computationally exploitable units of terminological knowledge.

5. From 2016 to 2026: Flexible Definitions, Interoperability, and the GenAI Transition

The decade from 2016 to 2026 marks a further shift in the study of terminological definitions. While the previous period had already connected definitions to ontologies, knowledge bases, and definition extraction, this more recent phase foregrounds flexibility, contextual adequacy, interoperability, and computational generation. The classical intensional model remains an important reference point, but it is increasingly problematized from several directions. Definitions are no longer treated only as stable statements of necessary and sufficient characteristics, but as context-sensitive representations of conceptual content, shaped by thematic perspective, communicative purpose, user needs, ideological assumptions, and, more recently, the affordances and risks of large language models.

A first major development concerns the consolidation of frame-based and flexible approaches to terminological definition writing. Durán-Muñoz and colleagues focus on the production of frame-based definitions, continuing the line of work associated with EcoLexicon and Frame-Based Terminology [24]. In this perspective, definitions are not isolated textual descriptions, but access points to a wider conceptual frame in which domain entities, processes, participants, and relations are organized. This approach is important because it moves definition writing away from a purely taxonomic model and toward the representation of situated specialized knowledge. The definition is expected to help users understand a concept by activating the relevant domain frame, not merely by placing the concept under a genus.

This contextual orientation becomes explicit in San Martín's work on the Flexible Terminological Definition Approach. In the 2022 study on thematic variation, terminological definitions are defined as natural-language descriptions of conceptual content, but the selection of that content is made dependent on thematic perspective and user needs [25]. The critique of the classical model is twofold: first, concepts are often multidimensional and may therefore be categorized under different genera; second, it is frequently impossible or unhelpful to distinguish necessary and sufficient characteristics from encyclopedic or contextual knowledge. The proposed solution is not to abandon definition writing, but to replace the search for context-independent characteristics with a corpus-based selection of relevant characteristics.

The same argument is further developed in San Martín's discussion of contextual constraints in terminological definitions [26]. Here, context is treated as indispensable for meaning construction, and terminological definitions are expected to account for the contextual factors that determine which part of a term's semantic potential is activated. The notions of *premeaning* and *precontext* are introduced to describe the interaction between the knowledge potentially associated with a term and the constraints that guide its interpretation. This perspective provides a theoretical basis for context-sensitive definitions: the definition is not a universal representation of all possible knowledge associated with a term, but a controlled representation of the knowledge that is relevant under specified contextual conditions.

The 2025 work on contextonym analysis operationalizes this contextual view through corpus methods [27]. Contextonyms are the terms that typically co-occur with the term to be defined, independently of a specific syntactic or semantic relation. By identifying the optimal configuration for extracting contextonyms in Sketch Engine, this work proposes a practical method for supporting the terminologist in selecting relevant definitional information. The study concludes that, for English, a 50-token window ranked by frequency is an effective configuration for extracting contextonyms useful for definition writing. This contribution is significant because it connects the flexible approach to reproducible corpus-based procedures: contextual relevance becomes something that can be explored empirically rather than only judged intuitively.

Another line of research concerns the role of conceptual relations in drafting definitions. Carvalho, Costa, and Roche show how conceptual information can be organized in a template-like format to support the writing of natural-language definitions in the biomedical domain [28]. Their approach is grounded in a double dimension of terminology work, where the conceptual system provides the backbone for the definition-writing process. The position of the concept in the system, its characteristics,

and its relations to other concepts are used to structure the definition before it is expressed in natural language. This work maintains the importance of conceptual relations, but it also shows how conceptual modeling can become a practical support for the drafting of definitions.

The connection between definitions and semantic annotation is also developed by Brač and Anić, who compare conceptual information from terminological definitions with semantically annotated corpus sentences in the aviation domain [29]. Their work connects terminological databases with semantic frame resources and semantic role labeling. Definitions in a terminological resource are not treated as final textual products only, but as sources of structured conceptual information that can be compared with corpus-based evidence. This is relevant for computational terminology because it suggests that definitions can contribute to the design of semantic roles and frame structures in specialized knowledge resources.

