The Role of Conceptual Relations in the Drafting of Natural Language Definitions: an Example from the Biomedical Domain
Sara Carvalho, Rute Costa, Christophe Roche

To cite this version:
Sara Carvalho, Rute Costa, Christophe Roche. The Role of Conceptual Relations in the Drafting of Natural Language Definitions: an Example from the Biomedical Domain. LREC 2018, 2018, Miyazaki, Japan. hal-01829913

HAL Id: hal-01829913
https://hal.archives-ouvertes.fr/hal-01829913
Submitted on 4 Jul 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The Role of Conceptual Relations in the Drafting of Natural Language Definitions: an Example from the Biomedical Domain

Sara Carvalho\textsuperscript{1,2}, Rute Costa\textsuperscript{2}, Christophe Roche\textsuperscript{3}

1 ESTGA – School of Technology and Management – University of Aveiro  
R. Comandante Pinho e Freitas, 28 3750-127 Águeda - Portugal  
2 NOVA CLUNL – Faculty of Social Sciences and Humanities – Universidade NOVA de Lisboa  
Avenida de Berna, 26-C 1069-061 Lisboa – Portugal  
3 Condillac Research Group – LISTIC – Université de Savoie Mont Blanc  
Campus Scientifique 73376 Le Bourget du Lac – France

E-mail: sara.carvalho@ua.pt, rute.costa@fcssh.unl.pt, christophe.roche@univ-savoie.fr

Abstract

Within the scope of the EndoTerm project, described in more detail in (Carvalho, Costa, & Roche, 2016; Carvalho, Roche, & Costa, 2015), this paper aims to explore Terminology’s key role in supporting one of the fundamental forms of concept representation - the definition -, namely by assuming a double dimensional perspective in which the conceptual backbone supports the writing process. In particular, the article will focus on how conceptual information (i.e. the concept’s position in the concept system, its characteristics, as well as the relationships linking it to other concepts) can be organised into a template-like format which would constitute the foundation of the natural language definition drafting process.

Keywords: conceptual relations, natural language definition, biomedicine

1. Introduction

In recent decades, the biomedical domain has undergone substantial changes: on the one hand, ageing population and the considerable decrease of the old-age support ratio have put more pressure on public health expenditure, raising concerns about the sustainability of social security systems and their role in health care; on the other hand, patients are playing an increasingly active and empowered role in their own healthcare; furthermore, technological innovation has been fostering an exponential growth in healthcare that is embodied not only in the widespread use of computerized examinations, procedures, prescriptions, and health records, but also in breakthroughs such as nanotechnology, 3D printing, robotic surgery, genomics, wearable technology, as well as the use of virtual, augmented and/or mixed reality. At the core of this healthcare revolution are the current biomedical terminological systems and their role in health care; on the other hand, patients are playing an increasingly active and empowered role in their own healthcare; furthermore, technological innovation has been fostering an exponential growth in healthcare that is embodied not only in the widespread use of computerized examinations, procedures, prescriptions, and health records, but also in breakthroughs such as nanotechnology, 3D printing, robotic surgery, genomics, wearable technology, as well as the use of virtual, augmented and/or mixed reality.

The ability to process and exchange clinical knowledge. The ability to provide secure, reliable, efficient and cost-effective ways to process and exchange clinical data, information, and knowledge. The ability to provide secure, reliable, efficient and cost-effective ways to process and exchange clinical information among the various stakeholders has become the foundation of eHealth action plans and programs worldwide, supported mainly by interoperability, i.e. “the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged” (HIMSS, 2013). Therefore, this paper aims to explore Terminology’s contribution to knowledge representation, knowledge organisation, and knowledge sharing in the biomedical

domain. Anchored in a double dimensional approach to terminology work, the article will focus on how conceptual information can support the natural language definition drafting process. As regards its structure, the paper will be organised as follows: Section 2 presents an overview of the current biomedical terminological systems and their increasing need for natural language definitions, and Section 3 reviews the aforementioned double dimension perspective and its impact on the creation of natural language definitions. Section 4 is dedicated to the methodological approach underlying the EndoTerm resource, with a case study based around the concept of Laparoendoscopic single-site surgery\textsuperscript{4} and encompassing both human- and machine-oriented formats, whereas section 5 provides examples of natural language definitions for Laparoendoscopic single-site total hysterectomy and Laparoendoscopic single-site ovarian cystectomy. The final section summarises the main findings and outlines future lines of research.

