Assessing EcoLexiCAT: Terminology Enhancement and Post-editing

Pilar León-Araúz, Arianne Reimerink, Pamela Faber

Department of Translation and Interpreting
University of Granada
E-mail: {pleon, arianne, pfaber}@ugr.es

Abstract

EcoLexiCAT is a freely available online application, which integrates all 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. EcoLexiCAT is powered by MateCat and the external sources include EcoLexicon, BabelNet, the EcoLexicon English Corpus (powered by Sketch Engine) and IATE, as well as other common resources (e.g. Wordreference, Wikipedia, Linguee, etc.). Machine translation (MT) can also be optionally added. In order to evaluate the functionalities and performance of the tool, two experiments were carried out. In the first, one subject group used EcoLexiCAT and the other used MateCat, acting as the control group. In the second, both subject groups used EcoLexiCAT and only one used MT. Both experiments shed interesting light on user behaviour, performance and satisfaction while using EcoLexiCAT.

Keywords: EcoLexiCAT; CAT tools; terminology management; MT post-editing

1. Introduction: EcoLexiCAT

Today, machine translation (MT) and computer-assisted translation (CAT) are a crucial part of the professional translation workflow. Nevertheless, the post-editing of MT output has only recently started to become more widely accepted, and terminology management is often not seamlessly integrated into the translation process. As a possible solution to this problem in the field of environmental translation we developed EcoLexiCAT, a terminology-enhanced CAT tool that provides easy access to domain-specific terminological knowledge in context and MT (León-Araúz, Reimerink & Faber, 2017; León-Araúz & Reimerink, 2018; León-Araúz, Reimerink & Faber, 2019).

The integration of MT post-editing and terminology enhancement in a CAT environment constitutes the core of what has recently been termed "augmented translation" (De Palm & Lommel, 2017; Lommel, 2018, 2017). Augmented translation is a technological approach that leverages various technologies to support and augment translators' mental processes while translating. Such technologies include translation memories, terminology management, adaptive machine translation, and automatic content enrichment (ACE). EcoLexiCAT can thus be regarded as an augmented

translation system for the environmental domain, since it combines to a certain extent all of the above, especially in terms of ACE, which is the whole idea underlying terminology enhancement. Similar approaches can be found in TaaS[1] (Terminology as a Service), SCATE (Smart-Computer-Aided Translation Environment) and the Ocelot plug-in developed in the project FREME[2].

EcoLexiCAT is freely available for any user interested in translating English or Spanish environmental texts[3]. It integrates all 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, León-Araúz & Reimerink, 2014, 2016); (2) BabelNet, an automatically constructed multilingual encyclopaedic dictionary and semantic network (Navigli & Ponzetto, 2012); (3) the EcoLexicon English Corpus (EEC), powered by Sketch Engine, the well-known corpus query system (Kilgarriff et al., 2004); (4) IATE, the multilingual terminological database of the European Union; and (5) other external resources that can be customized by users (i.e. Wikipedia, Wordreference, Linguee, etc.).

EcoLexiCAT is powered by MateCat[4], which runs as a web server and communicates with other services through open APIs. It allows communication with pre-existing TMs, terminological databases, concordance searches within the TMs and machine translation (MT) engines, from which the MT provider MyMemory (a combination of Google Translate and Microsoft Translator) is freely available[5].

The main interface (Figure 1) 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 (text comprehension). The right-hand section, which is where the target text is produced, is an editor where the source text appears split into different segments (text production).

Figure 2 shows a segment within the editor. First of all, the source segment is enriched with information from EcoLexicon. This is done by lemmatizing all the words in the segment and matching them against the term entries in the TKB.

