Ontotermology meets Lexicography: the Multimodal Online Dictionary of Endometriosis (MODE)

Sara Carvalho123, Rute Costa23, Christophe Roche22

1 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
Av. 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@fsh.unl.pt, christophe.roche@univ-savoie.fr

Abstract

With the advent of the Semantic Web and, more recently, of the Linked Data initiative, the need to operationalise lexicographic resources, i.e. to represent them in a computer-readable format, has become increasingly important, as it contributes to pave the way to the ultimate goal of interoperability. Moreover, the collaborative work involving Terminology and ontologies has led to the emergence of new theoretical perspectives, namely to the notion of Ontotermology, which aims to reconcile Terminology’s linguistic and conceptual dimension whilst preserving their core identities. This can be particularly relevant in subject fields such as Medicine, where concept-oriented and ontology-based approaches have become the cornerstone of the most recent (bio)medical terminological resources, and where non-verbal concept representations play a key role. Due to the lack of specialised lexicographic resources in the field of endometriosis, this paper aims to present the MODE project, i.e. the Multimodal Online Dictionary of Endometriosis, a multilingual resource comprising several types of data, namely video articles, a new type of scholarly communication in Medicine. It is believed that introducing a medical lexicographic resource supported by ontotermological principles and encompassing scientific video articles may constitute a relevant window of opportunity in the research field of Lexicography.

Keywords: terminology; ontotermology; e-lexicography; multimodal dictionary; ontology; endometriosis

1. Introduction

With the advent of the Semantic Web1 and, more recently, of the Linked Data initiative 2, the notion of operationalisation, i.e. the creation of computer-readable representations, has become increasingly important, as it contributes to pave the way to the ultimate goal of interoperability. Moreover, the collaborative work involving Terminology and ontologies – in the sense of Knowledge Engineering (KE) – has led to the emergence of new theoretical perspectives, one of them being Ontotermology (Roche et al. 2009), which aims to reconcile Terminology’s linguistic and conceptual dimensions whilst preserving their core identities (Roche (2012, 2015); Costa (2013); Santos & Costa (2015)). This can be particularly relevant in subject fields such as Medicine, where concept-oriented and ontology-based approaches have become the cornerstone of the most recent (bio)medical terminological resources, and where non-verbal representations play a key role. It is believed that ontotermological principles may provide a relevant theoretical and methodological contribution to the research field of Lexicography by supporting the creation of specialised online lexicographic resources, especially in domains that lack those resources, as is the case with endometriosis. Therefore, this paper aims to present the MODE project, i.e. the Multimodal Online Dictionary of Endometriosis, a multilingual resource comprising several types of data, namely video articles, a new type of scholarly communication in Medicine.

This article will thus be structured as follows: section 2 will focus on the theoretical background, specifically regarding Terminology’s double dimension, the Ontotermology approach and how both can relate to Lexicography; section 3 will provide a brief overview of endometriosis, not only as a subject field per se, but also in what concerns the existing specialised lexicographic resources; section 4 will be dedicated to the MODE project, with a description of its supporting principles and core structure, followed by a final section consisting of concluding remarks.

2. Terminology, ontologies and Lexicography

2.1 Terminology’s double dimension

As mentioned above, this approach, which encompasses a linguistic and a conceptual dimension that are interrelated, has been described by Roche (2012, 2015), Costa (2013)

---

and by Santos & Costa (2015). According to Roche (2015: 136), Terminology is “both a science of objects and a science of terms”. For Costa (2013), it is precisely this double dimension, as well as the study of the relationship between one and the other, that grants Terminology its place as an autonomous scientific subject.

This double dimension perspective implies, therefore, that both the experts’ conceptualisations of a given subject and the discourses produced by them must be taken into account in terminology work. In a nutshell, the cornerstone of this approach lies in the complementarity of these two fundamentally different dimensions, as two sides of the same coin.

