Evaluating EcoLexiCAT: a Terminology-Enhanced CAT Tool

Pilar León-Araúz, Arianne Reimerink
University of Granada
Department of Translation and Interpreting, c/Buensuceso, 11 18002 Granada (Spain)
{pleon, arianne}@ugr.es

Abstract

EcoLexiCAT is a web-based tool for the terminology-enhanced translation of specialized environmental texts for the language combination English-Spanish-English. It uses the open source version of the web-based CAT tool MateCat and enriches a source text with information from: (1) EcoLexicon, a multimodal and multilingual terminological knowledge base on the environment (Faber et al., 2014; Faber et al., 2016); (2) BabelNet, an automatically constructed multilingual encyclopedic dictionary and semantic network (Navigli & Ponzetto, 2012); (3) Sketch Engine, the well-known corpus query system (Kilgarriff et al., 2004); (4) IATE, the multilingual glossary of the European Commission; and (4) other external resources (i.e. Wikipedia, Collins, Wordreference, Linguee, etc.) that can also be customized by the user. The tool was built with the aim of integrating terminology management – often considered complex and time-consuming – in the translation workflow of a CAT tool. In this paper, EcoLexiCAT is described along the procedure with which it was evaluated and the results of the evaluation.

Keywords: computer-assisted translation; terminology management; specialized translation

1. Introduction

Currently, computer-assisted tools (CAT) do not seamlessly integrate terminology management – often considered complex and time-consuming – into the translation workflow. Furthermore, most terminological resources do not take into account the real search behavior of translators (Tudhope et al., 2006; Durán Muñoz, 2012: 78). Most terminological modules in CAT tools do not go beyond a simple glossary of source and target terms and access to corpora is rarely, if ever, provided. This leads to an inevitable loss of translation quality and a waste of precious time.

To fill this void, we developed EcoLexiCAT (León-Araúz et al., 2017), a terminology-enhanced CAT tool that provides easy access to domain-specific terminological knowledge in context. This application integrates different features of the professional translation workflow in a stand-alone interface where a source text is interactively enriched with terminological information (i.e. definitions, translations, images, compound terms, corpus access, etc.) from different external resources: (1) EcoLexicon, a multimodal and multilingual terminological knowledge base (TKB) on the environment (Faber et al., 2014: Faber et al., 2016); (2) BabelNet, an automatically constructed multilingual encyclopedic dictionary and semantic network (Navigli and Ponzetto, 2012); (3) Sketch Engine, the well-known corpus query system (Kilgarriff et al., 2004); (4) IATE, the multilingual glossary of the European Commission; and (5) other external resources (i.e. Wikipedia, Collins, Wordreference, Linguee, etc.) that can also be customized by the user.

The motivations behind the integration of these resources lie in the needs and expectations of translators regarding terminology management. According to Durán Muñoz (2012: 82), translators consider that the most important ISO fields (ISO 12630: 1999) in the microstructure of terminological resources are the following: clear and concrete definitions, equivalents, derivatives and compounds, domain specification, examples, phraseological information, definition in both languages for bilingual resources and abbreviations and acronyms.

Likewise, translators believe that terminological resources should be able to (1) permit exportability and/or importability in different formats; (2) include more pragmatic information about usage and tricky translations (old usage, false friends, specific usage in a domain or region, etc.); (3) offer links to other resources to improve or increase results; (4) improve search options; and (5) provide examples taken from real texts. Quite surprisingly, although the translators in this study (Durán Muñoz, 2012) did not show much interest in having access to corpora, they did highlight the need for more phraseological information, pragmatic information and examples taken from real texts. Even though this information can be extracted from corpora, translators were probably reticent to use them because it can take a long time if the right query methods are not provided.

EcoLexiCAT takes into account all of the above and includes the essential fields mentioned, links to other resources and improved search options for corpus analysis that provide the necessary pragmatic information and real text examples. Furthermore, it is a single-platform web-based CAT environment that has the capabilities of importing and exporting different file types and formats. Therefore, the next logical step is to evaluate the functionalities and performance of the tool based on the experience of prospective users in order to assess whether it meets the expectations of professional translators.

The remainder of this paper is organized as follows. Section 2 describes the web-based open source CAT tool MateCat on which EcoLexiCAT is based as well as the external resources used for terminology enhancement. Section 3 provides a description of EcoLexiCAT. Section 4 shows the procedure with which EcoLexiCAT was evaluated and the results of the evaluation. Finally, Section 5 presents the conclusions and future research.

