

How translators can improve multilingual terminology in a link: teaching case study examples

Carmen Gomez-Camarero
Universidad de Málaga
gomez@uma.es

Rocío Palomares Perraut
Universidad de Málaga
perraut@uma.es

Abstract

This work describes a teaching method to enrich and improve the translators' and interpreters' multilingual terminology to translate specialized texts from a single link. This method consists of using controlled vocabularies such as thesauri, classification schemes, subject heading systems and taxonomies that employ Linked Open Data (LOD) technology in the framework of Semantic Web.

1 Introduction

The Web has become an unavoidable tool to get information. It is a platform where nodes such as objects, people and identities are interconnected. What they called *The Web of Things* (Atzori *et al.*, 2010). User translators collaborate and exchange information and knowledge in order to obtain the most accurate and appropriate data to ensure the quality of a translation work. To do that, however, the translator's literacy skills are crucial in determining the suitability to choose a selected appropriate term or expression. According to ACRL (Association of College and Research Libraries, 2016), Information Literacy (IL) "forms the basis for lifelong learning. It is common to all disciplines, to all learning environments, and to all levels of education. It enables learners to master content and extend their investigations, become more self-directed, and assume greater control over their own learning".

In this context, IL is taught in Translation Studies at the University of Málaga (Spain) to developing in students the skills they need to get the accurate and precise information to translate. One of those skills is to access to the information required effectively and efficiently. But previously, they have to know which terms are the most accurate. Why? Because keywords are essential for searching online and for narrowing results. As well, they are vital part of any scientific text paper because someone has a previous and fast view of the content of the text (Vrkic, 2014; Yang, 2010). Students are familiar with dictionaries and terminological databases to look up terms and concepts to translate but they are not used to managing controlled vocabularies as a terminological tools (Lin *et al.*, 2009; Dancette, 2011; Ramírez-Polo, 2012; Dancette, 2015). We consider they are very interesting and useful as we argue below.

It is thus in this work we describe a method to enrich and improve the translators' and interpreters' multilingual terminology to translate specialized texts from a single link. This method consists of using controlled vocabularies such as thesauri, classification schemes, subject heading systems and taxonomies. Those controlled vocabularies employ Linked Open Data (LOD) technology in the framework of Semantic Web (Pastor-Sánchez *et al.*, 2009; Pastor-Sánchez, 2011). This standard is about using the Web to connect related data that wasn't previously linked in a way that, in the context of terminology, each concept is uniquely identified by an URIs (Uniform Resource Identifier) and SKOS (Simple Knowledge Organization System) to support the use of knowledge organization systems.

This way of publishing and connecting data on the Web has myriad possibilities and challenges for accessing and identifying quality and up-to-date information and knowledge. Particular, in the case of controlled vocabularies, they are betting in recent years in the rapidly expanding LOD landscape because is the latest advancement in the natural evolution

of the Semantic Web. Controlled vocabularies valued as authoritative are being structured and published to make them openly accesible and shareable on the Semantic Web (Summers *et al.*, 2008; Subirats-Colls, 2013).

These controlled vocabularies have various functions. These include, among others: gathering together the richness of variant terms and synonyms for concept and to link concepts in a logical order, sorting terms into categories, promoting consistency in preferred terms, improving and enriching the translators' multilingual and specialized terminology. Other very important propose of these controlled vocabulary is to organize information and provide terminology to catalogue and retrieve information. In doing so, the student can also increase their knowledge in accessing to the documents that have been catalogued and indexed by those controlled vocabularies.

2 Teaching practices using controlled vocabularies

To that end, some examples employed in class in order to teach these functions and usefulness of controlled vocabularies are described. The specialized texts used are taken from SINC (Servicio de Información y Noticias Científicas), a Spanish website to disseminate scientific information. Students must identify and select the proper keywords looking up these different controlled vocabularies: UNESCO Thesaurus, Spanish Cultural Heritage Thesauri (Tesauros del Patrimonio Cultural de España), Authorities Catalogue of the National Library of Spain (Catálogo de Autoridades de la Biblioteca Nacional de España –BNE-), Répertoire d'Autorité-Matière Encyclopédique et Alphabétique Unifié (RAMEAU) of the National Library of France and the Library of Congress Subject Headings (LCSH).

