‘Defined_by’ relation
as structuring principle of terminologies

Tomara Gotkova†, Francesca Ingrosso‡, Polina Mikhel† and Alain Polguère‡,*†

1Université de Lorraine, CNRS, ATILF (Nancy, France)
2Université de Lorraine, CNRS, LPCT (Nancy, France)

Abstract
We present an innovative strategy for the study and modelling of core scientific terminologies. Our approach is non-taxonomic and relies on semantic embedding of Terms in lexicographic definitions – ‘defined_by’ rather than ‘is_a’ relations – to identify the underlying terminological systems of sciences. It is demonstrated on the core terminology of (general) chemistry as modelled in Lexical Systems where terminologies are fully embedded within a model of the corresponding general language.

Keywords
core terminology of chemistry, Explanatory Combinatorial Lexicology, Lexical System, ‘defined_by’ relation, non-taxonomic approach

1. Towards non-taxonomic structuring of terminologies

1.1. Issues with lexical taxonomies

The mainstream approach to the structuring of terminological models is undoubtedly one based on formal (computerized) ontologies [1]. Such models are predominantly organized as ‘is_a’ class hierarchies [2]. Ontology-based terminological models are Concept-driven, rather than Term-driven: they graft the modelling of Terms onto a taxonomy of domain Concepts. Underlying such models lies the Linnaean principle of taxonomic organization of the World and the credo that taxonomies are the most efficient information structures for organizing the various realms of “Things.” From the so-called ‘Tree of Porphyry’ [3] inspired by Aristotle’s defining principles to modern formal ontologies, via Linnaeus’ classification of species [4], the taxonomic approach has established itself as the frame of reference for putting into order the natural messiness of Things.

It is worth noting that the taxonomic approach often finds its limits once applied and confronted to the complexity of natural Things as it tries to project onto them a unique classifying principle that is not devoid of philosophical and ideological prejudices [5]. Additionally, one of the drawbacks of this approach is that it favors relations between hypothesized classes of

2nd International Conference on ‘Multilingual digital terminology today. Design, representation formats and management systems’ (MDTT) 2023, June 29–30, 2023, Lisbon, Portugal
*Corresponding author.
†These authors contributed equally.

© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
Things over relations between actual Things themselves [6]. In contrast, other relational models are increasingly being constructed and applied: ones that represent given domains as huge sets of Things individually connected by multiple types of relations forming networks with specific topological properties known as small-world networks [7]. We adhere to the hypothesis that such structures are better suited than taxonomies to model all the complexity of the relational systems that connect natural Things – see the notion of terrain network coined by B. Gaume [8]. For lack of space, we will not attempt to justify further this hypothesis at the general level, but rather focus on natural language lexicons, and terminologies in particular.

The taxonomic approach has for a long time been applied to lexical units of natural languages through the recourse to hyperonymic / superordinate relations [9, sec. 1.5.2]. In contemporary lexical research, the most well-known instance of the approach is, for the English language, the Princeton WordNet [10, 11] with its multiple offsprings for other natural languages (e.g., [12]).

In our opinion, the main problems posed by the taxonomic / hyperonymic structuring of lexical units in general, and Terms in particular, are the following.

1. Hyperonymy applies mainly to nominal lexical units – cf. the separate structuring principles used in WordNet for Nouns vs. Verbs, Adjectives and Adverbs [13]; however, Terms are of other main parts of speech as well, such as verbs and adjectives [14].
2. As for all types of taxonomic models, it favors relations between postulated lexical classes over relations between individual lexical units (see earlier remark on taxonomies).
3. The ‘is_a’ (or ‘kind_of’) relation accounts only for a small subset of all relations holding between Terms, and between the corresponding domain notions. Terminological taxonomies may therefore be insufficient from a pedagogical point of view, where the acquisition of scientific and technical notions is at stake.
4. Finally, the structuring of terminologies into specific domain taxonomies leads to an isolationist approach where terminologies are modelled in close circuit systems whereas they interact with and are fully integrated to the system of the general language.

Terminological and writing conventions. In this paper, we take a Concept – with a capital C – to be any unit of reasoning, whether institutionalized (in the semiotic sense) or created on-the-fly in a given instance of inference performed by an individual. A Term – with a capital T – is a lexical unit that denotes an institutionalized Concept belonging to a conceptual system of a knowledge domain (or a set of knowledge domains). What is commonly called notion is a Term ↔ Concept association – names of notions are written in sans serif font. For instance, to fully master the chemical notion of element, one has (i) to master the corresponding scientific Concept (its role in chemistry, elements identified in the Periodic Table, etc.) and (ii) to master the linguistic properties of the English Term element (mainly) general language vocable element (see sections 2 and 3).