2. EcoLexiCAT sources

2.1 MateCat

MateCat, acronym of Machine Translation Enhanced Computer Assisted Translation, was originally aimed at improving the integration of machine translation (MT)

and human translation (Federico et al., 2014: 129). This application is not only an industrial tool but also an open source platform[1]. The fact that it has an open-source version as well as its high level of flexibility made it a suitable option for the development of EcoLexiCAT. In addition, the features and operation of MateCat are basically the same as those found in most CAT tools used nowadays, such as a text editor that divides the text to be translated in source and target segments and saves them along with their translation in a translation memory (TM). Moreover, it supports 59 different types of source files.

MateCat runs as a web server and communicates with other services through open APIs. It allows communication with pre-existing TMs and the collaborative TM MyMemory, terminological databases, concordance searches within the TMs and machine translation (MT) engines, from which the MT provider, also named MyMemory (a combination of Google Translate and Microsoft Translator), is freely available. Therefore, professional translators will not need to invest much time in learning how to use the tool and will benefit from the interoperability of CAT-related formats (TBX for glossaries, XLIFF for bilingual files, TMX for TMs, etc.). This enables them to use the resources generated during the translation process in other similar tools and reuse pre-existing resources (i.e. glossaries, bilingual files and TMs) in EcoLexiCAT.

2.2 EcoLexicon

EcoLexicon[2] is a multilingual and multimodal terminological knowledge base on environmental science (Faber, León-Araúz & Reimerink 2014; 2016). It is the practical application of Frame-based Terminology (FBT; Faber et al., 2011; Faber, 2012, 2015), a theory of specialized knowledge representation that uses certain aspects of Frame Semantics (Fillmore, 1982; Fillmore & Atkins, 1992) to structure specialized domains and create non-language-specific representations.

EcoLexicon currently has 3,601 concepts and 20,211 terms in Spanish, English, German, French, Modern Greek, and Russian. Regarding the languages included in EcoLexiCAT, EcoLexicon currently contains 5,290 terms in English and 4,898 terms in Spanish. This terminological resource is conceived for language and domain experts as well as for the general public. It targets users such as translators, technical writers, and environmental experts who need to understand specialized environmental concepts with a view to writing and/or translating specialized and semi-specialized texts. The resource contains definitions, semantic networks, equivalents, images, phraseological information and access to corpus information

2.3 BabelNet and Babelfy

The multilingual encyclopedic dictionary and semantic network BabelNet[3] was created by integrating the lexicographic and encyclopedic knowledge contained in WordNet and Wikipedia (Navigli & Ponzetto, 2012: 218). It connects concepts and named entities in a network of semantic relations, made up of about 14 million entries, called Babel synsets. Each Babel synset represents a given meaning and contains all the synonyms expressing that meaning in a range of different languages. Wikipedia and WordNet are integrated through automatic mapping and by filling in lexical gaps in resource-poor languages with MT.

BabelNet is an enormous information resource that can be accessed through an open API, and was considered to be a valuable addition to EcoLexiCAT in those cases where EcoLexicon, a manually-built resource, did not include sufficient information or information regarding general language issues or for texts that combine environmental issues with other domains of expertise. Furthermore, the BabelNet researchers created their own algorithm, called Babelfy, for the disambiguation of polysemic words when found in the context of a particular text (Moro, Raganato & Navigli, 2014; Moro, Cecconi & Navigli, 2014). In EcoLexiCAT, the source text is disambiguated through Babelfy before matching the terms with BabelNet.

2.4 Sketch Engine

Sketch Engine (Kilgarrif et al., 2004) is an online corpus query system with a very efficient search engine and a statistical component for enhanced precision. It contains over 300 corpora in over 60 languages and allows end users to create their own corpora as well. One very interesting module is information extraction through word sketches. Word sketches are summaries of collocational information of a search term, where the term is analyzed according to the verbs, modifiers and other usual constructions that accompany it in real texts. Word sketches are created through sketch grammars that launch specific queries to a corpus. End users can create their own grammars for word sketches and therefore adapt the tool to their specific needs.

Sketch Engine also hosts a set of freely available open corpora that can be queried with full Sketch Engine functionalities with no need of having a subscription. This made a perfect option for EcoLexiCAT, since the English EcoLexicon Corpus was uploaded as an open corpus[4] and can be accessed from EcoLexiCAT through its API.

2.5 IATE

IATE, Inter-Active Terminology for Europe, is the UE's inter-institutional terminology database that has been used in the EU institutions and agencies since 2004, enhancing standardization and promoting an official EU terminology. It has around 8.4 million terms. IATE cannot be accessed through APIs but can be downloaded[5]. Therefore, we downloaded the set of English and Spanish terms and stored them in a database to interact with EcoLexiCAT as a fourth external resource.