2. Biomedical terminological resources and interoperability: is there still a place for natural language definitions?

As stated earlier, interoperability has become one of the ‘hot topics’ in healthcare, insofar as a successful implementation of interoperable solutions can contribute to enhance the quality and outcomes in the sector, while decreasing costs (Coiera, 2015). Yet, interoperability has also become one of the most challenging topics, due to the underlying complexity of delivering “the right information,

\footnotesize
\textsuperscript{1} A type of surgical procedure that is becoming increasingly prevalent in several medical specialties, including gynaecology. It is also known as LESS surgery.

\footnotesize
\textsuperscript{2} Throughout this paper, concepts will be capitalised and written between single chevrons, while terms will be presented in lower case and between double quotation marks (cf. (Roche, 2015)).
at the right time, to the right place” (Benson & Grieve, 2016). Thus, one of the key priorities in recent years has been to devise systems and applications that allow machines, rather than humans, to accurately communicate with each other (Sicilia & Balazote, 2013).

In this regard, the most recent versions of biomedical terminological systems (e.g. the Disease Ontology, the Unified Medical Language System (UMLS), or SNOMED CT) have been focusing predominantly on finding a solid conceptual foundation supported by formal (i.e. logic-based and computer-processable) concept definitions, as well as by Semantic Web standards, such as RDF and OWL, so as to enable inter-resource mapping. Within this framework, one might wonder whether there is still room in such resources for natural language definitions of concepts. It would appear so.

One of the short-term objectives of the Disease Ontology, for instance, is to expand the number of textual definitions until reaching full coverage (Kibbe et al., 2015). The 11th version of the International Classification of Diseases (ICD-11), to be released this year, will include “a short concise textual definition” for each entity, a feature that does not exist in the existing ICD-10 (WHO, 2011, p. 17). That will also be the case with the International Classification of Health Interventions (ICHI), currently awaiting its official release and where definitions will be used to “describe the intervention” and “assist the user in selecting the most appropriate [intervention] code” (ICHI, 2018). Moreover, and despite the fact that the current version of SNOMED CT lacks natural language definitions, it is also likely that this issue will be addressed soon. On the one hand, 63% of SNOMED CT users stated, in a 2010 survey, that textual definitions would be extremely relevant (Elhanan, Perl, & Geller, 2011). On the other hand, SNOMED CT’s expected widespread use at an international level (e.g. it will fully replace the Read Codes in the UK National Health Service’s Primary Care System by April 2018) will presumably gather various stakeholders with different areas of expertise and subsequently raise particular needs, one of them probably being natural language definitions.

Notwithstanding this growing interest in textual definitions, no unequivocal guidance has been explicitly provided by the aforementioned biomedical terminological resources or their respective guidelines on how to draft such definitions. In ICD-11, for instance, contributors proposing a definition are advised to “describe the entity clearly and concisely” (WHO, 2011, p. 19), as well as to resort to existing definitions as much as possible. However, no further, more specific, drafting recommendations are outlined. The overall picture is not very different in the remaining biomedical terminological resources. In fact, one of the few - and pertinent - references to the governing principles of such definitions is to be found at the Draft ICHI Guidelines, which state that the definitions should “reflect the (...) axis categories from which the code is constructed”, thereby pointing towards the conceptual core structure of the classification as a useful starting point in the development of natural language definitions. Yet, once again, no additional information is given.

Bearing all of this in mind, it is believed that the current work can provide a contribution to systematising the natural language definition drafting process within this subject field, as will be further explored in the following sections.

3. Terminology: a matter of concepts and a matter of terms

At the heart of the work being carried out in this research project is the assumption that Terminology has a double dimension4, linguistic and conceptual, in an approach that regards it as both a “science of objects and a science of terms” (Roche, 2015, p. 136). Therefore, terminology work needs to consider not only the analysis of discourses produced by experts but also the formal (or semi-formal) representations of the shared knowledge regarding their respective domains. For (Costa, 2013), the specificity of Terminology as an autonomous scientific subject lies precisely in these two dimensions and in studying the way they interrelate and become complementary. In short, the analysis of specialised texts, on the one hand, and the collaborative work with experts, on the other hand, play a key role in terminology work, supported by a theoretical and methodological framework that allows the terminologist to maximise the potential within each dimension and the synergies resulting from their interaction.