The question of representation and interchange becomes increasingly important in the same period. Eckart de Castilho and colleagues compare different designs for linguistic annotation representation and interchange, with attention to the conceptual and technical differences between frameworks [30]. Although their focus is not terminological definition writing as such, their work is relevant to the broader computational environment in which definitions, annotations, and semantic resources must circulate. As terminological definitions become connected to NLP pipelines, corpora, semantic annotation, and knowledge graphs, the problem of interoperable representation becomes central. Definitions must be not only human-readable but also compatible with computational workflows and annotation infrastructures.

Kuznetsova's work on definition in the semantic structure of scientific text provides another relevant perspective [31]. The paper argues that models of scientific texts organized around textual definitions may represent content more adequately than models based only on terms and their relations. This view is important for the present review because it assigns definitions a structural role in scientific discourse. A definition does not merely explain a term after it has been extracted; it can function as a semantic unit through which the content of a scientific text is organized, interpreted, and processed automatically.

In parallel with these modeling and computational developments, recent work also expands the social, ideological, and inclusive dimensions of terminological definitions. Pei and Cheng analyze terminological variation and ideology in data protection laws, showing how defined legal terms vary across legal systems and historical contexts [32]. Their corpus-assisted approach reveals that terminological variation cannot be explained only as a linguistic phenomenon. It is also connected to discourse communities, institutional histories, and ideological stances. For terminological definitions, this means that the act of defining a legal term is not neutral: it may encode assumptions about rights, obligations, institutional authority, and social values.

Song and Ju apply sociocognitive terminology principles to the translation of traditional Chinese medicine classics [33]. Their four-dimensional model emphasizes linguistic economy, systematic relevance, conceptual retrospection, and knowledge dissemination. Although the focus is translation, the paper is relevant to terminological definitions because it treats specialized terms as historically, culturally, cognitively, and communicatively embedded units. Definitions and translations must therefore account for conceptual connotation, symbolic representation, diachronic stratification, and communicative context. This reinforces the broader movement away from abstract, context-free definition models toward situated and culturally aware terminological description.

The social dimension becomes especially explicit in Chiocchetti and Ralli's study of gender-sensitive language in terminological definitions of Italian legal concepts [34]. Their central question is whether gender should be treated as a characteristic of concepts or as a property of objects. This has direct consequences for the rewriting of definitions in terminology databases, especially in grammatical-gender languages. The paper shows that gender-sensitive language is not merely a question of choosing inclusive terms; it also affects the conceptual, linguistic, and comparative levels of terminological description. In this sense, gender becomes a dimension of concept representation and therefore a matter for terminological definition writing.

The final part of the decade is marked by the explicit entrance of large language models into definition-oriented terminology work. Di Nunzio, Gallina, and Vezzani investigate the use of ChatGPT for extracting difficult terms and writing terminological definitions in the context of SimpleText 2024

[35, 36, 37]. Their work is positioned at the intersection of automatic term extraction, text simplification, and terminological definition generation. The proposed methodology relies on iterative prompting, evaluation of extracted terms and definitions, and manual correction. This is representative of the GenAI transition: LLMs can produce fluent explanatory definitions, but their output still requires terminological control, prompt design, expert evaluation, and post-editing.

This GenAI transition does not replace earlier concerns; rather, it makes them more urgent. If LLMs are used to generate terminological definitions, then the problems identified throughout the literature become evaluation criteria: conceptual adequacy, relevance of selected characteristics, contextual fit, consistency with domain knowledge, sensitivity to variation, interoperability with terminological resources, and transparency of the drafting process. The recent literature therefore suggests that automatic definition generation should not be evaluated only in terms of fluency or readability. It should also be assessed against the theoretical and methodological requirements developed in terminology studies.

Overall, the period from 2016 to 2026 shows that terminological definitions have become increasingly dynamic objects. They are conceptual descriptions, but also corpus-informed representations, frame-based access points, semantic annotation resources, instruments of legal and ideological positioning, sites of inclusive language policy, and outputs of human–LLM interaction. The decade therefore completes the transition from definition as a relatively stable terminographic field to definition as a computationally and socially situated knowledge object. This provides the immediate background for current shared tasks on definition generation, where the central question is not only whether a system can generate a plausible definition, but whether it can generate a definition that is terminologically, conceptually, contextually, and socially adequate.