[1] http://www.taas-project.eu/

[2] http://www.freme-project.eu/

[3] Temporarily hosted at http://manila.ugr.es:9966

[4] https://www.matecat.com/open-source

[5] https://www.matecat.com/support/managing-language-resources/machine-translation-engines/

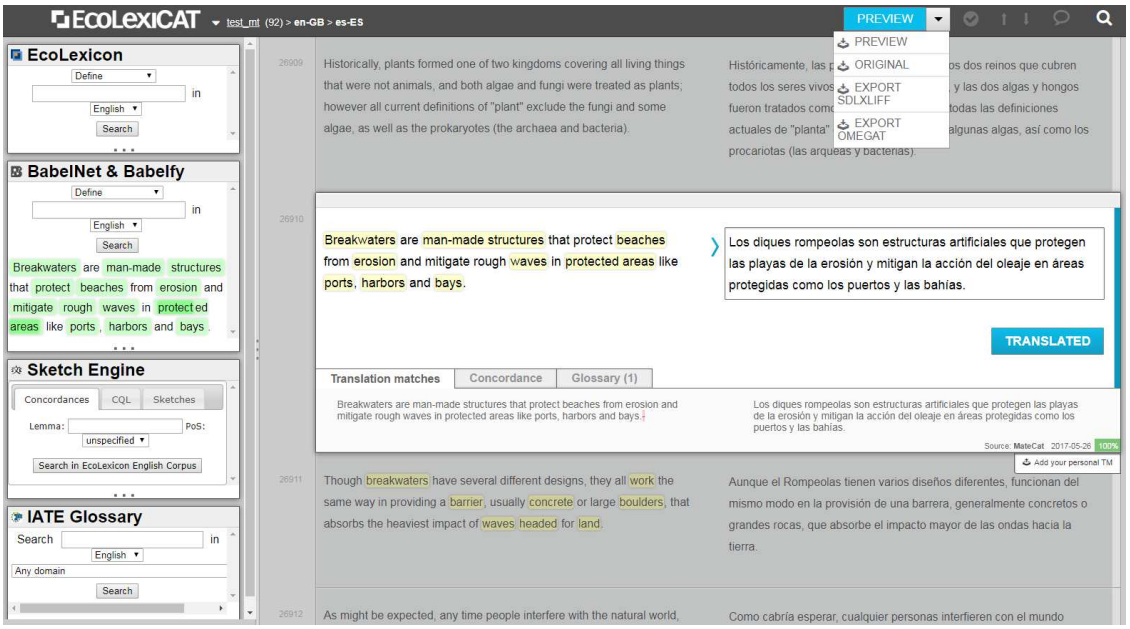

Figure 1: Main user interface of EcoLexiCAT.

All matching terms are highlighted in yellow. In the BabelNet box, the source text is matched against the contents of the KB. After applying the Babelfy algorithm for disambiguation, matches are marked in green. If users 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. In the case of EcoLexicon, these options correspond to the data categories in the TKB that are useful for text comprehension: translations, synonyms, definitions, semantic relations and images. The data categories of BabelNet included in EcoLexiCAT are definitions, translations, compound words, semantic relations, and images.

In the Sketch Engine box, the behaviour of a term selected in the source or target segments can be analysed in the EcoLexicon English Corpus (EEC; León-Araúz et al., 2018) hosted in Sketch Engine Open Corpora. Three different query modes are provided: lemma-based concordances, word sketches, and CQL (Corpus Query Language). In the IATE box, the set of English and Spanish terms downloaded from the database interacts with EcoLexiCAT as a fourth external resource.

Finally, other common language resources (e.g. Wikipedia, Wordreference, Linguee, etc.) are integrated as a pop-up box right under the active segment. Their results are shown as they appear online, since these resources are integrated as embedded websites.

In turn, the target segment is enriched with a predictive typing feature based on the matches from EcoLexicon. In addition, as in the source segment, users can right-click on any term typed in the target segment and send queries to all resources in the

opposite language directionality[6].