1.2. The Lexical System approach

We present below an alternative to taxonomic terminological models where ‘defined_by’ relations holding between Terms – rather than ‘is_a’ relation holding between Concepts [15] – is the
organizational principle of terminologies, motivated by pedagogical / acquisition considerations. In our approach, the ‘defined_by’ relation is one among many different types of relations holding between lexical units of a language, forming what is termed a Lexical System [16]. Let us briefly characterize this type of model.

Lexical Systems are designed to be compatible with fully relational models of the Mental Lexicon as postulated from both psycholinguistic and lexicological perspectives [17, 18, 19]. They are lexicographically built according to theoretical and descriptive principles of Explanatory Combinatorial Lexicology / Lexicography [20], except for the fact that they are lexical network models rather than “textual” dictionaries, such as [21, 22]. Formally, they are small-world networks (see 1.1). Their nodes are mainly lexical units of the language; their arcs are lexical relations of various types, listed below by order of importance to the small-world network structuring of Lexical Systems:

- paradigmatic (semantic) and syntagmatic (combinatorial) relations corresponding to so-called Meaning-Text lexical functions [23];
- copolysemy relations (extension, metonymy, metaphor, ...) holding between senses of polysemous vocables [24];
- ‘defined_by’ relations that connect each given lexical unit \( L_1 \) to other lexical units \( L_2, L_3 \ldots L_n \), whose meaning is embedded in the meaning of \( L_1 \) – see details in section 2 below;
- formal inclusion relations – e.g., the lexico-syntactic structure [25] of the idiom “push buttons” “irritate someone” is formally built with the lexemes push and button.

2. Nature and role of ‘defined_by’ relations

Though Meaning-Text paradigmatic and syntagmatic lexical functions are the leading principle for the structuring of Lexical Systems, ‘defined_by’ relations proved to be essential in the organization of terminologies, especially from a pedagogical perspective.¹

A ‘defined_by’ relation holds between two lexical units \( L_1 \) and \( L_2 \) if \( L_1 \) is lexicographically defined in terms of \( L_2 \) – i.e., if \( L_2 \) appears in the lexicographic definition of \( L_1 \). Such relation is noted \( L_1 \overset{\text{Def}}{\rightarrow} L_2 \).

Let us illustrate the notion of ‘defined_by’ relation with data in Table 1: the lexicographic definition of the chemistry Term \( \text{element III.3a} \) proposed in [27] – see section 3 below for more details.² Note that such lexicographic definition follows the format and structuring principles of Explanatory Combinatorial Lexicology detailed in [28].

In Table 1, terminological ‘defined_by’ relations are signalled by underlying the Terms that (i) participate in the lexicographic definition of \( \text{element III.3a} \) and (ii) are themselves accounted for in the Lexical System of English: i.e., \( \text{element III.3a} \overset{\text{Def}}{\rightarrow} \{ \text{atom I.2, proton, nucleus I.2} \} \). Two types of ‘defined_by’ relations need to be distinguished in natural language lexicons.

¹For instance, the set of lexicological notions introduced in the manual [26] is entirely structured according to this principle and it determines the flow of notions making up the teaching program implemented by the manual.
²The lexicographic numbering used in the paper has been established in [27] and takes into consideration the polysemy of the corresponding vocables both for general and specialised senses – see [27, Chap. 6].
Table 1
Lexicographic definition of element III.3a [27, Chap. 6]

<table>
<thead>
<tr>
<th>element III.3a X</th>
<th>type of atoms</th>
<th>that is identified by the number X corresponding to the quantity of protons in the nucleus of the atoms</th>
</tr>
</thead>
</table>

Ex.: How is an atom of the element 54 (Xe) likely to act during a chemical reaction?

Generic ‘defined_by’ relations. A generic \( L_1 \xrightarrow{\text{Def}} L_2 \) relation is such that \( L_2 \) is the central component \([28, \text{Sec. 2.4}]\) of \( L_1 \)’s definition. Such is the case of atom in Table 1, though the definition stipulates that chemical elements are not atoms per se, but types of atoms, and therefore belong to a more abstract level of conceptualization of chemical entities. Generic ‘defined_by’ relations between Terms are, by definition, closely related to ‘is_a’ taxonomic relations between Concepts.