6. Conclusion and Future Directions: Toward Terminology-Augmented Generation

This review has shown that terminological definitions have progressively become structured, evaluable, and reusable knowledge objects. The historical trajectory reconstructed in the previous sections suggests that definition writing is no longer only a terminographic activity aimed at human consultation. It is increasingly connected to knowledge representation, interoperability, text simplification, terminology extraction, translation, and machine-assisted generation.

A promising direction for future work is represented by Terminology-Augmented Generation (TAG), a paradigm that aims to integrate curated terminology resources into generative AI workflows [38]. TAG can be understood as a terminology-oriented counterpart to Retrieval-Augmented Generation (RAG), but with a different epistemic and technical basis. Whereas RAG typically retrieves unstructured or semi-structured text chunks on the basis of semantic similarity, TAG relies on structured terminology resources, such as termbanks, ontologies, glossaries, multilingual concept systems, and terminological databases. This distinction is crucial for definition generation, because terminological adequacy cannot be guaranteed by fluent generation alone. A generated definition must be aligned with a concept system, use the correct term, respect domain-specific relations, and remain consistent with expert-curated terminological knowledge.

From this perspective, automatic definition generation should not be framed merely as a natural language generation task. It should be treated as a terminology-aware generation task in which the model is constrained or guided by concept-oriented resources. TAG offers a way to make explicit the information that should guide generation: concept identifiers, preferred and deprecated terms, multilingual equivalents, domain labels, definitions, usage notes, conceptual relations, and contextual constraints. This is particularly relevant for terminological definitions because the main risk of LLM-based generation is not only hallucination in the general sense, but also terminological drift: the production of definitions that sound plausible but misrepresent the concept, blur distinctions between neighboring concepts, or introduce uncontrolled variants.

The evaluation of TAG is therefore a central issue. Lackner, Vega-Wilson, and Lang show that

terminology-augmented generation can improve terminology adherence in the machine translation use case, especially when LLMs are supplied with structured terminological information at run time [39]. Their results also show that the format in which terminology is provided to the model matters. Standard formats such as TBX remain important for terminology exchange and interoperability, but LLM-oriented formats such as Markdown, JSON, or YAML may be more effective as prompt-level representations. This suggests that future work on definition generation should consider not only which terminological data are used, but also how those data are serialized, filtered, and presented to the generative model.

TAG also provides a way to reconnect definition generation with earlier theoretical concerns in terminology studies. The literature reviewed in this paper repeatedly shows that terminological definitions depend on conceptual relations, feature relevance, domain context, user needs, and linguistic realization. A TAG-based system for definition generation should therefore be evaluated according to criteria that reflect this heritage: conceptual correctness, terminological precision, relevance of selected characteristics, consistency with the concept system, adequacy for the target user, and traceability of the sources used during generation. Fluency, readability, and grammaticality remain necessary, but they are not sufficient criteria for evaluating terminological definitions.

Recent work on generative AI and semantic analysis for digital lexicography and terminology extraction further shows that LLMs can support the identification of frames, concepts, and terminological units from user-generated discourse [40]. This is relevant because future terminological definition work will probably not be limited to highly controlled expert corpora. Social media, online communities, institutional texts, and specialized communication environments can all provide evidence of how concepts are framed, evaluated, and linguistically realized. However, such evidence must be transformed into structured terminological resources before it can reliably support generation. TAG can therefore be seen as a bridge between corpus evidence, semantic frames, terminology resources, and controlled generation.

The main implication is that the future of terminological definitions should not be reduced to fully automatic generation. Rather, it should be oriented toward hybrid workflows in which LLMs assist terminologists, translators, domain experts, and lexicographers under explicit terminological constraints. In such workflows, LLMs may propose draft definitions, identify candidate terms, suggest conceptual relations, or adapt definitions to different user groups. However, terminological resources and expert validation remain essential for controlling the output. TAG is promising precisely because it does not replace terminology work; it makes terminology work operational within generative AI systems.

Acknowledgments

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 is 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.

Declaration on Generative AI

During the preparation of this work, the author(s) used Chat-GPT-5 in order to: Grammar and spelling check. After using these tools, the author(s) reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] F. Vezzani, G. M. Di Nunzio, V. Bonato, G. Silvello, Overview of the international workshop on definition and term extraction challenge (detech) 2026, in: Proceedings of the International

Workshop on Definition and Term Extraction Challenge (DETECH) 2026, CEUR.org, Zadar, Croatia, 2026.