One of the areas of terminology work where the impact of this complementary approach can become more visible is precisely the definition, one of the core forms of concept representation and a topic that has been widely debated in Terminology for quite some time (de Bessé, 1997; Löckinger, Kockaert, & Budin, 2015; Rey, 1995; Sager, 1990, 2000; Sager & Ndi-Kimbi, 1995; Seppälä, 2007; Temmerman, 2000). According to the 1087-1 and 704 ISO standards (ISO, 2000, 2009), a terminological definition should allow a concept to be differentiated from other related concepts, either by stating its superordinate concept and the respective delimiting characteristics (intensional definition - regarded as preferential by ISO whenever possible) or by enumerating all its subordinate concepts under a given criterion of subdivision (extensional definition).

However, other approaches to Terminology (cf. (Meyer, Bowker, & Eck, 1992; Temmerman, 2000)) have highlighted the limitations and the lack of flexibility of such definitions, especially in more multi- or interdisciplinary subject fields, proposing, instead, a ‘definitional template’ that reflects the position that a given concept occupies in the conceptual system it belongs to. This has also been the case in Frame-based approaches to terminology work (Durán-Munoz, 2016; Faber, 2012, 2015) and to lexicography (Maks, 2006; Swanepoel, 2011), plus work by Fillmore (e.g. (Charles J. Fillmore, 2003; C. J. Fillmore & Atkins, 1994)).

Therefore, and within the scope of the EndoTerm project, examples will be provided in the following sections of how conceptual information (i.e. the concept’s position in the

---


5 This approach has been described in more detail by (Costa, 2013; Roche et al., 2009; Roche, 2012, 2015; Santos & Costa, 2015).
4. EndoTerm: a double dimensional approach to terminology work within the biomedical field

The EndoTerm project\(^6\) aims at the creation of a terminological resource focusing on medical terminology, namely on Endometriosis, a benign gynecologic condition affecting approximately 10% of women of reproductive age worldwide (Adamson, Kennedy, & Hummelshoj, 2010; Dunselaar et al., 2014). Destined to future experts, experts of other, related domains, and also to expert patients, this research seeks to integrate both the linguistic and the conceptual dimensions in terminology work by relying on specialised corpus collection and analysis, as well as on a formal ontology, respectively. The latter constitutes the backbone of the aforementioned resource, combining hierarchical and non-hierarchical concept relations that allow a more accurate representation of the shared knowledge within this particular domain, as will be further explored in this section.

The development of EndoTerm led to the study of single-port surgery, a relatively recent type of surgical procedure that has been gaining significant ground regarding the treatment of gynecologic diseases, endometriosis being among them. A more detailed analysis of specialised resources from the subject field, including verbal, non-verbal, and multimedia content, pointed towards a lack of terminological consensus among the expert community, having identified more than 20 different terms in the literature (Carvalho, Costa, & Roche, 2016). In order to solve this terminological dispersion, the multidisciplinary Laparoendoscopic Single-Site Surgery Consortium for Assessment and Research (LESSCAR) issued a White Paper (Gill et al., 2010) that aimed to standardise the terminology in the field, proposing the term “laparoendoscopic single-site surgery” as the one that most accurately depicted this surgical procedure.

The analysis of the aforementioned sources, together with the feedback of senior expert gynaecologists who are also surgeons, helped ground the development of a micro-concept system concerning the main types of surgery performed in cases of endometriosis. As can be seen from Figure 1 below, this micro-concept system allows <Laparoendoscopic single-site surgery> to be positioned within the broader concept of <Surgical procedure> by resorting to a specific difference, Aristotelian-based convention.

The figure depicts the initial stage of that conceptualisation process, i.e. a semi-formal concept representation developed with CMap Tools\(^7\). Moreover, three main axes of analysis were set up, thereby allowing the following specific differences to be outlined at each stage: i) degree of invasiveness\(^6\): /invasive/ vs. /minimally invasive/; ii) existence of skin incision: /with skin incision/ vs. /without skin incision/; iii) number of skin incisions: /single skin incision/ vs. /multiple skin incisions/.

Through this conceptual representation, it is possible to conclude that the existence of a single skin incision constitutes the essential characteristic (ISO, 2000) of <Laparoendoscopic single-site surgery>. Furthermore, it also allows a clearer distinction between different surgical approaches, i.e. the routes used to access the procedure site. In this case, <Laparotomy> is an example of an open abdominal approach, <Laparoscopy> and <Laparoendoscopic single-site surgery> of a percutaneous endoscopic approach (either intraluminal or transluminal), whereas a procedure such as the <Natural orifice transluminal endoscopic surgery> (also known as NOTES) resorts to a per orifice transluminal approach\(^10\).