3. EcoLexiCAT: a terminology-enhanced CAT tool

When users start a new project in EcoLexiCAT they first access the project settings interface where they can do the following: (1) name the project; (2) choose directionality (so far, English-Spanish or Spanish-English); (3) select a

[1] https://www.matecat.com/open-source/
[2] ecolexicon.ugr.es
[3] babelnet.org
[4] https://the.sketchengine.co.uk/open/
[5] http://iate.europa.eu/tbxPageDownload.do

particular domain within the environment – these are in consonance with the domains according to which EcoLexicon is organized and are included in this first step as a way to classify projects and TMs for later reuse; (4) choose between general and patent segmentation rules, for the source text to be segmented accordingly; (5) optionally add an MT provider for post-editing – MyMemory is freely available, but others (e.g. Moses, DeepLingo, IP Translator) can also be added if users have an account with them; (6) optionally add users' own TMs and/or glossaries – otherwise a collective TM stored in the system will be used; and (7) upload the source text. These steps, except for (3), are default options in MateCat.

Once the source text is processed and converted into a bilingual format (XLIFF), users can access the main interface (Figure 1), which is divided into two main sections. The left-hand section is where the four external resources (i.e. EcoLexicon, BabelNet/Babelfy, Sketch Engine and IATE) provide the terminological enhancement of the translation process. The right-hand section is where the target text is produced, an editor where the source text appears split into different segments.

In the right upper part of the editor, users may download the target or the source text in their original format, and export the bilingual file in SDLXLIFF (SDL Trados Studio's native format) or the whole project in OmegaT's native format, another desktop open source CAT tool. This, together with the possibility of downloading the TM and the glossary created during the project, ensures the interoperability of different formats across different CAT tools, an issue that professional translators must often deal with.

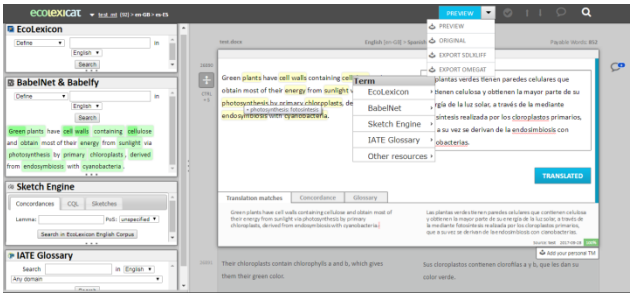

Figure 1: User's interface of EcoLexiCAT.

Figure 2 shows a segment within the editor, which offers the usual editing features of any CAT tool, and how the terms from EcoLexicon are recognized (highlighted in yellow). Users can also split or merge segments, copy the source text in the target segment, benefit from a QA (quality assurance) system that detects missing spaces or tags, create on-the-fly glossary entries, search for concordances within the TM and get suggestions from previously stored segments in the TM or, if added, from an MT engine. Once a segment is confirmed, it is stored in the users' TM or in the collaborative TM, from which other users can benefit. This converts the tool into a collaborative environment. Of course, translators working under confidentiality agreements should always choose their own TM, which will prevent their segments from getting stored in the collaborative TM.

However, the difference between an ordinary CAT tool and EcoLexiCAT is that EcoLexiCAT is a terminology-enhanced translation tool. This means that the editor interacts with external terminological resources that can assist the translator during the different phases of the translation workflow. First of all, the source segment is enriched with information from EcoLexicon. As shown in Figure 2, all matching terms are highlighted in yellow, and users can interact with them in three ways: (1) if they hover the mouse over them, all possible translations (equivalent terms and synonyms) are displayed in an emerging box; (2) if they click on any of them, the EcoLexicon box of the left-hand side shows both the translations and the definition; and (3) if they right-click on any of them, a scroll-down menu gives access to all the different options provided by each of the resources of the left-hand section, sending ready-made queries based on the selected term.

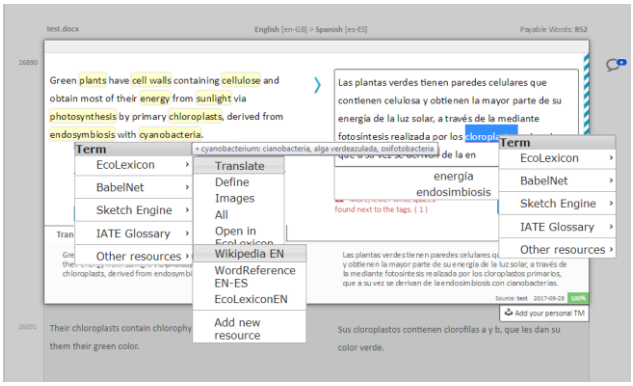

Figure 2: EcoLexiCAT editor.

In the case of EcoLexicon, these options correspond to the data categories in the TKB that usually serve for text comprehension: translations, synonyms, definitions, and images. Also from this menu, a new tab can be opened in the browser to access the EcoLexicon TKB for a more detailed analysis of the conceptual networks.