[2] B. de Bessé, 1.3.1 terminological definitions, in: S. E. Wright, G. Budin (Eds.), Handbook of terminology management, John Benjamins Publishing Company, 1997, pp. 63–74. URL: https://doi.org/10.1075/z.htm1.09bes. doi:doi:10.1075/z.htm1.09bes, tex.booktitle+duplicate-1: Volume 1: Basic Aspects of Terminology Management.

[3] F. W. Riggs, Onomantics and Terminology. Pt. 1: Their contributions to Knowledge Organization, Knowledge Organization 23 (1996) 25–37. URL: https://www.imrpress.com/journal/ko/23/1/10.5771/0943-7444-1996-1-25. doi:10.5771/0943-7444-1996-1-25.

[4] A. Ndi-Kimbi, Guidelines for terminological definitions: The adherence to and deviation from existing rules in BS/ISO 2382: Data Processing and Information Technology Vocabulary, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 1 (1994) 327–350. URL: https://www.jbe-platform.com/content/journals/10.1075/term.1.2.07ndi. doi:10.1075/term.1.2.07ndi.

[5] J. C. Sager, M.-C. L'Homme, A model for the definition of concepts: Rules for analytical definitions in terminological databases, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 1 (1994) 351–373. URL: https://www.jbe-platform.com/content/journals/10.1075/term.1.2.08sag. doi:10.1075/term.1.2.08sag.

[6] J. C. Sager, A. Ndi-Kimbi, The conceptual structure of terminological definitions and their linguistic realisations: A report on research in progress, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 2 (1995) 61–85. URL: https://www.jbe-platform.com/content/journals/10.1075/term.2.1.04sag. doi:10.1075/term.2.1.04sag.

[7] A. Ndi-Kimbi, The functional role of verbs in terminological definitions, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 2 (1995) 87–105. URL: https://www.jbe-platform.com/content/journals/10.1075/term.2.1.05ndi. doi:10.1075/term.2.1.05ndi.

[8] K. E. Eck, I. E. Meyer, Bringing Aristotle into the 20Th Century: Computer-Assisted Definition Construction in a Terminological Knowledge Base, in: Standardizing and Harmonizing Terminology: Theory and Practice, ASTM International, 1995, pp. 83–101. URL: https://store.astm.org/stp13748s.html. doi:10.1520/STP13748S.

[9] J. Heinsohn, A hybrid approach for modeling uncertainty in terminological logics, in: R. Kruse, P. Siegel (Eds.), Symbolic and Quantitative Approaches to Uncertainty, Springer, Berlin, Heidelberg, 1991, pp. 198–205. doi:10.1007/3-540-54659-6_89.

[10] V. M. Leitchik, S. D. Shelov, SOME BASIC CONCEPTS OF TERMINOLOGY: TRADITIONS AND INNOVATIONS, Terminology Science & Research / Terminologie : Science et Recherche 14 (2003) 86–101. URL: https://journal-eaft-aet.net/index.php/tsr/article/view/5803.

[11] P. Ten Hacken, The tension between definition and reality in terminology, in: Proceedings of the xiv euralex international congress, 2010, pp. 915–927.

[12] G. Löckinger, H. J. Kockaert, G. Budin, Intensional definitions, Handbook of terminology 1 (2015) 60–81.

[13] S. Seppälä, A Proposal for a Framework to Evaluate Feature Relevance for Terminographic Definitions, in: G. Sierra, M. Pozzi, J.-M. Torres (Eds.), Proceedings of the 1st Workshop on Definition Extraction, Association for Computational Linguistics, Borovets, Bulgaria, 2009, pp. 47–53. URL: https://aclanthology.org/W09-4408/.

[14] S. Seppälä, An ontological framework for modeling the contents of definitions, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 21 (2015) 23–50. URL: https://www.jbe-platform.com/content/journals/10.1075/term.21.1.02sep. doi:10.1075/term.21.1.02sep.

[15] W. Ceusters, B. Smith, A Unified Framework for Biomedical Terminologies and Ontologies, Studies in health technology and informatics 160 (2010) 1050–1054. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC3104298/.