Among the theoretical perspectives that have emerged in recent years involving Terminology and the role of ontologies, Ontoterminalogy is the one that best suits the objectives of the MODE project, and thus will be presented in more detail below.

2.2 Ontoterminalogy: a new approach to Terminology?

Proposed by Roche et al. (2009), Ontoterminalogy aims to reconcile Terminology’s linguistic and conceptual dimensions while maintaining their fundamental differences. Defined as a “terminology whose conceptual system is a formal ontology” (Roche et al., 2009: 325), this approach acknowledges the conceptualisation of a given domain as the starting point of any terminological project, hence corroborating ISO 704’s perspective that “producing a terminology requires an understanding of the conceptualisation that underpins human knowledge in a subject area” (2009: 3).

As mentioned before, even though the conceptual dimension plays a key role in Ontoterminalogy, due to the potential of operationalising the conceptualisations of a given subject field – thus enabling interoperability –, this does not mean that natural language should be excluded from terminology work. In fact, “to conceptualise, one must verbalise” (Roche, 2015: 149). Albeit with vagueness and inconsistencies, the discourses provide fundamental access to the expert community, especially in some areas of expertise where the main goal is knowledge stabilisation and dissemination, as is the case of endometriosis.

Consequently, both specialised texts and expert collaboration constitute invaluable resources in terminological work, provided that there is a supporting theoretical and methodological framework through which it can be possible to maximise the potential of each dimension, and mostly of the synergies resulting from their interaction.

What is important to emphasise, according to Ontoterminalogy, is that even though the conceptual and linguistic dimensions rely on two diverse semiotic systems that should not be confused, both of them have their place in projects and products supported by ontoterminalogy principles. As a matter of fact, this approach proposes the double semiotic triangle, an extension of Ogden and Richards’s proposition (1923) which allows a distinction between the definition of the term, written in natural language, and the definition of the concept, which may resort to either a formal or a semi-formal language (Roche, 2012). It is believed that when anchored in this approach, terminology work may contribute to further enhance the quality of specialised communication.

2.3 Ontoterminalogy and Lexicography: is collaboration possible?

As Terminology, in the last few decades, Lexicography has been searching for its identity as an autonomous scientific discipline in its own right, with an intense debate around the principles that should support lexicographic theory and practice (cf. Wiegand 1997, 1998; Bergenholtz & Tarp 2003; Atkins & Rundell 2008; Tarp 2008; Béjoint 2010; Hartmann 2010; Fuertes-Olivera & Bergenholtz 2011; Granger & Paquot 2012; Fuertes-Olivera & Tarp 2014).

Part of this discussion pointed, understandably, towards delimiting and positioning Lexicography scientifically, as well as its branches, namely Specialised Lexicography.

In this context, a lot has been written about the need to distinguish Specialised Lexicography from Terminology. They are indeed different, first and foremost because the former studies the units of the specialised lexicon and the way they behave in discourse, whereas the latter focuses not only on the linguistic dimension, but also on a conceptual dimension that cannot be underestimated and is in fact embodied in terms⁴.

However, and despite the differences, some consider that Specialised Lexicography and Terminology are not necessarily incompatible and that both areas could benefit from collaborative work (cf. Humbley 1997; Costa 2013). Fuertes-Olivera & Tarp (2014) refer to the existing interaction between Specialised Lexicography and Terminology in the conception and production of a number of reference works, particularly within the scope of the Function Theory of Lexicography (FTL), although they do not further specify how this interaction actually takes place.

As previously stated, this paper intends to show how Terminology, and particularly Ontoterminalogy, may contribute to the work carried out by Lexicography without undermining both research fields.

First of all, terminological work, as lexicographic practice, relies on a key premise: to have users and their respective needs in mind. In fact, the social responsibility [gesellschaftliche Verantwortung] that, according to Wiegand (1997), should characterise Lexicography as a scientific discipline could also be applied to Terminology. However, it should be noted that, in Terminology, the user may not necessarily be human – at least the primary user –, which will consequently determine the purpose, structure and content of the resource to be developed, as well as the also be perceived as units of representation of the concepts of a given subject field. As such, they have the capacity to exist outside of discourse, pointing towards the concept and providing access into the subject field (cf. Carvalho, Roche, & Costa, 2015).