2.1 A teaching practice using Authorities Catalogue of the Spanish National Library (BNE)

We will describe in this work two practices. The first one consists in offering the student a text from SINC titled “La dependencia y la edad son los principales factores de riesgo de maltrato en personas mayores“ (<http://www.agenciasinc.es/Noticias/La-dependencia-y-la-edad-son-los-principales-factores-de-riesgo-de-maltrato-en-personas-mayores>) using one of the controlled vocabularies. In this practice, we employ the Authorities Catalogue of the National Library of Spain (Catálogo de Autoridades de la Biblioteca Nacional de España). This is a tool, normally used by National Libraries, to control documents for storing and retrieve information on the catalogue. The terms are known as Subject Headings and they should be looked up by the user first before starting a researching for information.

Shorly, the text is an Spanish medical research about the high risk in older people to be abused because of their age and grade of pshycological dependency.

The aims of this practice are:

- a) Identify the main keywords of the text, selecting only three terms that represent the content of the text. The three keywords should be as accurate as possible. These terms can be chosen from the title, from the abstract or even from the text.
- b) Search these three terms in the Authorities Catalogue and verify if those are considered authorized and non-authorized entries.
- c) Compare and choose authorized entries according to the content and meaning of the text .
- d) Identify non-authorized entries in order to enrich their vocabulary.

- e) Finally, in this practice, through LOD (Authorities Catalogue of the National Library of Spain is linked with LCSH and RAMEAU) students have to search others controlled vocabularies in other languages. This is our teaching goal in this practice: improve their multilingual terminology specialized of the text they are working on.

The second step of this practice consists in describe the keywords and their semantic relationships according to the Authorities Catalogue of the Spanish National Library (BNE). In this practice, we describe the three chosen keywords by students:

Abuse (Maltrato): Searching this term on Authorities Catalogue of BNE does not appear alone but the searching result is a list of a non-authorized terms that include this word, such as: *maltrato a las personas mayores, maltrato animal, maltrato entre alumnos, maltrato infantil, maltrato psicológico en el trabajo*. In this case, the student must choose the closest non-authorized term to the meaning of the text. That it is, *maltrato a las personas mayores*. This term is non-authorized term but is linked with an authorized term: *Ancianos-Malos Tratos*.

Variants of this concept (in Spanish Usado Por –UP-) and non-authorized terms are: *Agresión a los ancianos, Ancianos—Abusos, Malos tratos a las personas mayores, Maltrato a las personas mayores* and *Violencia contra los ancianos*.

Finally, LOD technology offers the possibility to check the equivalents of this concept in other controlled vocabularies: *Older people-Abuse* in English (LCSH) and *Personnes âgées-Violence envers* (RAMEAU) in French. At the same time, the student can browse and learn the same semantic relationships in other languages. Therefore, the student enriches widely his vocabulary in other languages in a link. That is why we think this way of working the specialized terminology is very interesting for translating.

The image shows a screenshot of the BNE Authorities Catalogue interface. At the top, there is a search bar with the text 'Nueva búsqueda' and a 'Cambiar Formato' button. The current format is set to 'Etiquetado'. The main content area displays the entry for 'Ancianos -- Malos tratos'. Under the heading 'Usado por:', there is a list of terms: 'Agresión a los ancianos', 'Ancianos -- Abusos', 'Malos tratos a las personas mayores', 'Maltrato a las personas mayores', and 'Violencia contra los ancianos'. Under 'Término relacionado:', it lists 'Ancianos maltratados'. Under 'Fuentes:', it lists 'LCSH; (Older people-Abuse of)' and 'RAMEAU; (Personnes âgées-Violence envers)'. At the bottom, under 'Otro identificador normalizado:', there is a URL: 'http://id.loc.gov/authorities/subjects/sh85002088 lcsh'.

Figure 1: Term “Abuse of older people”

Dependency (Dependencia): When searching for “dependencia” on Authorities Catalogue of BNE, the result is a list with different meanings for the same word. It is an authorized term but the meaning is not the equivalent according to the context of the text. It is referred to “political dependence”. Therefore, the student applies his critical thinking skill (one of the IL

skills) to find out the right one among the other options. It is has to be said that any term is controlled by their scope. In our example, the meaning of “dependencia” and the right one is “psychological dependence”.

As described above, the student has the link to LCSH to find the equivalent of this term in English and to RAMEAU to find the equivalent of that term in French.