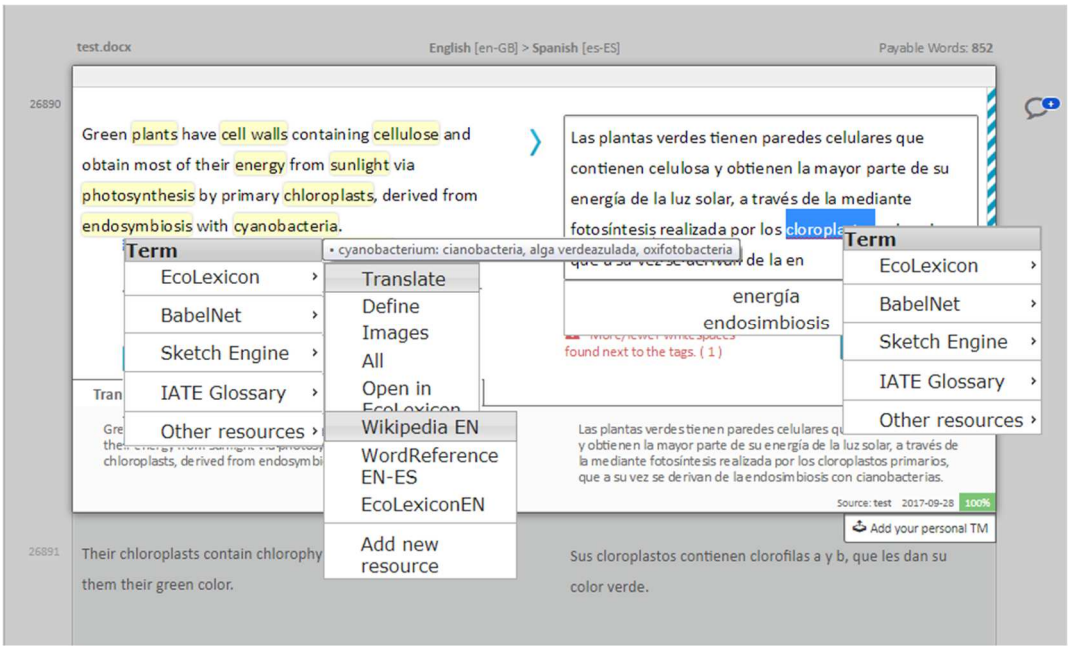

Figure 2: EcoLexiCAT editor.

After designing, creating and testing EcoLexiCAT, the next logical step was 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 translators.

In the remainder of this paper, we present the experimental setup (Section 2) and the results of two experiments carried out to evaluate the tool, focusing on user expectations (Section 3), user behaviour (Section 4), user performance (Section 5) and user satisfaction (Section 6). In the first experiment (León-Araúz, Reimerink & Faber, 2019), one subject group used EcoLexiCAT and the other used MateCat, acting as the control group. In the second, both subject groups used EcoLexiCAT, but only one used MT. Accordingly, in the first experiment we studied the benefits of terminology enhancement, whereas in the second we focused on the benefits of MT post-editing. Finally, Section 7 presents the conclusions derived from this research.

2. Experimental setup

EcoLexiCAT was evaluated in two experiments conducted one year apart. This means that during the second experiment the tool had already been improved based on the results of the first.

[6] For a more detailed account of the functioning of EcoLexiCAT, consult León-Araúz, Reimerink & Faber (2017), León-Araúz & Reimerink (2018) and León-Araúz, Reimerink & Faber (2019).

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.

The subject groups of the first experiment (EcoLexiCAT translators vs. MateCat translators) were students from the master's degree in Professional Translation of the Faculty of Translation and Interpreting of the University of Granada (Spain). In contrast, the subject groups of the second experiment (EcoLexiCAT translators vs. EcoLexiCAT post-editors) were students from both the master's degree and the final year of the Undergraduate Programme in Translation of the same faculty.

In the first experiment a total of 19 students, 22 to 37 years of age, were included in the evaluation: 10 EcoLexiCAT translators and nine MateCat translators. All subjects except for one were native speakers of Spanish; 11 subjects had English as their first foreign language, and five as their second foreign language. One subject was a native speaker of both English and Spanish, and two did not include English as one of their official working languages during their undergraduate degree, but had sufficient proficiency. The majority had a translation degree (84%); the others had degrees in modern languages or related areas. Only four subjects mentioned previous professional translation experience.