In the BabelNet & Babelfy box the source text, previously disambiguated by the Babelfy algorithm, is also matched against the BabelNet network. This enables the system to propose statistically relevant candidate translations, which is a significant advantage taking into account that BabelNet covers any specialized or general domain and ambiguity can be frequently encountered.

All matched terms are highlighted in green and behave in the same manner as the terms in the source segment with regard to EcoLexicon (hovering the mouse, clicking and right-clicking). The options available from BabelNet, after right-clicking, correspond to the data categories that have been considered most interesting for translators: definitions, translations, compound words and images. Also, from the definitions option, a new tab can be opened in the browser to access the semantic networks in BabelNet.

The Sketch Engine box can be used to select a term from both the source and target segments and analyze its behavior in the EcoLexicon corpus. So far, only the EcoLexicon English Corpus is hosted in Sketch Engine Open Corpora. The EcoLexicon Spanish Corpus is still in the compilation phase. The corpus can be queried through basic or CQL queries as well as through word sketches. Word sketches are based on the default sketch grammars offered by Sketch Engine but also on a set of customized sketch grammars specifically built for the extraction of semantic relations (León-Araúz et al., 2016).

Corpus information can be very useful during the text production phase (e.g. searching for modifiers or verbs that collocate with a particular noun, looking for synonyms or frequent syntactic structures, etc.). However, corpora can also help translators to understand how concepts interrelate with each other within the domain (hence, the customized semantic word sketches).

For this reason, corpus queries are enabled from both source and target segments. The output of the queries can be opened in a new tab that sends users to the website of Sketch Engine Open Corpora for a more detailed analysis. In this way, they can use all the functionalities of the tool (e.g. Context, Word list, Thesaurus, Sketch Diff, etc.) and make more specific queries filtered by the features according to which the corpus is tagged (i.e. year, genre, contextual domain, user type and linguistic variant).

The IATE box gives access to the information that the downloadable dump provides (equivalents and domains). The queries can be limited by domain and the system delivers all possible equivalents in both directionalities.

In turn, the target segment is enriched with a predictive typing feature. As soon as users start typing a word that has been matched as the translation of one of the terms in the source segment, all possible translations are shown in a drop-down list. In addition, as in the source segment, users can right-click on any term they type in the target segment and send queries to the four resources in the opposite language directionality.

Finally, from the scroll-down menu that can be accessed from both source and target segments, a last option is provided to interact with a customized list of other external resources that can be expanded by the user. This appears as an emerging window based on the URLs of the resources that users may usually consult, such as WordReference, Wikipedia, Linguee, etc. This works as the SDL Trados Studio plug-in Web Lookup or the MemoQ web search feature.

4. Evaluating EcoLexiCAT

4.1 Experimental setup

EcoLexiCAT was evaluated by comparing the behavior and products of two subject groups, one using EcoLexiCAT and the other one acting as a control group and using MateCat. Both groups were made up of students from the Master's Degree in Professional Translation of the Faculty of Translation and Interpreting of the University of Granada (Spain).

Prior to the translation task, participants of both groups were asked to fill out a brief questionnaire in order to collect data about their professional/training background, their expectations of terminological resources and CAT tools and their habits regarding the use of dictionaries, corpora, terminological resources, etc. when confronted with a translation assignment.

Then, the group using EcoLexiCAT was made acquainted with the tool, so that subjects did not waste their time getting used to its specific functionalities. Both groups were already familiar with MateCat. For this experiment, the collaborative memory option nor machine translation were used.

Afterwards, subjects were presented with a translation task consisting of two short specialized translation assignments, one English-Spanish and the other Spanish-English, of extracts of scientific papers on the topic of coastal engineering, a domain widely covered in EcoLexicon. The reason for having chosen both directionalities is, firstly, to see if behavior and results change according to directionality; secondly, because the only corpus available so far is the EcoLexicon English Corpus and usage examples are usually demanded during the text production phase.

Students were required to deliver publishable texts in two hours. Therefore, the length of each source text was under 200 words (EN-SP 194 and SP-EN 168 words). Other features of the source texts were high term density, syntactically complex sentences and collocational specificities that called for both a deep understanding of domain knowledge and expression. Students were thus confronted with several challenges both during the comprehension and production phases of the translation workflow.

Both groups were asked to note down all the problems encountered and the resources that helped them solve each problem. The EcoLexiCAT group was allowed to use resources outside EcoLexiCAT only if they did not find the answer within the tool.

Finally, after finishing the assignments, EcoLexiCAT users filled out another questionnaire on the tool's usability, functionality and efficiency, three main parameters established by the ISO 9128 standard for software product evaluation. They were also asked to highlight the problems encountered and propose possible improvements.