As previously mentioned, this conceptual backbone can provide a valuable contribution to the development of natural language definitions, or to the enhancement of existing definitions. However, it is insufficient to distinguish between different surgical procedures that use the same surgical approach (e.g. <Laparoendoscopic single-site hysterectomy> is a <Laparoendoscopic single-site surgery> vs. <Laparoendoscopic single-site ovarian cystectomy>

\(^6\) Described in more detail in (Carvalho, Costa, & Roche, 2016; Carvalho, Roche, & Costa, 2015).

\(^7\) A freely available software developed by the Florida Institute for Human and Machine Cognition (IHMC) and available at https://cmap.ihmc.us/cmaptools.

\(^8\) Following the existing lexicographic and terminological definitions, it has been assumed that all surgical procedures are, to some extent, invasive.

\(^9\) As referred to earlier regarding the concepts and terms, the aforementioned differences also follow a typographical convention, being represented, in this case, between forward slashes.

\(^10\) This results from a systematisation of the approaches listed on a set of current procedure classifications and other related biomedical terminological systems, such as SNOMED-CT, the IHC, the ICD-10-PCS (Procedure Codes), used in the United States, the Canadian Classification of Health Interventions (CCI), and the French Classification Commune des Actes Médicaux (CCAM).
is_a <Laparoendoscopic single-site surgery> is_a <Minimally invasive surgical procedure with single skin incision>.

Therefore, and within the scope of the work that has been developed for EndoTerm, it is proposed that the preceding conceptualisation can be enhanced not only via hierarchical, but also non-hierarchical relationships\textsuperscript{11}, as well as a systematised categorial structure\textsuperscript{12} for terminological systems of surgical procedures (ISO, 2012). The table below illustrates EndoTerm’s conceptual framework regarding surgical procedures, in line with the ISO 1828: 2012, and includes the core top-level concepts, a set of is_a and non-hierarchical relationships and, lastly, the authorised Source Concept - Relationship - Target Concept combinations\textsuperscript{13}.

<table>
<thead>
<tr>
<th>SOURCE CONCEPT</th>
<th>RELATIONSHIP</th>
<th>TARGET CONCEPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Surgical procedure&gt;</td>
<td>is_a</td>
<td>&lt;Procedure&gt;</td>
</tr>
<tr>
<td>&lt;Surgical procedure&gt;</td>
<td>has_method</td>
<td>&lt;Surgical action&gt;</td>
</tr>
<tr>
<td>&lt;Surgical procedure&gt;</td>
<td>has_procedure_site</td>
<td>&lt;Human anatomy&gt;</td>
</tr>
<tr>
<td>&lt;Surgical procedure&gt;</td>
<td>has_morphology</td>
<td>&lt;Lesion&gt;</td>
</tr>
<tr>
<td>&lt;Surgical procedure&gt;</td>
<td>has_surgical_approach</td>
<td>&lt;Procedural approach&gt;</td>
</tr>
<tr>
<td>&lt;Surgical procedure&gt;</td>
<td>uses_access_device</td>
<td>&lt;Device&gt;</td>
</tr>
<tr>
<td>&lt;Lesion&gt;</td>
<td>has_procedure_site</td>
<td>&lt;Human anatomy&gt;</td>
</tr>
<tr>
<td>&lt;Device&gt;</td>
<td>has_procedure_site</td>
<td>&lt;Human anatomy&gt;</td>
</tr>
<tr>
<td>&lt;Device&gt;</td>
<td>has_procedure_site</td>
<td>&lt;Lesion&gt;</td>
</tr>
</tbody>
</table>

Table 1: EndoTerm’s categorial structure.

The following micro-concept systems - built around the concepts of <Laparoendoscopic single-site total hysterectomy> (Figure 2) and <Laparoendoscopic single-site ovarian cystectomy> (Figure 3)\textsuperscript{14}, respectively - demonstrate how the template structure referred to above can help overcome the limitations of fully hierarchical concept representations, while providing a logical foundation that can prevent logical errors, especially at an initial, semi-formal stage where automatic reasoning may not be available.

To further substantiate the preceding approach, all of EndoTerm’s micro-concept systems were then tested using TeDI (for OntoTerminology EDItor), a software environment created by C. Roche dedicated to the development of multilingual ontoterminologies\textsuperscript{15}. In this case, and via TeDI, it was possible to validate EndoTerm’s semi-formal concept systems and convert them into a formal ontology, also benefiting from the tool’s built-in reasoner and from the subsequent logical verification that takes place during the ontology development process. The image below (Figure 4) shows a glimpse of TeDI’s concept editor, namely from the concept <Laparoendoscopic single-site total hysterectomy>, its position in the hierarchy, the specific differences, as well as one of the non-hierarchical relationships (has_procedure_site).