---

³ The lexical networks extracted from corpora may not always match the conceptual systems resulting from the collaboration of subject field experts – “Saying is not modelling” (Roche, 2007).
⁴ If, on the one hand, terms are in fact units of discourse, they can
medium. Secondly, it is believed that the added value of Terminology in the conception and development of specialised lexicographic resources lies precisely in its double-dimensional nature, and in the fact that the conceptual dimension – substantiated in its knowledge organisation potential – may, in turn, support the linguistic dimension, namely by assisting in the drafting of natural language definitions. The ontotermological approach aims to take this contribution to the next level: by placing the ontotermology at the heart of a given resource, it intends to provide a stable conceptual backbone of a subject field, built in collaboration with subject field experts, and which may become the basis of other, derived products, such as terminology databases, specialised dictionaries, thesauri, etc. The types and amount of data to be made available would then depend on the user profile, on his/her needs, as well as on the social situations and contexts, yet this conceptual core structure, which might or might not be visible to the human user, would remain the same⁵.

As described in the previous subsection, Ontoterminology does not underestimate the linguistic dimension: in fact, it values it, by allowing linguistic diversity to be registered, which is seldom the case in ontology-based approaches. Section 4 will provide an example as to how synonymy and equivalence, for instance, can – and should – have their place within this project. To sum up, Terminology can play a role in the creation of specialised lexicographic products, both from a linguistic and a conceptual perspective. Within the framework of Ontoterminology, the latter constitutes a valuable foundation which may contribute to enhance the quality of specialised communication.

3. Endometriosis: facts and figures
Endometriosis is defined as “the presence of endometrial-like tissue outside the uterus, which induces a chronic, inflammatory reaction” (Kennedy et al. 2005). The exact prevalence of the disease is unknown, but it is believed to affect an estimated 176 million women of reproductive age worldwide (Adamson, Kennedy, & Hummelshoj 2010). While its aetiology is uncertain, it is likely to be multifactorial, including genetic, immunological, endocrinological and environmental influences.

Women with endometriosis typically have a range of pain-related symptoms, such as dysmenorrhea, dyspareunia, dyschezia, dysuria, non-cyclical pelvic pain, as well as chronic fatigue (Dunselman et al. 2014). A recent study conducted in 10 countries throughout the world has reported an overall diagnostic delay of 6.7 years (Nnoaham et al. 2011). Moreover, the World Endometriosis Research Foundation (WERF) EndoCost study (Simonsens et al. 2012) has shown that the costs arising from women with endometriosis treated in referral centres are substantial (an average annual total cost per woman of €9,579), an economic burden that is at least comparable to the costs of other chronic diseases, such as diabetes, Crohn’s disease, or rheumatoid arthritis.

Taking into account the estimated 10% prevalence of the disease among women of reproductive age around the world, which is significant, it is surprising to realise that there are very few specialised language-related resources dedicated to it – lexicographic or of any other nature. In fact, there is, to our knowledge, only one reference work published under the name “Dictionary of Endometriosis” (Parker & Parker 2003), yet this resource is more of an annotated bibliography and a research guide to Internet references concerning the disease. The “dictionary” section is actually a monolingual glossary, in English, with about 1,300 terms and their respective definitions, taken, according to the authors, both from the National Institutes of Health⁶ and the European Union, although it is never mentioned where exactly from the EU these definitions stem from. An extensive search of resources on endometriosis concluded that the few that actually exist correspond mostly to the notion of glossary, perceived as a “list of designations and definitions in a particular subject field” (ISO 1087-1, 2000: 12). These lists are almost exclusively monolingual (with English as the most frequently used language), depicting a widely variable number of terms (ranging from 20 to 1,500), usually containing no sources in what concerns the definitions, and with hardly any supplementary material, namely images or videos. In addition, these resources have been built by and are destined to different types of people, and have therefore fairly distinct levels of specialisation. Some examples include the European Society for Human Reproduction and Embryology’s Guideline on the Management of Women with Endometriosis (Dunselman et al., 2013) (expert > expert or > semi-expert) and the American Society for Reproductive Medicine’s Endometriosis Guide for Patients (ASRM 2012) (expert > non-expert).