Older people (Personas mayores): This multi-term concept is not on Authorities Catalogue of the BNE. Again, the student has to work his critical thinking to find a controlled and authorized synonym. One of them is “viejos” but non-authorized term, among others. But according to the meaning and scope of the text, there is “ancianos dependientes” which is the proper one. Once again, searching for an equivalent in English the student will find “frail elderly” on LCSH and in French “*personnes âgées dépendantes*” on RAMEAU.



The image shows a screenshot of the BNE Catalogue interface. At the top, there is a header with the text "Catálogo BNE" and a search bar labeled "Nueva búsqueda". Below the search bar, there is a button "Cambiar Formato" and the text "Formato: Etiquetado". The main content area is titled "Ancianos dependientes" and contains the following information:

- Usado por:** Ancianos frágiles
- Término genérico:** Ancianos
- Fuentes:** LCSH; (Frail elderly)
CSIC; (Ancianos dependientes)
RAMEAU; (Personnes âgées dépendantes)
- Otro identificador normalizado:** <http://id.loc.gov/authorities/subjects/sh89004704> lch
<http://lemac.sgcb.mcu.es/Autoridades/LEMAC201213996/concept lemac>
- Nº Registro:** XX550063

At the bottom left, there is a link labeled "Obras".

Figure 2: Term “*Frail elderly*”

As we said above, students can extend their knowledge of the meaning of the text by clicking on the link “Obras”, a link where you can reach the documents that are stored in the Spanish National Library (BNE) and indexed by the subject heading “Ancianos dependientes”. In this example we have retrieved 4 documents, as we can see in the image below.

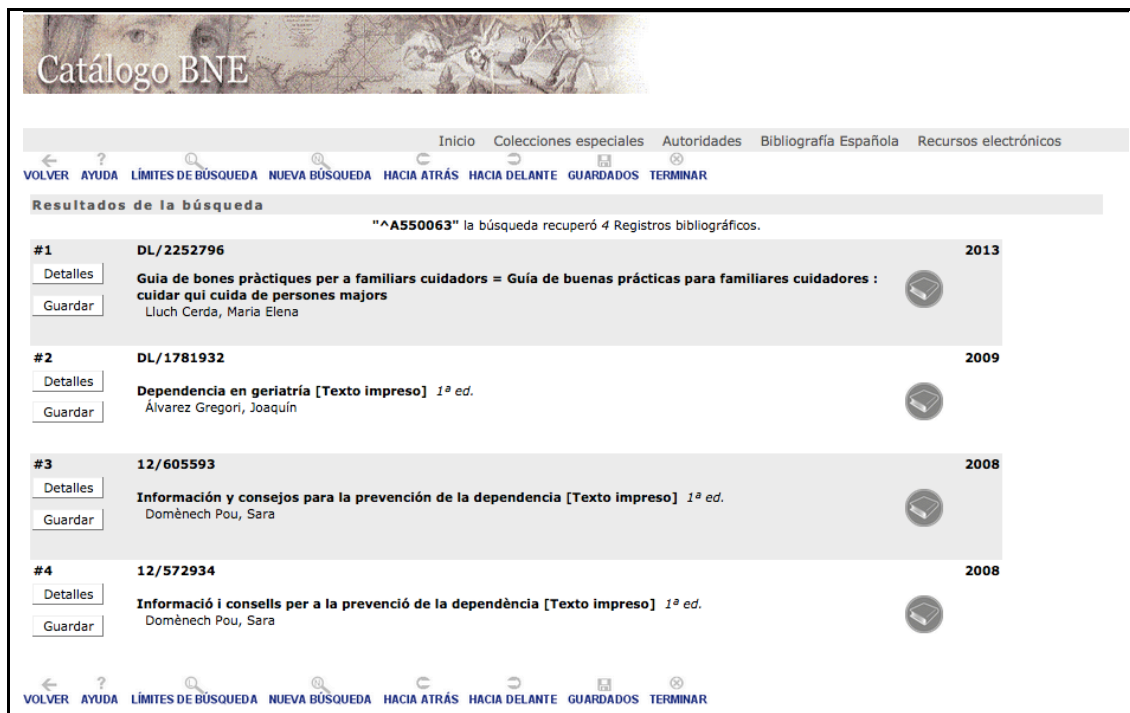


Figure 3: Documents about the term “*Ancianos dependientes*”

2.2 A teaching practice using the UNESCO Thesaurus

The aim of this second practice is to train the students’ skills in using the multilingual UNESCO Thesaurus. We offer to the students a scientific popular text of approximately 500 words (<http://www.agenciasinc.es/Noticias/Los-ninos-que-se-culpan-a-si-mismos-y-a-sus-familias-por-haber-sufrido-abusos-sexuales-tienen-mayores-tasas-de-estres-post-traumatico>). After a quick reading of it, they have to choose three keywords which are supposed to represent the main subjects.