In the second experiment a total of 20 students, 20 to 54 years of age, participated in the evaluation: 10 EcoLexiCAT translators and 10 EcoLexiCAT post-editors. All subjects were native speakers of Spanish, 16 subjects had English as their first foreign language, and four as their second foreign language. Among the master's students, 90% had a translation degree and 70% had previous professional translation experience. In both experiments these characteristics were evenly divided over both groups.

In both experiments the subjects were presented with the same translation task. It consisted of two short, specialized translation assignments, one English-Spanish (EN-ES) and the other Spanish-English (ES-EN). The texts were extracts of scientific papers on the topic of Coastal Engineering, a domain widely covered in EcoLexicon. The reason for having chosen both directionalities was first to see whether behaviour and results varied according to directionality, and second, because the only corpus available so far is the EEC and usage examples are usually requested during the text production phase.

Subjects were required to deliver publishable texts in two hours. Therefore, the length of each source text was less than 200 words (EN-ES 194 and ES-EN 168 words). Other features of the source texts were high term density, syntactically complex sentences and collocational specificities that called for a deep understanding of both domain knowledge and written expression. Subjects were thus confronted with various

challenges during the comprehension and production phases of the translation workflow.

Moreover, in the two experiments both groups were asked to list all the problems encountered and the resources that helped them solve each problem. EcoLexiCAT translators and post-editors were allowed to use resources other than those in EcoLexiCAT only if they did not find the answer within the tool.

Finally, after finishing the assignments, EcoLexiCAT users filled out another anonymous questionnaire on the tool's usability, functionality and efficiency, which are three parameters established by the ISO 9126 (2001) standard for software product evaluation. They were also asked to highlight any issues related to the functioning of the tool and to propose possible improvements.

Apart from discovering the expectations of our prospective users, the purpose of this evaluation was threefold. We were not only able to assess user satisfaction but also user behaviour and performance. The first parameter was assessed based on the answers given by EcoLexiCAT translators and post-editors in the last questionnaire. The second parameter was based on the analysis of the subjects' behaviour according to Google Analytics. The third parameter was assessed by comparing the time employed and the average quality of the target texts delivered by all groups. Quality assessment was based on a scale where both translation and linguistic errors and accurate choices were accounted for. The editing logs of EcoLexiCAT and MateCat were used to see how long subjects took to translate each text.

3. 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 revision options. The results in Figure 3 for both experiments show that the most important features were found to be format interoperability, terminological resources, and QA and revision options. Access to corpora was regarded as slightly more essential than desirable, whereas access to MT engines was only desirable. This might be due to the fact that post-editing of MT is still not widely accepted by the translation community.

When asked about other features not included in the above list, most subjects could not identify anything else that they considered to be relevant in CAT tools. Exceptions were image editors and customizable QA rules and target text preview.