[16] P. Faber, R. Mairal, P. J. M. Redondo, Linking a domain-specific ontology to a general ontology,

in: R. C. Murray, P. M. McCarthy (Eds.), Proceedings of the twenty-fourth international florida artificial intelligence research society conference, may 18-20, 2011, palm beach, florida, USA, AAAI Press, 2011. URL: http://aaai.org/ocs/index.php/FLAIRS/FLAIRS11/paper/view/2523, tex.bibsource: dblp computer science bibliography, https://dblp.org tex.timestamp: Wed, 26 Oct 2022 08:35:20 +0200.

[17] P. L. Araúz, A. S. Martín, Multidimensional categorization in terminological definitions, in: R. V. Fjeld, J. M. Torjusen (Eds.), Proceedings of the 15th EURALEX international congress, Department of Linguistics and Scandinavian Studies, University of Oslo, Oslo,Norway, 2012, pp. 578–584.

[18] M. R. Vendrell, J. F. Domínguez, Emergent neologisms and lexical gaps in specialised languages, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 18 (2012) 9–26. URL: https://www.jbe-platform.com/content/journals/10.1075/term.18.1.02rol. doi:10.1075/term.18.1.02rol.

[19] O. Cenac, On Romanian Technico-Scientific Terminology, Procedia - Social and Behavioral Sciences 63 (2012) 85–92. URL: https://www.sciencedirect.com/science/article/pii/S1877042812047489. doi:10.1016/j.sbspro.2012.10.015.

[20] G. Ciobanu, Contribution of a Terminology Centre in Managing Computer Terminology, Procedia - Social and Behavioral Sciences 46 (2012) 4678–4681. URL: https://www.sciencedirect.com/science/article/pii/S1877042812020538. doi:10.1016/j.sbspro.2012.06.317.

[21] M. Dobrowolska, S. Szpakowicz, Terminology in WordNet and in plWordNet, in: H. Orav, C. Fellbaum, P. Vossen (Eds.), Proceedings of the Seventh Global Wordnet Conference, University of Tartu Press, Tartu, Estonia, 2014, pp. 299–303. URL: https://aclanthology.org/W14-0141/. doi:10.18653/v1/W14-0141.

[22] M. C. Suárez-Figueroa, G. A. d. Cea, A. Gómez-Pérez, Lights and shadows in creating a glossary about ontology engineering, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 19 (2013) 202–236. URL: https://www.jbe-platform.com/content/journals/10.1075/term.19.2.03sua. doi:10.1075/term.19.2.03sua.

[23] H. Saggion, R. J. Gaizauskas, Mining on-line sources for definition knowledge, in: V. Barr, Z. Markov (Eds.), Proceedings of the seventeenth international florida artificial intelligence research society conference, miami beach, florida, USA, AAAI Press, 2004, pp. 61–66. URL: http://www.aaai.org/Library/FLAIRS/2004/flairs04-015.php, tex.bibsource: dblp computer science bibliography, https://dblp.org tex.timestamp: Wed, 26 Oct 2022 08:35:30 +0200.

[24] I. Durán-Munoz, Producing frame-based definitions: A case study, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 22 (2016) 223–249. URL: https://www.jbe-platform.com/content/journals/10.1075/term.22.2.04mun. doi:10.1075/term.22.2.04mun.

[25] A. San Martín, A Flexible Approach to Terminological Definitions: Representing Thematic Variation, International Journal of Lexicography 35 (2022) 53–74. URL: https://doi.org/10.1093/ijl/ecab013. doi:10.1093/ijl/ecab013.

[26] S. Martín, Antonio, Contextual Constraints in Terminological Definitions, Frontiers in Communication 7 (2022). URL: https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2022.885283/full. doi:10.3389/fcomm.2022.885283.

[27] A. San Martín, Optimizing Contextonym Analysis for Terminological Definition Writing, Information 16 (2025) 257. URL: https://www.mdpi.com/2078-2489/16/4/257. doi:10.3390/info16040257.

[28] S. Carvalho, R. Costa, C. Roche, The role of conceptual relations in the drafting of natural language definitions: an example from the biomedical domain, in: Proceedings of the LREC 2018 workshop "globalex 2018 – lexicography & WordNets, Proceedings of, I. Kernerman & S. Krek (eds.)„ Miyazaki, Japan, 2018. URL: https://hal.science/hal-01829913, tex.hal_id: hal-01829913 tex.hal_version: v1.