This formal concept definition can also be exported into W3C-compliant formats (RDF/XML), which can pave the way to a potential integration into existing biomedical, concept-oriented terminological resources.

\textsuperscript{11} Despite their secondary role in the current ISO standards related to terminology and terminology work (ISO, 2000, 2009), non-hierarchical concept relationships are regarded as “equally important and more revealing about the nature of the concepts” (Sager, 1990, p. 34), as well as extremely relevant in the biomedical domain (cf. McCray & Bodenreider, 2002; A. L. Rector et al., 1997; Smith et al., 2005)).

\textsuperscript{12} i.e. a “minimal set of domain constraints for representing concept systems in a subject field” (ISO, 2007).

\textsuperscript{13} In Description Logic, the source and target concepts are also known as domain and range, respectively, and they are also subject to constraints (Baader, 2003; A. Rector & Rogers, 2006).

\textsuperscript{14} Hysterectomy, often seen as a last resort in cases of severe endometriosis (Peter Rogers et al., 2016; Rogers et al., 2009), and ovarian cystectomy, i.e. the removal of ovarian endometriotic cysts or endometriomas (Working group of ESGE, ESHRE, and WES et al., 2017), are two common surgical procedures as regards the management and treatment of endometriosis.

\textsuperscript{15} An ontoterminology is “a terminology whose conceptual system is a formal ontology” (C. Roche & Calberg-Challot, 2009). More information on the software can be found at http://christophe-roche.fr/tedi.
Figure 5: Formal definition of <Laparoendoscopic single-site total hysterectomy> in RDF/OWL.

5. Terminological definitions in EndoTerm: two examples

Based on the validated conceptualisation explored above, a template-based natural language definition can be put forward for each of the analysed LESS surgery concepts, with direct reference to both the specific difference approach and to the non-hierarchical relationships, supported by the categorial structure.

Concept 1: Type of <Surgical procedure> has_method <Surgical action> has_procedure_site <Human anatomy> uses_access_device <Device>

Hence, for the concept of <Laparoendoscopic single-site total hysterectomy>, the proposed definition is the following:

<Minimally invasive surgical procedure> which consists of the <Excision> of the <Uterus> and <Cervix>, using a <Laparoscope> as an access <Device> via a /single skin incision/.

Concept 2: Type of <Surgical procedure> has_method <Surgical action> has_morphology <Lesion> has_procedure_site <Human anatomy> uses_access_device <Device>

Regarding the concept of <Laparoendoscopic single-site ovarian cystectomy>, the proposal would read:

<Minimally invasive surgical procedure> which consists of the <Excision> of a <Cyst> located in the <Ovary>, using a <Laparoscope> as an access <Device> via a /single skin incision/.

Finally, it is believed that EndoTerm’s knowledge organisation proposal, grounded by the outlined methodology and theoretical background, will enable an integration with some of the existing biomedical terminological resources dedicated to procedures, especially the ICHI and SNOMED CT. Despite the fact that these resources do not currently encompass any natural language definitions, nor any guidelines or drafting principles, as stated earlier, their solid concept orientation will undoubtedly constitute a valuable framework in that almost inevitable process. And when that happens, it is expected that EndoTerm can help to enhance the yet rather marginal presence of <Laparoendoscopic single-site surgery> - and other endometriosis-related concepts - in existing biomedical terminological resources.

6. Concluding remarks

This paper aimed to demonstrate that conceptual representations, in this case an ontology supported by a combination of the specific difference approach and a categorial structure for procedure concepts, can make a valuable contribution to the current lack of natural language definitions in most of the biomedical terminological resources. By providing an organised and clear framework of interrelated concepts, relationships, and domain constraints, these conceptualisations can become useful allies against the limitations of the so-called traditional terminological definitions. The ongoing changes regarding the way medical information and knowledge are produced, used, stored and shared require efficient and reliable solutions, in a society that demands immediate and multi-platform access to all digital content. If one of the main postulates of terminology work is to provide tools and services that can respond to the concrete needs of a given target audience, at a certain moment in time, within a specific domain, and under particular circumstances, then terminological projects developed within the subject field of healthcare, especially those focusing on knowledge representation, knowledge organisation and knowledge sharing, must take the above-mentioned background into consideration.

7. Acknowledgements

This research has been financed by Portuguese National Funding through the FCT – Fundação para a Ciência e Tecnologia as part of the project Centro de Linguística da Universidade Nova de Lisboa – UID/LIN/03213/2013.

8. Bibliographical References