Roughly, the text is about the role of coping strategies and attributions of blame on the psychological adjustment of victims of child sexual abuse. According to that, the three main keywords chosen by the students are: Children, Sexual abuse and post traumatic stress disorder.

Children (Niños): On searching for the term “niños” in plural, the student find out that that term like this does not exist. So, they are trained to apply other flexible abilities and critical thinking skills to search the concept in other way like, for example, trying to find the concept in singular and not in plural. In doing that, students get 22 results related to the concept of children, as we can see in the picture below (Figure 4).

But “niño” is a descriptor too broad and it is expressed in the context of several narrower subjects, such as Addiction, Delinquency, Discrimination, and so on. Again, students have to practice their information literacy skills in trying to find the term “niño” closer to the concept “sexual abuses”. UNESCO Thesaurus, as a controlled vocabulary, offers a list of hierarchical and related concepts of the concept “sexual abuse of children”. Luckily, they find a related concepts (RT) “Child abuse”. Students have to find a term “niño” related to “sexual abuses”.

Each descriptor has a file card where are shown and are provided all the conceptual relationships between descriptors and non-descriptors and the equivalents in English and French as well.

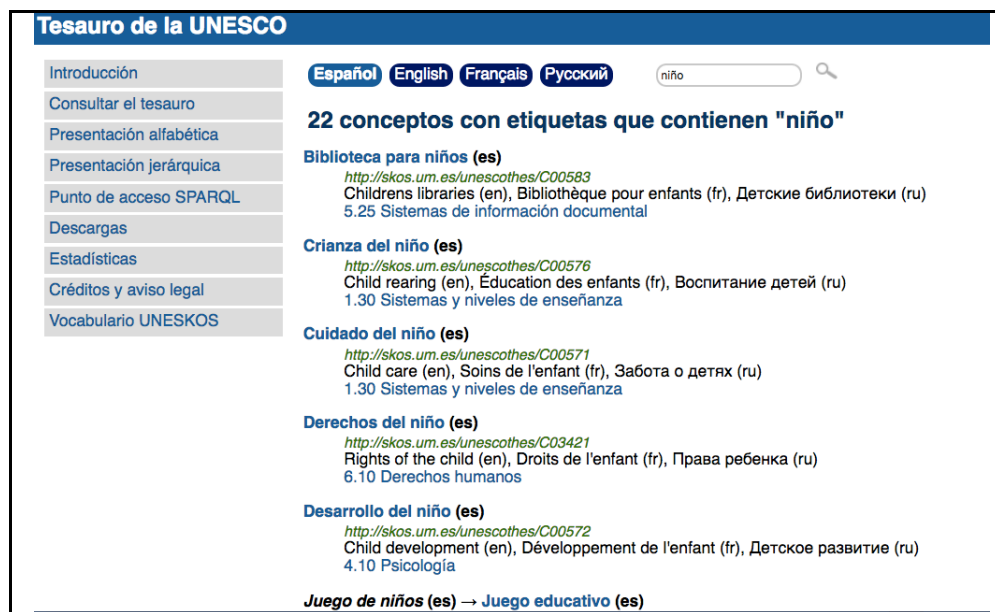


Figure 4: Concept of “children”

Sexual abuses (Abusos sexuales): This term does not exist in plural. So, the student has to work the possibilities that UNESCO Thesaurus offers. In this case, the term “abuso sexual” is in singular. On the other hand, the student can check other terms for “abuso a menores” that are synonyms but non-descriptors. They are represented in the thesaurus with the symbol UP/UF (Usado Por/Used For): “violencia sexual”, “violación”, “acoso sexual”.

Post traumatic stress disorder (Estrés post-traumático): Finally, this descriptor is not represented in UNESCO Thesaurus. In this case, the student, working their information literacy skills, has to choose for another solution. Searching for any other synonym such as “estrés mental”, the student will find out other terms, such as: “tension mentale” in French or “Mental stress” in English. In the same way, he can discover related and broader terms of this concept.

Using the Unesco Thesaurus the student can retrieve all documents stored in the Repository of the UNESCO known as UNESDOC, “a database that contains the full text and bibliographic records of documents and publications published by UNESCO since 1945 as well as bibliographic records of library acquisitions” (Unesco <http://www.unesco.org/ulis/en/faq.html>).