4. The MODE project
As previously mentioned, the main goal of this paper is to present the Multimodal Online Dictionary of Endometriosis (MODE), a project of a multilingual resource based around the concept of <Endometriosis>⁷, which is currently at its conception stage and aims to integrate several types of data, including medical video articles⁸.

⁵ Assuming that the knowledge in that particular domain is stable enough to be represented via a semi-formal or formal conceptualisation.
⁶ That belong to the U.S. Department of Health and Human Services and integrate the National Library of Medicine, responsible for issuing and updating PubMed/MEDLINE, Medical Subject Headings (MeSH), and MedlinePlus.
⁷ In this paper, concepts will be capitalised and written between single chevrons, whereas terms will be presented in lower case and between double quotation marks (cf. Roche, 2015).
⁸ This peer-reviewed and indexed resource has been described more thoroughly in Carvalho, Roche, & Costa (forthcoming). MODE’s guiding principles have been defined within the scope of the EndoTerm project, presented in Carvalho, Roche, & Costa.
Even though the inclusion of images and multimedia content is not new in medical lexicographic resources (whether in CDs and DVDs, in online editions, or more recently in apps, as is the case of the renowned Stedman’s Taber’s and Dorland’s Medical Dictionaries, just to name a few), MODE can offer added value supported by three essential axes: the inclusion of medical video articles and the emphasis on their potential as a new type of scholarly communication in Medicine; the choice of the subject field itself, which lacks specialised resources; and finally, its ontoterminological principles, grounded in Terminology’s double dimension. This resource, aimed primarily at future experts (medical students) or experts of other, related domains (such as nursing staff, for example), can make a valuable contribution in specialised training, which is why expert collaboration plays a critical role in helping identify relevant and realistic needs in this particular subject field. As for the situations which may lead to the consultation of such a lexicographic resource, and using the terminology adopted by the FTL (cf. Tarp, 2008; Fuertes-Olivera & Tarp, 2014), it is believed that MODE’s potential users will be mainly interested in acquiring knowledge about a particular subject (cognitive situation), rather than, for example, trying to solve a communication problem (communicative situation).

In the next few paragraphs, MODE’s core structure, as well as a methodological proposal, will be put forward. Due to space constraints, the examples to be provided will focus on the concept of <Laparoendoscopic single-site surgery>, a relatively recent type of surgical procedure that is becoming more and more prevalent in several medical specialties, and that accounts for a significant amount of endometriosis-related surgeries (Gill et al. 2010). The conceptual structure of the domain is MODE’s “beating heart”, providing, as stated in section 2, a backbone that supports the remaining components. As such, it constitutes the first by-product of the project, and has been built using OTE Soft ©, a concept modelling tool created by the Condilac Research Group (cf. Roche, 2015; Carvalho, Roche, & Costa 2015). Based on information provided by textual and multimedia sources, by current (bio)medical terminological resources (such as MeSH, UMLS and SNOMED CT), as well as by the feedback and validation of a team of senior expert gynaecologists, a set of concept maps have been created, one of which is shown below.

Although this goes beyond the scope of this paper, it is believed that the notions that have characterised the types of users of specialised lexicographic or terminological products, namely the distinction between subject field experts, semi-experts and non-experts or laypeople (cf. Bergenholz & Kaufmann 1997), are becoming more and more blurred, at least in some areas, and should therefore be discussed. Within the (bio)medical domain, for instance, one could assume that a patient would belong to the last group. However, growing digital literacy has brought the patients into the driver’s seat and has led them to play a more active – and empowered – role. In fact, patient empowerment has been at the heart of the most recent healthcare policies and initiatives, particularly at a European level (http://www.eu-patient.eu). One of the most promising projects in this respect is the European Patient Academy (EUPATI), which provides Patient Expert Training Courses destined to increase “the capacity and capability of patients to understand and contribute to medicines’ R&D” and also to improve “the availability of objective, reliable, patient-friendly information for the public”. More information at https://www.eu-patient.eu.