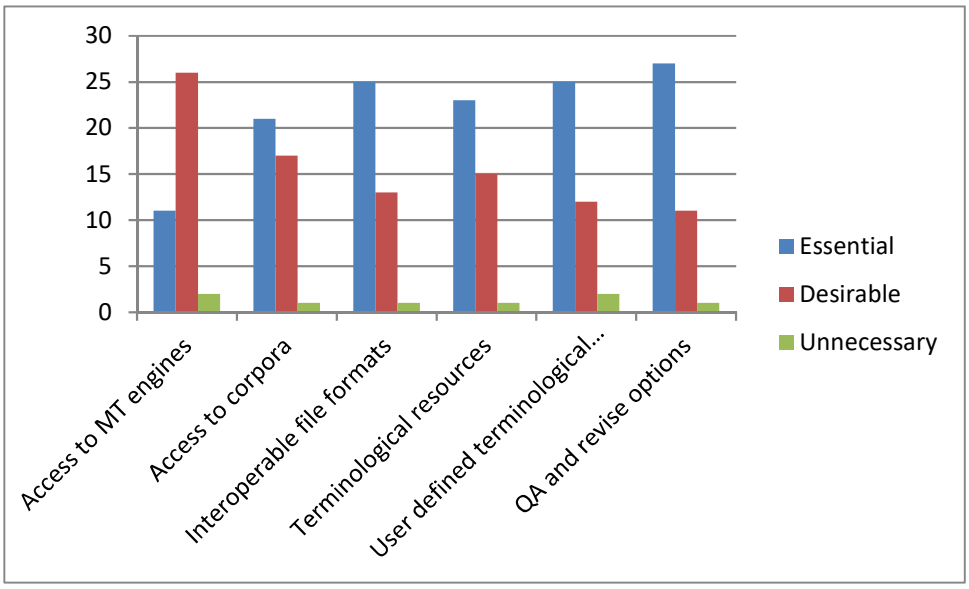

Figure 3: User expectations about CAT tools.

The 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, also merged from both experiments, show that definitions, translations, synonyms and variants, context and usage examples, phraseology and collocations and access to corpora are the most relevant data categories. Desirable categories include conceptual relations, register, images, etymology and compounds and derivatives. Pronunciation is the category most often regarded as unnecessary.

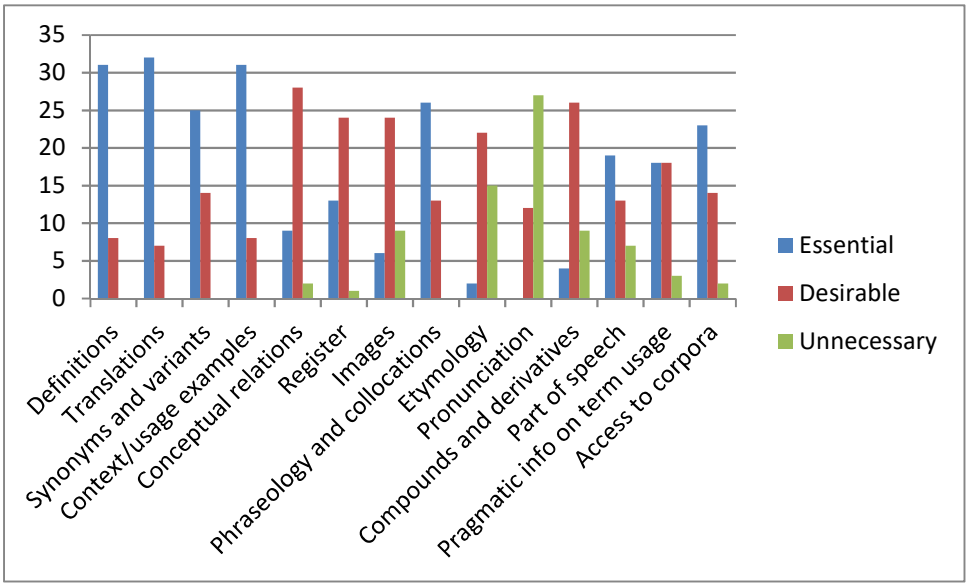

Figure 4: User expectations about terminological resources.

When asked about other features not included in the above list, most subjects could not identify any other that they regarded as relevant for terminological resources. Exceptions were specialized reference works and term use frequency, connotations, and false friends. The resources that subjects used the most for their translation assignments were as follows: Wordreference, Linguee, Reverso Context, IATE, Merriam-Webster, Oxford dictionaries, Collins, Cambridge Dictionary, RAE, esTenTen and enTenTen corpora in Sketch Engine, the BNC, CREA, the web as a corpus, CORPES XXI, Pons and Termium Plus, Glosbe, DeepL, ProZ forum, WIPO Pearl, and Medline Plus.

The subjects' answers indicated that EcoLexiCAT meets most user needs and expectations, but they also highlight how to improve the tool as well as EcoLexicon. For instance, currently there is a phraseology module (essential for most subjects) under construction in EcoLexicon that will be linked to EcoLexiCAT in the future. Part-of-speech is currently included as a data category in EcoLexicon but not in EcoLexiCAT. Therefore, based on the fact that most users considered it essential or desirable, it will be included in the next version. Furthermore, some of the resources reported by users had already been included based on the feedback received after experiment 1. However, it was impossible to include others because they do not allow embedding.