The purpose of this evaluation was two-fold. We intended to assess user satisfaction but also user performance. The first parameter was assessed based on the answers given by the EcoLexiCAT user group in the last questionnaire, whereas the second parameter was assessed by comparing the time employed and the average quality of the target texts delivered by both groups. Quality assessment was based on a scale where both translation and linguistic errors and wise choices are accounted for. The editing log of EcoLexiCAT and MateCat was used to see how long students took to translate each text.

4.2 Results and discussion

A total of 19 students, aged from 22 to 37, were included in the evaluation, 10 in the EcoLexiCAT group and 9 in the control group. All subjects except for one were native speakers of Spanish, 11 subjects have English as their first foreign language and 5 as their second foreign language. One subject is a native speaker of both English and Spanish and 2 did not include English as one of their official working languages during their undergraduate degree, but did have sufficient proficiency. The large majority has a translation degree (84%), the others have degrees in modern languages or related areas. Only four subjects mentioned previous professional translation experience. The different characteristics were evenly divided over both experimental groups.

4.2.1 User expectations

In the first questionnaire, the participants were asked to classify the following features in CAT tools as essential, desirable or unnecessary: access to MT engines, access to corpora, interoperable file formats, access to terminological resources, access to terminological resources defined by users, and QA and revise options. The results in Figure 3 show that, according to the subjects, the most important features are format interoperability, terminological resources and QA revise options. Access to corpora was regarded essential and desirable in the same proportion, whereas access to MT engines was deemed only desirable. This might be due to the fact that post-editing of MT is still not widely accepted by the translators' community. MT engines, access to corpora and user-defined terminological resources were the only features that were unnecessary according to a few participants.

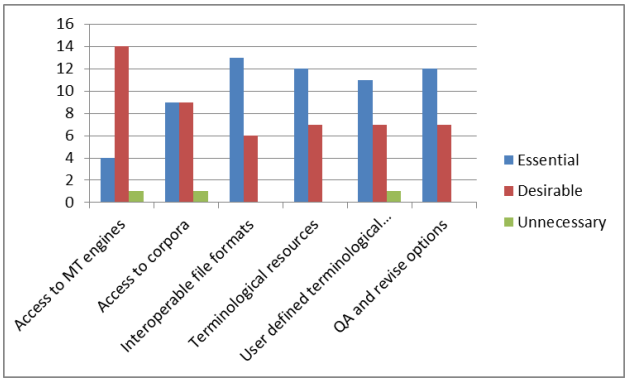

Figure 3: Users's expectations about CAT tools.

When asked about other features not included in the above list, most students could not identify any other feature that they would consider relevant in CAT tools. Exceptions were image editors and customizable QA rules.

Participants were also asked to do the same with a set of data categories usually included in terminological resources. The data categories were: definitions, translations, synonyms and variants, context and usage examples, conceptual relations, register, images, phraseological and collocational information, etymology, pronunciation, compounds and derivatives, part of speech, pragmatic information on term usage, and access to corpora.

The results in Figure 4 show that definitions, translations, context and usage examples, and access to corpora are the most relevant data categories. Among desirable categories, conceptual relations, register, images and compounds and derivatives stand out. Etymology and pronunciation are the categories most often regarded as unnecessary.

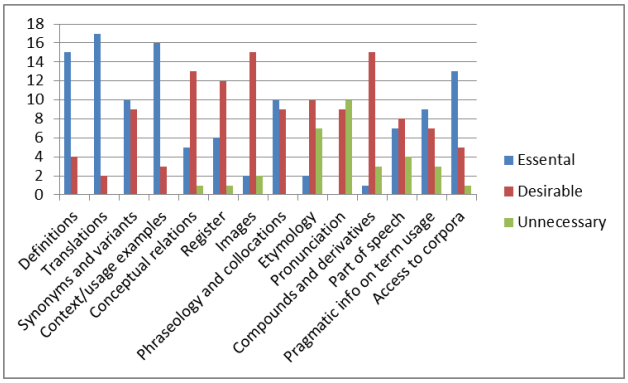

Figure 4: Users's expectations about terminological resources.

When asked about other features not included in the above list, most students could not identify any other feature that they would deem relevant in terminological resources. Exceptions were specialized reference works and term use frequency. Among the resources that students use the most for their translation assignments, the following stand out: Wordreference, Linguee, Reverso Context, IATE, Merriam-Webster, Oxford dictionaries, Collins, esTenTen and enTenTen corpora in Sketch Engine, the BNC, CREA, the web as a corpus, Pons and Termium Plus.

The analysis of subjects' answers indicates that EcoLexiCAT meets most of users' needs and expectations, but it also points to how to improve the tool and even EcoLexicon. For instance, currently there is a phraseology module under construction in EcoLexicon that will undoubtedly be linked to EcoLexiCAT in the future.