Figure 1: Concept map of <Laparoendoscopic single-site surgery>.

The project is currently at the beginning of its second stage, consisting of corpus collection and analysis, in order to see whether the selected texts contain designations that point towards the previously identified concepts. There are three working languages (English – eng, European Portuguese – pt and French – fr) involved and in this respect, the experts play a critical role, namely in advising as to the texts that are deemed representative and/or mandatory in the subject field of endometriosis. Based on their feedback, a text typology, i.e. “la réunion et la classification d’un ensemble des textes sous une même étiquette” (Costa & Silva 2008: 6) has been created, integrating 3 main types of texts: a) academic (comprising scientific articles and theses); b)
normative (guidelines, White Papers, and standards); and c) teaching materials (textbooks, handbooks or course books). The subsequent corpus treatment and analysis are to be conducted using AntConc® and a set of candidate terms is to be presented to the experts for validation. The next step consists of the development of the dictionary entries. At the moment, a study is being carried out regarding the layout of those upcoming entries, specifically the structure that could best suit the resource’s guiding principles, including the proposition of one entry per concept and the need for interoperability. Thus, and as it is currently not possible to present an actual entry of the MODE, the example below resorts to CMap Tools—a freely available software developed by the Florida Institute for Human and Machine Cognition (IHMC) and available at http://cmap.ihmc.us.

In order to solve this terminological dispersion, the aforementioned LESSCAR proposed the term “laparoendoscopic single-site surgery” as the one that most accurately depicted this surgical procedure. The remaining designations can be perceived as synonyms, which, from a terminological point of view, raises the dilemma of whether apples are indeed being compared to apples, i.e. whether or not all these terms are in fact representing the same concept. A more thorough analysis of this subject, which will occur after the corpus analysis is completed, is necessary in order to confirm this hypothesis and further develop it. Therefore, our “entry” proposal contains two randomly selected terms as synonyms for the term in English.

Concerning the equivalents, the data gathering accomplished thus far has confirmed the significant discrepancy between the English-speaking corpus and the ones in French and European Portuguese, which can be explained by the predominance of English in specialised communication, particularly in the academic world. Moreover, in the fr and pt texts compiled so far, no equivalents of “laparoendoscopic single-site surgery” have been found. Consequently, further research of academic texts (theses and scientific articles), as well as of teaching materials, was conducted in those two languages. The search was carried out via Google’s advanced search and the results indicate that in pt, the most frequent designation was “cirurgia laparoscópica por porta umbilical única” [single-port umbilical laparoscopic surgery], mainly within the medical specialty of Urology, whereas in fr, the term “chirurgie par accès unique” [single-access surgery] was the one most widely used. Potential synonyms have also been found in both languages, but appear to raise the same dilemmas as those mentioned above: “laparoendoscopia de incisão única” [single-incision laparoendoscopy], “cirurgia por incisão única” [single-incision surgery] (pt); “chirurgie laparoscopique par accès omégal unique” [single umbilical access laparoscopic surgery], “chirurgie laparoscopique à trocart unique” [single-port laparoscopic surgery], “chirurgie par orifice unique” [single-orifice surgery] (fr).

As mentioned before, the ontoterminal approach enables the existence of both a term and a concept definition. However, as the collection of the English corpus has not been completed up to the present moment, a natural language definition cannot be provided. Still, the designed micro-concept map containing the concept’s essential and delimiting characteristics (cf. ISO 1087-1, 2000; ISO 704, 2009) may contribute to enhance the quality of an existing natural language definition or to actually create a new one if none exists.