4. User behaviour

While completing their assignments, EcoLexiCAT subjects were monitored through Google Analytics. Prior to the evaluation task, we defined a series of "Events" based on the kind of actions that we wished to monitor. These "Events" in Google Analytics can be tracked according to a three-level structure consisting of Category (e.g. EcoLexicon), Action (e.g. definition by clicking on the terms) and Label (e.g. breakwater), which would mean that when users search for the definition of breakwater in EcoLexicon by clicking in the editor, the event is stored as such. This allowed us to compare the real use of each resource and the kind of queries that subjects make through a certain kind of action (e.g. definitions, translations, images, etc. from the right-click menu, by clicking in the editor, in the search form of each left-hand box, etc.). Table 1 shows a summary of the main actions tracked within each resource.

In experiment 1, a total of 5,693 events were stored during the completion of the assignments. Obviously, most of them took place within MateCat (4,874), but of the other resources, EcoLexicon stands out with 473 events (58%). EcoLexicon is followed by BabelNet, with 262 events (32%); other resources, with 47 (6%); IATE, with 27 (3%); and Sketch Engine with 10 (1%) (Figure 5).

In experiment 2, a total of 8,650 events were stored, and this higher number makes sense since both subject groups worked with EcoLexiCAT. Again, most of them took

place within MateCat (7,694). EcoLexicon, with 695 events (74%), was followed by other resources, with 88 events (9%); Sketch Engine, with 72 (8%); BabelNet, with 60 (6%); and IATE, with 41 (4%) (Figure 6). The number of events for other resources is higher than in experiment 1 (from 6% to 9%), and this is probably because new resources were added after the first experiment. The use of Sketch Engine is much higher than in experiment 1 (from 1% to 8%), which is undoubtedly an indication of the subjects' competence in corpus analysis. What is surprising is that the use of BabelNet dropped dramatically (from 32% to 6%).

Matecat	Insert-text, open-segment, delete-text, translate, search, download-original, download-translation
EcoLexicon	Definitions-click, definitions-menu, definitions-form, translations-click, translations-menu, translations-form, showAll-menu, showAll-form, images-form, images-menu, open-menu (EcoLexicon in a browser), relations-form, relations-menu
BabelNet	Definitions-click, definitions-menu, definitions-form, translations-click, translations-menu, translations-form, compound_words-menu, compound_words-form, images-menu, images-form, relations-form, showAll-menu, showAll-form, relations-menu
Sketch Engine	Concordance-menu, concordance-form, sketches-menu, sketches-forms, CQL-form
IATE	Search-menu, search-form
Other resources	Load-Linguee EN-ES, Load-Linguee ES-EN, load-WordReference EN-ES, load-WordReference ES-EN, load-Cambridge EN-ES, load-Cambridge ES-EN, load-EcoLexicon EN, load-EcoLexicon ES, load-MetaGlossary, load-TermiumPlus, load-Wikipedia EN, load-Wikipedia ES, load-Onelook, load-Majstro, load-RAE

Table 1: Main actions tracked within each resource in EcoLexiCAT.

