Tailoring Terminological Resources to the Users' Needs: a Corpus-based Study on Appositive Constructions in Italian and English

Giulia Speranza, Maria Pia Di Buono and Johanna Monti

University of Naples "L'Orientale" - UNIOR NLP Research Group, Via Chiatamone, 61/62, 80121 Napoli NA

Abstract

Terminological resources are indispensable tools for accessing a specialized domain of knowledge. Several actors gravitate around a specialized domain and, therefore, the creation of terminological resources for different kind of users is a challenging task. In this paper, we propose a methodology for extracting terms and linguistic information useful for different receivers, hinging on appositive constructions.

Keywords

Terminological Resources, Users, Appositive Constructions

1. Introduction

Terminology, intended as the set of different terms composing the vocabulary of a specific discipline [1], is the core of every Language for Specialized Purposes (LSPs), which convey the most informative content in written or spoken communication.

Terminology gathering often involves the processing of actual texts, since they contain terms as the users of that specific language would use them [2]; therefore, investigating large collection of electronic texts, i.e., corpora, constitutes a valuable solution for the analysis of LSP and for the creation of a terminological resource [3].

In this study, we propose a methodology for designing different types of terminological resources intended for different target users who gravitate around LSPs, i.e., domain experts, non-experts, and translators.

Taking into consideration the users' needs in relation to technical terminology is fundamental for the creation of a suitable resource [4]. Indeed, the intended users may resort to a terminological resource for different purposes: laypeople mainly aim at understanding the term they perceive as new, difficult, and semantically obscure to decipher with the sole background knowledge at their disposal; on the other hand, translators, interpreters, and language experts in general, might more be interested in terminological equivalents in different languages, terminological variants, synonyms, part-of-speech, definitions as well as examples of the term in the context of a sentence; finally, technical writers, who fall under the category of experts in the

1st International Conference on "Multilingual Digital Terminology Today. Design, representation formats and management systems", 16 - 17 June 2022, Padua, Italy

gsperanza@unior.it (G. Speranza); mpdibuono@unior.it (M. P. Di Buono); jmonti@unior.it (J. Monti)

0000-0002-9249-492X (G. Speranza); 0000-0001-7284-6556 (M. P. Di Buono); 0000-0002-4563-5988 (J. Monti)

domain of knowledge, may use such resources as an aid to better convey technical concepts from a purely linguistic point of view.

2. Theoretical Background

In the communication between experts, terms are used by a restricted niche to commonly designate and refer to agreed technical concepts.

Nonetheless, communication on a specialized topic can also be addressed to non-experts in the field, following the vertical dimension of specialized languages [5, 6], where the degree of technical specialism is usually mediated and lowered.

Indeed, specialized texts, especially when addressed to a diversified, non-expert audience, usually show the presence of specific linguistic structures aimed at better conveying technical concepts, such as appositive constructions.

Appositive constructions have been studied and described by several scholars [7, 8, 9] as the juxtapposition of (usually) two elements: the *anchor* and the *supplement* [10].

From a graphical point of view, the supplement can be enclosed between punctuation marks such as commas, dashes or brackets.

On the semantic and pragmatic level, appositive constructions are usually employed with the aim of reformulating, paraphrasing, explaining or adding information by means of supplements with reference to the anchors.

3. Methodology

Our hypothesis is that, provided that the domain corpus of reference is composed of specialized text addressed to non-experts, appositive constructions retrieval can be beneficial for the extraction of terms [11] and the creation of different terminological resources responding to different declared aims and users.

As a case study, we select the domain of archaeology, a sub-field of Cultural Heritage (CH), which, despite making use of a highly technical language, rich in terminology, is one of the least investigated among the LSPs, if compared to other fields of knowledge.

As a linguistic base for our study we compile an Italian-English parallel corpus composed of texts in the form of brochures, leaflets, and museums and archaeological sites guides, for a total of 200k tokens for each side of the parallel corpus. The parallel corpus is also chosen as a suitable source for dealing with terminology from the point of view of translation and interpreting.