4.2.2 User performance

All target texts were evaluated by one reviser to ensure the same criteria were applied in all cases. To assess the quality of the target texts of both groups, ten translation problems were identified for both the English-Spanish and the Spanish-English assignment. The problems identified were based on those that the students mentioned repeatedly and on the reviser's expertise in the text type and domain. Depending on how well the students solved these problems they could obtain up to 10 translation points. On the other hand, the language errors in both Spanish and English were deducted from a maximum grade of 10. The final grade was then the average between the translation points obtained and the linguistic quality of the target text (see Table 1).

For example, one translation problem of the English-Spanish assignment was finding the correct terminological equivalent in Spanish for the different types of current (longshore, tidal and rip current). Another problem was understanding the exact location of a groyne in "perpendicular or slightly oblique to the shoreline

extending into the surf zone (generally slightly beyond the low water line)". An example of a translation problem in the Spanish-English assignment was understanding that "bocana" and "desembocadura" are synonyms and can both be translated as "river mouth".

The EcoLexiCAT group outperformed the control group in both directionalities, although only slightly in the SP-EN assignment. The average quality of the target texts of both groups was not very high. This is understandable because most subjects of both groups did not have any professional translation experience or previous knowledge about the environmental domain. The results are promising, though, as EcoLexiCAT helps to obtain a better target text in less time.

	EcoLexiCAT EN-SP	Control EN-SP	EcoLexiCAT SP-EN	Control SP-EN
	6.8	6.1	6.9	7.1
	4.5	3.9	7.5	6
	8.4	7.6	8.1	6.6
	7.4	3.9	6.4	5
	8.0	4	7.1	4.6
	7.8	4.8	5.4	6.5
	7.4	6.3	6.1	7.3
	6.5	6.5	4.5	6.6
	7.6	6.1	7.6	6.1
	5.3		4.5	
Average	6.9	5.5	6.4	6.2

Table 1: User's interface of EcoLexiCAT.

It is also interesting to point out that the control group used very similar resources to solve the translation problems to those included in EcoLexiCAT: EcoLexicon, BabelNet, Wordreference, IATE, Linguee, and Wikipedia.

In terms of the time invested, in both directionalities the EcoLexiCAT group outperformed the control group. Suprisingly, the EcoLexiCAT group took longer in the Spanish-English assignment than in the English-Spanish one, whereas the control group took longer in the Spanish-English assignment, which is striking because even though it was a shorter source text, the assignment implied translating into a non-mother tongue of most of the participants.

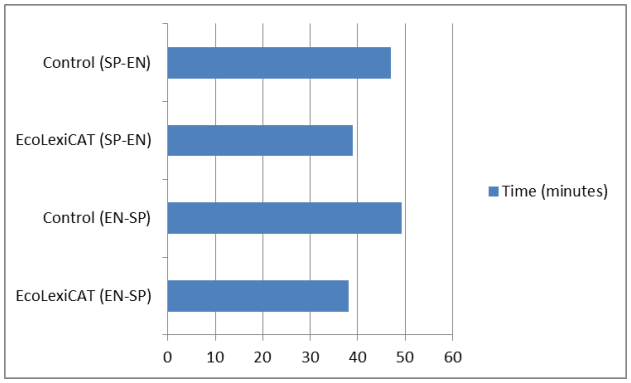

Figure 5: Time invested during SP-EN and EN-SP assignments by the EcoLexiCAT and control groups.

4.2.3 User satisfaction

Generally speaking, the subjects belonging to the EcoLexiCAT group believed that the tool is very useful (60%) or useful (40%). No subjects answered "not very useful" or "useless" when asked about the general usefulness of the tool for the translation of environmental texts.

The parameters of functionality, usability and efficiency were evaluated based on the rating of different items in a 1-to-5 scale, where 1 was the lowest and 5 the highest rate.

Regarding functionality (Figure 6), the subjects were asked whether the tool contained suitable features for: (1) the translation of environmental texts (80% answered 4 and 20% 5); (2) the comprehension phase of an environmental text (80% answered 4 and 20% 3); and (3) the production of an environmental text (50% answered 4, 40% 3 and 10% 2). This implies that EcoLexiCAT is more comprehension-oriented and that future improvements should head for the assistance in production-oriented tasks.

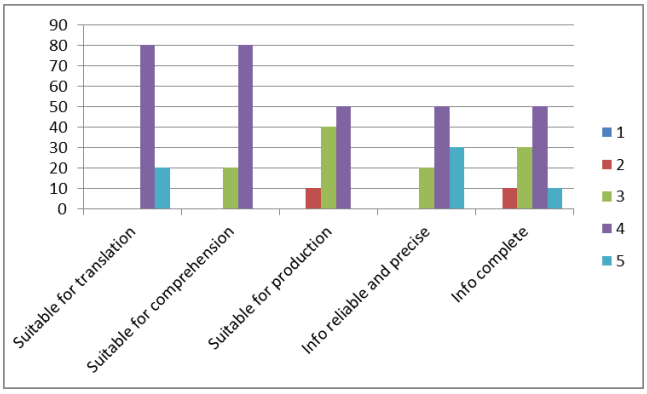

Figure 6: Functionality of EcoLexiCAT.