Specific ‘defined_by’ relations. A specific \( L_1 \xrightarrow{\text{Def}} L_2 \) relation is such that \( L_2 \) belongs to a peripheral (= non-central) component of \( L_1 \)’s definition. Such is the case of proton and nucleus in the definition of element III.3a.

It is essential to note that generic and specific ‘defined_by’ relations are equally important from the point of view of notion acquisition. Because notions are Term ↔ Concept associations (cf. terminological remark in section 1.1), the definition in Table 1 tells us that the mastering / understanding of the notion of chemical element necessitates the mastering / understanding of not only the notion of atom, but also the notions of atom’s nucleus and of proton.

We advocate a descriptive approach where terminological models are organized by a system of (generic and specific) ‘defined_by’ connections that aims at accounting for a notion-building perspective on terminologies. Simultaneously, however, the terminographic description of each individual Term should be embedded into the Lexical System of the corresponding natural language – for a justification of such integrated approach, see \([29]\) and section 4 below.

We illustrate below our descriptive strategy with work done on the core terminology of chemistry \([27]\).

3. Core terminology of chemistry

In \([27]\), the core terminology of chemistry is characterized as constituted of Terms that (i) are taught in core courses in general chemistry and (ii) are “shared by most subdomains of chemistry” without belonging to a given subdomain \([27, \text{p. 14}]\). We can add that those terms determine the notional foundation of the discipline: i.e., notions – such as atom, (chemical) bond, molecule, etc. – from which the bulk of the notional system of general chemistry is derived. The study, performed in a multilingual perspective, led to the definition of over a hundred core chemistry Terms for each of the three languages considered: English, French and Russian.\(^3\)

\(^3\)For terminological gaps between these languages, see \([27, \text{sec. 7.1.2}]\).
As mentioned earlier, the theoretical and description foundation of the work is Explanatory Combinatorial Lexicology, the lexico-semantic component of Meaning-Text linguistics. In this respect, it relates to previous terminological work anchored in the same linguistic framework, see [30]. A distinctive feature, however, is the fact that the core terminology of chemistry has been modelled in the context of the lexicography of Lexical Systems (section 1.2) where Terms are integrated in the small-world network of the general language. The core of the description of each Term is, of course, its lexicographic definition, of which one example is given above (Table 1, section 2): the definition of element \( \text{III.3a} \). From the three ‘defined by’ relations embedded in this definition and elicited earlier – \( \text{III.3a} \stackrel{\text{Def}}{\rightarrow} \{ \text{ATOM 1.2}, \text{PROTON}, \text{NUCLEUS 1.2} \} \) – one can infer the bottom-up notion building organization shown in the right-hand side graph, where an \( N_1 \rightarrow N_2 \) link indicates that the acquisition of notion \( N_1 \) is required for the acquisition of notion \( N_2 \). Note that this graph takes only into consideration the acquisition of the notion of element and of the three other notions it is directly related to: atom, proton, and nucleus. These notions, in their turn, presuppose other notions, via the lexicographic definition of their corresponding Term. Proton, for instance, presupposes the additional notions of subatomic particle, interaction and charge, via the lexicographic definition of the Term proton, given in Table 2.

Table 2: Lexicographic definition of proton [27, Chap. 8]

| proton of X interacting with Y: subatomic particle \^ of the atom 1.2 X |
|---|---|
| \cdot \text{that interacts with the subatomic particle} Y of X |
| \cdot \text{that is positively charged 0} |

Ex.: In a hydrogen atom, a negative electron orbits a positive proton because of the electromagnetic, not the gravitational, force between the two particles.

At the level of the complete core terminology of chemistry, the whole set of Term definitions determine a hierarchical ‘defined by’ induced organization of corresponding core notions that function as a roadmap for the teaching / acquisition of chemistry, as a scientific discipline. An extract of this roadmap for English chemical notions is given in Appendix A.

In [27], core chemical terminologies have been modelled for each of the three languages considered. The respective numbers of Terms vs. notions are as follows:

- English: 107 defined core Terms \( \rightarrow \) 53 corresponding core notions;
- French: 103 defined core Terms \( \rightarrow \) 53 corresponding core notions;
- Russian: 102 defined core Terms \( \rightarrow \) 52 corresponding core notions.