The methodology we follow comes from the corpus linguistics field and makes use of the Corpus Query Language (CQL) provided in Sketch Engine [12], as it results a powerful tool enabling the setting of specific criteria for lexical pattern identification and investigation. We set queries able to retrieve appositional constructions in the parallel corpus and visualize them as Keyword in Context (KWIC) in the Parallel Concordance section. The queries hing on the syntactic structure and nature of appositional constructions, especially where brackets are registered, making use of both Part of Speech (PoS) and Regex.

4. Results

Examples of appositive constructions in Italian and English, mainly enclosed between brackets, retrieved from our corpus can be classified as follows:

- term and its simplified synonym/variant: i.e., *fibula (spilla)* / fibula (brooch)
- term and its description: i.e., *erma (ritratto su pilastro)* / herm (portrait on the pilaster)
- term and its description by means of the function: i.e., *praefurnium (forno per il riscaldamento)* / praefurnium (oven for heating)
- term and its description by means of the morphology: i.e. *rhyton (coppa a forma di corno)* / rhyton (a horn-shaped cup)

Making use of only the anchors or the anchor in combination with its supplement we can obtain different resources, i.e., the anchors can be used to create a term base of bilingual equivalent terms (erma(it) - herm(en)); the supplement can be used to provide a resource containing explanation of the technical term for non-experts in the field, both from a monolingual and a bilingual perspective.

5. Conclusions and Future Works

Retrieved results show that appositive constructions can be extracted and employed to:

- create a terminological resource containing the term and its simplification/reformulation (the so-called supplement) with the aim of making technical terminology intelligible for a public of non-experts in the field, following an "informational perspective", i.e. *praefurnium* is, more generally, a oven (for heating).
- identify bilingual terminological equivalents (candidate terms) and monolingual terminological variants in order to create a terminological resource intended for translators, interpreters, or language experts, according to a "professional perspective", i.e., *erma* in Italian is the equivalent of "herm" in English. Furthermore, an alternative term for "fibula" is "brooch".
- help technical writers by providing them a linguistic and semantic based guide with the most frequent and productive ways of dealing with terminology, from an "educational perspective", i.e., technical terms can be exemplified by a description of their function (purpose), or their morphology (shape) or a more generic synonym.

As future work it would be interesting to enlarge the domain corpus in order to obtain even more results. Furthermore, a parallel corpus in more than two languages would also be useful for the creation of a multilingual resource.

Acknowledgments

This research has partly been supported by POR Campania FSE 2014/2020 funds and by the National Operative Program (Programma Operativo Nazionale - PON).

References

[1] J. C. Sager, Practical course in terminology processing, John Benjamins Publishing, 1990.

[2] M. T. Cabré, Terminology: Theory, methods, and applications, volume 1, John Benjamins Publishing, 1999.

[3] L. Bowker, J. Pearson, Working with specialized language: a practical guide to using corpora, Routledge, 2002.

[4] G. Rondeau, Introduction à la terminologie, 1984.

[5] M. Cortelazzo, Lingue speciali: la dimensione verticale, Unipress, 1994.

[6] M. Dardano, Linguaggi settoriali e processi di riformulazione, Parallela 3 (1987) 134–145.

[7] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, A comprehensive grammar of the english language, Longman, 1985.

[8] C. F. Meyer, Apposition in contemporary English, Cambridge University Press; Cambridge [England]; New York, 1992.

[9] N. Burton-Roberts, Parentheticals, Keith Brown (ed.), Encyclopaedia of language and linguistics, 2nd edn., vol. 9, (2006) 179–182.

[10] R. Huddleston, G. Pullum, The cambridge grammar of the english language, Zeitschrift für Anglistik und Amerikanistik 53 (2005) 193–194.

[11] G. Speranza, M. P. Di Buono, J. Monti, Terms and appositions: What unstructured texts tell us, in: Formalizing Natural Languages: Applications to Natural Language Processing and Digital Humanities, Springer, 2021, pp. 219–230.

[12] A. Kilgarriff, V. Baisa, J. Bušta, M. Jakubíček, V. Kovář, J. Michelfeit, P. Rychlỳ, V. Suchomel, The sketch engine: ten years on, Lexicography 1 (2014) 7–36.