3 Conclusion

In conclusion, this kind of teaching practice is very effective because contributes to students’ acquisition of some information literacy skills for their future as translators. These information literacy skills are:

- Extracting and identifying the main ideas of a text by keywords
- Identifying the necessary information sources for their translating work. In this case, selecting terminological sources such as Subject Headings and Thesauri.
- Getting to know, understanding and managing controlled vocabularies.
- Getting to know how to search for information more effectively.
- Being flexible, creative and resourceful searchers to find data and information.

- Being capable to combine, reformulate, research, infer and experiment new ways of searching on the Web or any database because there is a saying “All the roads leads to Rome”.

References

- Atzori, Luigi, Antonio Iera and Giacomo Morabito. 2010. The Internet of Things: A survey, *Computer Networks*, 4: 2787–2805. DOI: 10.1016/j.comnet.2010.05.010c
- ACRL. 2016. *Information Literacy Competency Standards for Higher Education*. Chicago: American Library Association.
- Catálogo de Autoridades BNE*. 2016. Madrid: Biblioteca Nacional de España.
- Dancette, J. 2015. A context-rich dictionary with a relational structure: A tool for economic translation. *inTRAlinea*. Special Issue: New Insights into Specialised Translation. Retrieved from http://www.intraline.org/specials/article/a_context_rich_dictionary_with_a_relational_structure (september 2016).
- Dancette, J. 2011. L’intégration des relations sémantiques dans les dictionnaires spécialisés multilingues : du corpus ciblé à l’organisation des connaissances, *Meta*, 56 (2):284-300. URI: <http://id.erudit.org/iderudit/1006177ar>. DOI: 10.7202/1006177ar.
- Lin, J., G.C. Murray, B.J. Dorr, J. Hajič, and P. Pecina. 2009. A cost-effective lexical acquisition process for large-scale thesaurus translation. *Language Resources and Evaluation*, 43 (1): 27-40.
- Pastor-Sánchez, Juan-Antonio. 2011. *Tecnologías de la web semántica*. Barcelona: UOC.
- Pastor-Sánchez, Juan-Antonio, Francisco-Javier Martínez-Méndez and José-Vicente Rodríguez-Muñoz. 2009. Advantages of thesaurus representation using the simple knowledge organization system (SKOS) compared with proposed alternatives. *Information research*, 14 (4), paper 422.
- Servicio de Información y Noticias Científicas. 2016. *SINC: La ciencia es noticia*. Ministerio de Economía y Competitividad-FECYT. <http://www.agenciasinc.es/> (September 2016).
- Ramírez Polo, Laura. 2012. Los lenguajes controlados y la documentación técnica: mejorando la traducibilidad. *Revista Tradumática: Tecnologías de la traducción*, 10:192-204. Retrieved from http://ddd.uab.cat/pub/tradumatica/tradumatica_a2012n10/tradumatica_a2012n10p192.pdf (september 2016).
- SKOS-Tesaurus de la UNESCO*. 2016. <http://skos.um.es/unescothes/?l=es> (September 2016).
- Subirats-Coll, Imma. 2013. La Web Semántica y su aplicación en servicios de información : El caso de la Organización de las Naciones Unidas para la Alimentación y la Agricultura. In *Tercer Simposio Nacional De Patrimonio Bibliográfico y Documental*, Bogotá (Colombia), 30 September - 1 October 2013. [Presentation]. Retrieved from <http://eprints.rclis.org/22452/1/simposionacionalbogota-131024043410-phppap01.pdf> (September 2016).
- Summers, E., A. Isaac, C. Redding, C. And D. Krech. 2008. LCSH, SKOS and Linked Data. In *Proc. Int’l Conf. on Dublin Core and Metadata Applications*, pages 25-33. Retrieved from <http://dcpapers.dublincore.org/pubs/article/viewFile/916/912> (september 2016).
- Vrkic, Dina. 2014. Are they a perfect match? Analysis of usage of author suggested keywords, IEEE terms and social tags. In *37th International Convention on ,Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, Opatija (Croatia) 26-30 May 2014, pages 732-737. Doi: 10.1109/MIPRO.2014.6859662.
- Yang, Chan-Jin. 2010. Study on Keywords and Their Use of Academic Theses - Focused on Database Development and Information Link. *Humanities Contents*, 19: 395-416.