As regards the linguistic dimension, i.e. the term(s) designating the concept in question, a lack of terminological consensus among the expert community has been identified, with a plethora of terms coined by individual groups and organisations. In fact, more than 20 have been documented in the literature, as shown in the table below.

Table 1: Terms designating the concept of “Laparoendoscopic single-site surgery”.

<table>
<thead>
<tr>
<th>English term / Full term</th>
<th>Portuguese term</th>
<th>French term</th>
</tr>
</thead>
<tbody>
<tr>
<td>laparoendoscopic single site surgery</td>
<td>cirurgia laparoscópica por porta umbilical única</td>
<td>chirurgie laparoscopique par accès omégal unique</td>
</tr>
<tr>
<td>single-port umbilical surgery</td>
<td>single-incision surgery</td>
<td>single-port access</td>
</tr>
<tr>
<td>single-site surgery</td>
<td>single-incision laparoscopic surgery</td>
<td>single-port laparoscopic surgery</td>
</tr>
<tr>
<td>single-access surgery</td>
<td>single-incision laparoscopy</td>
<td>single-port laparoscopy</td>
</tr>
<tr>
<td>single-orifice surgery</td>
<td>single-orifice surgery</td>
<td>single-orifice surgery</td>
</tr>
</tbody>
</table>

10 The ISO 1951:2007 standard, for instance, may not suit our needs, as it explicitly mentions its “lexicographical lemma-oriented approach”, hence distancing itself from “concept-oriented works”.
11 A freely available software developed by the Florida Institute for Human and Machine Cognition (IHMC) and available at http://cmap.ihmc.us.
Having a conceptual framework as the basis of the MODE project can also contribute to facilitate and improve the insertion of supplementary material, such as images, diagrams, and videos, by acting as a sort of “tag”. Relying on a validated knowledge organisation proposal enables the inclusion of the aforementioned resources in a much more thorough way, which will undoubtedly be useful for a group of intended users seeking for detailed subject field knowledge. In addition, it can lead to a more effective customisation of the MODE.

Let us take the following example: as stated beforehand, the LESS technique is very often used in endometriosis-related surgical procedures, namely in hysterectomies, often seen as a last resort in cases where the disease strikes more severely. However, there are different types of hysterectomy (supracervical or partial, total and radical) and if a given video article describes, for instance, a supracervical hysterectomy using LESS, it is possible to add that video to the actual concept being depicted <LESS supracervical hysterectomy> and not to the more generic concept <LESS hysterectomy> or, going even further up, <Hysterectomy> (check Figure 3).

![Figure 3: Micro-concept map of the different types of <LESS hysterectomy> and example of video insertion.](image)

This issue is even more pressing when interoperability is at stake. In a study involving the creation of a dictionary for sign language, Kristoffersen and Troelsgard (2012) refer to the unsuitability, from a computational perspective, of video recordings as lemmas in a dictionary database, as they “would have to be represented by a transcription, a filename, a number, or some other sort of ID in order to be ordered or filtered” (296). A conceptual framework within the ontotermological approach would actually enable that operationalisation, i.e. that computational representation. Furthermore, and although this is not the focus of the current project, it is also believed that the experience resulting from the inclusion of medical video articles in MODE will constitute the starting point for further projects, substantiated in the content analysis and tagging of these video articles, which may then supply inputs regarding the classification, indexing and archive of these multimedia resources.

5. Concluding remarks

Through the presentation of the MODE project, this paper intended to show that the ontotermological approach can make a valuable contribution to the field of Lexicography. Rather than being perceived as incompatible, both areas combined provide added value to a research project and these synergies will certainly represent a window of opportunity in the conception and development of online specialised resources. As Gouws (2011: 29) points out, “looking to the future, (…) we must unlearn a great deal of what we know, and we must learn anew so that we can produce innovative reference tools, including dictionaries.”

6. 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.

7. Bibliographical References


8. Language Resource References