After that, they were consulted on the type of information provided and the usefulness of external resources. They were asked whether the information provided was: (1) reliable and precise (50% answered 4, 30% 5 and 20% 3); and (2) complete (50% answered 4, 30% 3, 10% 2 and 5).

When asked to rate the usefulness of external resources during their assignments (Figure 7), EcoLexicon, Linguee and Sketch Engine seem to be the best rated.

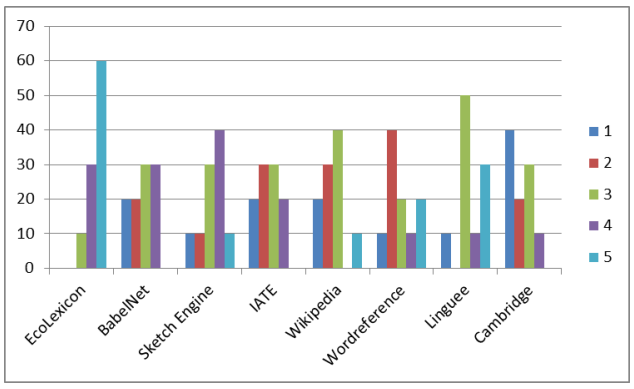

Figure 7: Usefulness of external resources.

As for usability (Figure 8), subjects were asked whether EcoLexiCAT: (1) was intuitive and easy to use (60% answered 4 and 40% 5); (2) had a functional design (50% answered 4, 40% 5 and 10% 3; and (3) provided an adequate interaction with the layout (40% answered 4; 30% 3, 20% 5 and 10% 2).

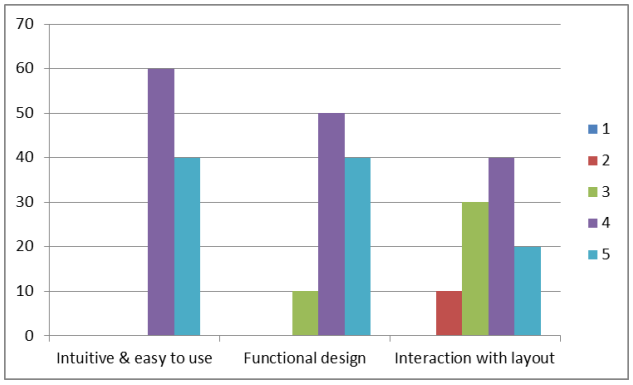

Figure 8: Usability of EcoLexiCAT.

Finally, efficiency (Figure 9) was assessed based on whether the information was loaded at the right speed (60% answered 4 and 40% 5) and fluency: (1) user interaction with the editor (50% answered 4, 20% 5, 20% 3 and 10% 2); (2) interaction of the editor with external resources (50% answered 4, 30% 3 and 20% 5); (3) user interaction with external resources (50% answered 3, 40% 4 and 10% 5).

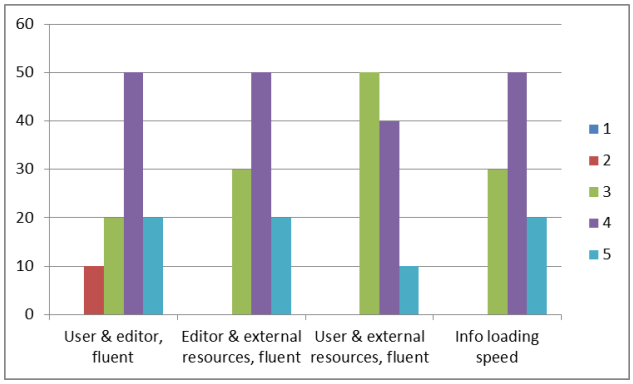

Figure 9: Efficiency of EcoLexiCAT.

The three parameters point to a favorable evaluation of EcoLexiCAT, although efficiency seems to be the one to be first improved in the future.

5. Conclusions

Although it is difficult to generalize de results obtained from the experiment described above, they do indicate that integrating terminology management in the translation workflow in a stand-alone interface improves the quality of the translation and reduces the time spent on the task. EcoLexiCAT users considered the tool to be useful and were satisfied with its functionality, usability and, although slightly less, efficiency.

In the near future, improvements will be made to the tool to make the interaction between the user and the tool and its external sources more fluent.

As the results also indicate that EcoLexiCAT was considered more useful for the comprehension phase than for the production phase, we plan to add other sources to improve on the latter. For example, Reverso Context will be added to the external sources. In addition, we are currently designing and implementing a phraseology module for EcoLexicon, which will be linked to EcoLexiCAT and will undoubtedly be very helpful for the production phase of the translation workflow.