The discrepancy between the number of core Terms and the number of core notions, for each language, is explained by the fact that a core Term \( T \) can have one or more semantic derivatives \( T', T'' \), ... with a unique corresponding core notion. For instance, all six core Terms ion, ionic, ionization, ionize\text{non-causative}, ionize\text{causative} and ionized are connected by semantic (and morphological) derivations, and they have one unique corresponding core chemical notion: ion.
4. In lieu of conclusion: carbon, the Great Escape from chemistry

We are in the process of implementing our core chemistry terminology – structured by the ‘defined_by’ relation – as an online resource for chemistry teachers and students. Hopefully, this will ultimately validate the pedagogical relevance of our approach.

To conclude, we use the case of the noun carbon to briefly justify the importance of adopting an integrated approach to terminological modelling, one where Terms cohabitate with general language lexical units within a unique Lexical System [29].

Carbon is omnipresent in today’s media and political discourse, daily conversation, even printed on goods’ labels and travel tickets. With the evidence of an ongoing environmental crisis, it has also become a buzzword with sometimes fluctuating and fuzzy semantic boundaries, as shown by a recent study of the use of carbon on social medias [31]. In addition, many Terms have been coined from carbon in relation to the Environment: (de-)carbonize, carbon(-iz-)ation, carbon footprint, carbon sequestration, etc. To sort out this plethoric presence of carbon in modern discourse one has to start with the English vocable carbon itself and its rich polysemy. We model it as follows (with approximate glosses rather than actual lexicographic definitions), applying the principles of relational polysemy presented in [24] and focusing on the first four senses, those that are most directly linked to the topic of the present paper.

I.1 spec4 ‘element with atomic number 6’
I.2 [Extension of I.1] quasi-spec ‘substance which is the materialization of carbon’ [Ex.: Carbon is a solid, with a color of blackish brownish resembling charcoal.]
II.1 [Metonymy of I.1] (spec) ‘type of gas containing carbon’ = CO₂ [Ex.: Coal-fired power plants, which produce the majority of Georgia’s electricity and emit the most carbon, would pay the most.]
II.2 [Metaphor of II.1] quasi-spec ‘symbolic polluting substance as if it were carbon’ [Ex.: Most people emit carbon every day simply by using a non-renewable resource, such as coal, natural gas, or oil.]
III.1 [Metonymic metaphor of II.1] = ”Carbon paper”
III.2 [Metonymy of III.1] = ”Carbon copy”

Though carbon is clearly a “terminological vocable” intimately associated to the field of chemistry via its basic lexical unit – cf. the spec(ialized) sense I.1 –, it contains both terminological and non-terminological senses. This polysemy is particularly tricky to handle from a terminology viewpoint due to the fact the “hardcore” terminological sense I.1 cohabitates in the vocable and frequently interacts in the Speaker’s mind with senses that possess a terminological flavor without being associated to a well-structured notional system: those identified by the usage note quasi-spec(ialized). To top it all, the vocable contains an optionally specialized sense II.1, marked as (spec): a “runaway” Term [32], i.e., a Term that fully belongs to an organized terminology but tends to be used equally in non-specialized discourse by Speakers who do not necessarily master the corresponding notion. Clearly, carbon literally escaped from the terminology of chemistry to develop into closely-related senses, this situation being potentially harmful for the proper acquisition and exploitation of corresponding notions. This illustrates well why it is necessary to have an integrated approach to the modelling of terminologies, one that takes into consideration the fact that terminologies are fully contained in natural language lexicons.

4See below for a discussion of usage notes spec, (spec) and quasi-spec.
Acknowledgments

This work was supported partly by the FRONTERME doctoral project, reference 18_GE4_022, of the Grand Est Region (France) and by the French PIA project “Lorraine Université d’Excellence,” reference ANR-15-IDEX-04-LUE. We wish to express our gratitude to our two MDTT 2023 reviewers for their rich and inspiring comments.

References


A. Notion building roadmap for core English chemical notions

Figure 1 below presents a small sample of the hierarchical system of core English chemical notions [27, p. 167] as a system of ‘defined_by’ relations connecting the corresponding terms. It has to be consulted from bottom to top: from the most “primary” notions to those directly or indirectly built on them via lexicographic definition of the corresponding Terms. The red color signals notions that correspond to chemical entities; amber signals notions that correspond to chemical facts.

Figure 1: Sample of the ‘defined_by’ hierarchy of core English chemical notions – from [27, p. 167]