[29] I. Brač, A. O. Anić, From concept definitions to semantic role labeling in specialized knowledge resources, in: Proceedings of the 13th international conference of the asian association for lexicography. Eds. Gürlek, mehmet; naim çiçekler, ahmet; taşdemir, yasin. Asos publisher. Istanbul., 2019.

[30] R. Eckart de Castilho, N. Ide, E. Lapponi, S. Oepen, K. Suderman, E. Velldal, M. Verhagen, Representation and Interchange of Linguistic Annotation. An In-Depth, Side-by-Side Comparison of Three Designs, in: N. Schneider, N. Xue (Eds.), Proceedings of the 11th Linguistic Annotation Workshop, Association for Computational Linguistics, Valencia, Spain, 2017, pp. 67–75. URL: https://aclanthology.org/W17-0808/. doi:10.18653/v1/W17-0808.

[31] Y. M. Kuznetsova, Definition in the Semantic Structure of a Scientific Text, Scientific and Technical Information Processing 44 (2017) 309–313. URL: https://doi.org/10.3103/S0147688217050070. doi:10.3103/S0147688217050070.

[32] J. Pei, L. Cheng, Mapping terminological variation and ideology in data protection laws, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 26 (2020) 159–183. URL: https://www.jbe-platform.com/content/journals/10.1075/term.00049.pei. doi:10.1075/term.00049.pei.

[33] Y. Song, B. Ju, Sociocognitive terminology-based principles for translating the classics of traditional Chinese medicine, Social Sciences & Humanities Open 8 (2023) 100616. URL: https://www.sciencedirect.com/science/article/pii/S2590291123002218. doi:10.1016/j.ssaho.2023.100616.

[34] E. Chiocchetti, N. Ralli, Gender-sensitive language in terminological definitions: A case study on Italian legal concepts, Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 32 (2026) 119–147. URL: https://www.jbe-platform.com/content/journals/10.1075/term.00093.chi. doi:10.1075/term.00093.chi.

[35] G. M. Di Nunzio, E. Gallina, F. Vezzani, UNIPD@SimpleText2024: A Semi-Manual Approach on Prompting ChatGPT for Extracting Terms and Write Terminological Definitions, in: G. Faggioli, N. Ferro, P. Galuščáková, A. G. S. d. Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), volume 3740 of *CEUR Workshop Proceedings*, CEUR, Grenoble, France, 2024, pp. 3230–3237. URL: https://ceur-ws.org/Vol-3740/#paper-313.

[36] G. M. Di Nunzio, F. Vezzani, V. Bonato, H. Azarbonyad, J. Kamps, L. Ermakova, Overview of the CLEF 2024 SimpleText Task 2: Identify and Explain Difficult Concepts, in: G. Faggioli, N. Ferro, P. Galuscáková, A. G. S. de Herrera (Eds.), Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2024), Grenoble, France, 9-12 September, 2024, CEUR Workshop Proceedings, CEUR-WS.org, 2024, pp. 3129–3146.

[37] L. Ermakova, E. SanJuan, S. Huet, H. Azarbonyad, G. M. Di Nunzio, F. Vezzani, J. D'Souza, J. Kamps, Overview of the CLEF 2024 SimpleText Track, in: L. Goeuriot, P. Mulhem, G. Quénot, D. Schwab, G. M. Di Nunzio, L. Soulier, P. Galuščáková, A. García Seco de Herrera, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer Nature Switzerland, Cham, 2024, pp. 283–307. doi:10.1007/978-3-031-71908-0_13.

[38] G. M. Di Nunzio, Terminology-Augmented Generation (TAG): Foundations, Use Cases, and Evaluation Paths, Journal of Digital Terminology and Lexicography 1 (2025) 97–104. doi:10.25430/pupj.jdtl.1752566034.

[39] A. Lackner, A. Vega-Wilson, C. Lang, An Evaluation of Terminology-Augmented Generation (TAG) and Various Terminology Formats for the Translation Use Case, Journal of Digital Terminology and Lexicography 1 (2025) 31–47.

[40] M. Panagiotou, K. Gkatzionis, E. Kaloudis, Framing Food Online Discourse: Employing Generative AI and Semantic Analysis for Digital Lexicography and Terminology Extraction, Journal of Digital Terminology and Lexicography 1 (2025) 1–12.