Furthermore, EcoLexicon is currently being linked to other encyclopedic (i.e. DBpedia) and environmental resources (i.e. GEMET, AGROVOC) by means of Linked Data. Once the TKB is fully integrated into the Linguistic Linked Open Data, EcoLexiCAT will also benefit from reliably disambiguated encyclopedic and specialized term entries.

We also plan to carry out further evaluation experiments with professional translators and the machine translation option.

Finally, when all new features are included in the tool, EcoLexiCAT will be made freely available for any user interested in translating English or Spanish environmental texts. Users will only need to register and indicate their educational background, translation experience and the purpose for which they will be using the tool. This will help us analyze user profiles and behaviour when interacting with the tool. Moreover, it will allow us to classify the resources generated (i.e. TMs), which can be used as a parallel corpus, thus enriching both the tool and the EcoLexicon Corpus.

6. Acknowledgements

The authors would like to thank the master students who participated in the evaluation of EcoLexiCAT.
This research was carried out as part of projects FF2014-52740-P, Cognitive and Neurological Bases for Terminology-enhanced Translation (CONTENT), and FFI2017-89127-P, Translation-oriented Terminology Tools for Environmental Texts (TOTEM), funded by the Spanish Ministry of Economy and Competitiveness.

7. Bibliographical References

Durán Muñoz, I. (2010). Specialized lexicographical resources: a survey of translators' needs. In S. Granger & M. Paquot (eds) (2010). *eLexicography in the 21st century: New Challenges, new applications. Proceedings of ELEX2009*. Cahiers du Cental. Vol. 7. Louvain-La-Neuve: Presses Universitaires de Louvain, pp. 55–66.

Faber, P., León-Araúz, P. & Reimerink, A. (2011). Knowledge representation in EcoLexicon. In N. Talaván, E. Martín Monje & F. Palazón (eds.) *Technological Innovation in the Teaching and Processing of LSPs: Proceedings of TISLID*, 10. Madrid: Universidad Nacional de Educación a Distancia, pp 367–385.

Faber, P. (ed.) (2012). *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin/New York: Mouton de Gruyter.

Faber, P. (2015) Frames as a framework for terminology. In H. J. Kockaert & F. Steurs (eds.) *Handbook of*

Terminology, 1. John Benjamins Publishing Company, pp. 14–33.

Faber, P., León-Araúz, P. & Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In *Languages for Specific Purposes in the Digital Era. Educational Linguistics*, 19. Springer, pp 267–301.

Faber, P., León-Araúz, P. & Reimerink, A. (2016) EcoLexicon: new features and challenges. In I. Kernerman, I. Kosem Trojina, S. Krek, & L. Trap-Jensen (eds.) *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in conjunction with the 10th edition of the Language Resources and Evaluation Conference*, Portorož, pp. 73–80.

Fillmore, C. J. (1982). Frame Semantics. In The Linguistic Society of Korea (ed.) *Linguistics in the Morning Calm*. Seoul: Hanshin, pp. 111–137.

Fillmore, C. J. & Atkins, B. T. S. (1992). Toward a Frame-based Lexicon: The Semantics of RISK and Its Neighbors. In A. Lehrer & E. Kittay (eds.) *Frames, Fields and Contrasts: New Essays in Semantic and Lexical Organization*. Hillsdale NJ: Erlbaum, pp. 75–102.

Kilgarriff, A, Rychly, P., Smrz, P. & Tugwell, D. (2004). The Sketch Engine. In *Proceedings of the 11th EURALEX International Congress*. Lorient: EURALEX, pp. 105–116.

León-Araúz, P., Reimerink, A. & Faber, P. (2017) EcoLexiCAT: a Terminology-enhanced Translation Tool for Texts on the Environment. In *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, edited by Kosem, I., Kallas, J., Tiberius, C., Krek, S., Jakubíček, M. & Baisa, V., pages 321-341. Brno: Lexical Computing CZ s.r.o.

León-Araúz, P., San Martín, A. & Faber, P. (2016) Pattern-based Word Sketches for the Extraction of Semantic Relations. In *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016)*, pages 73-82. Osaka, Japan: COLING 2016

Navigli, R. & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193, pp. 217–250.

Tudhope D., Koch T. & Heery R. (2006). Terminology Services and Technology: JISC state of the art review. Available at: http://www.ukoln.ac.uk/terminology/JISC-review2006.html

8. Language Resource References

LexiCon Research Group. (2017). EcoLexicon. Available at http/:/ecolexicon.ugr.es.

LexiCon Research Group. (2017). EcoLexicon English Corpus. Available at https://the.sketchengine.co.uk/open/