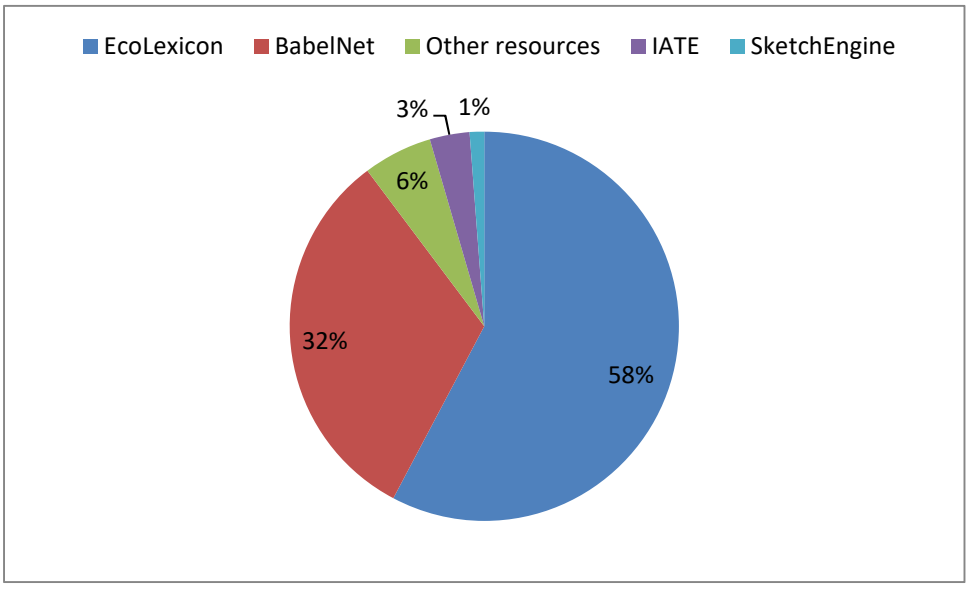

Figure 5: Events per resource in experiment 1.

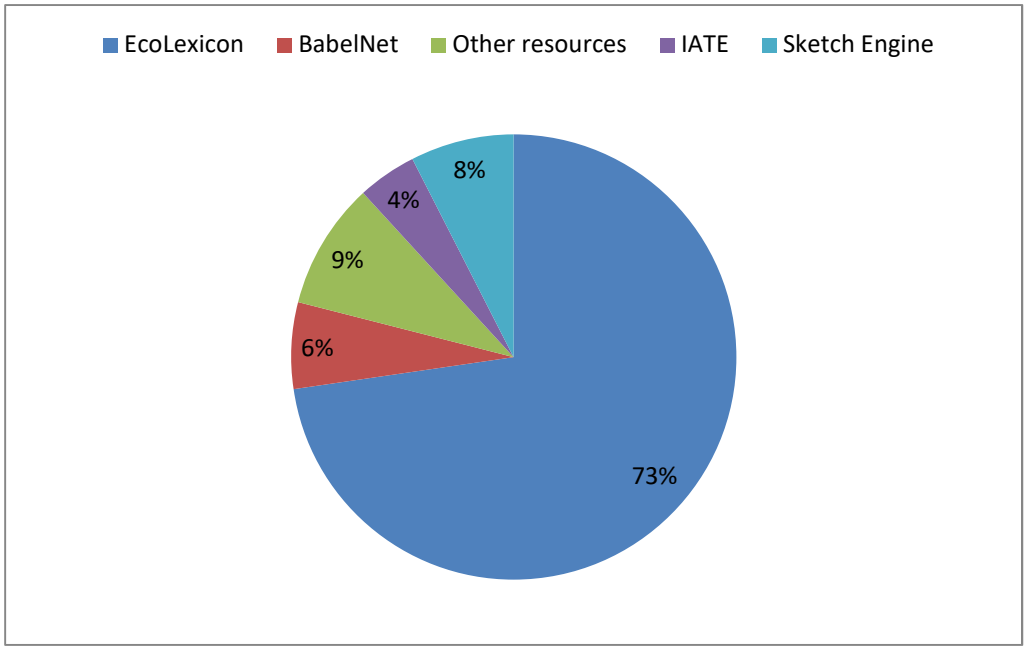

Figure 6: Events per resource in experiment 2.

From a quantitative point of view, the following figures (7-16) show the number and type of actions performed within each of the resources. This illustrates the usefulness of both the data categories of each resource (e.g. definitions, translations, images, etc.) and the way in which each category can be accessed (e.g. clicking, from the menu, writing the query in the box, etc.). For instance, in EcoLexicon (Figures 7-8) and BabelNet (Figures 9-10), definitions and translations are the preferred data categories. Clicking in the editor is clearly the preferred action in EcoLexicon in both experiments. However, in experiment 2, translations-form (writing the query in the box) and translations-menu (selecting from the right-click menu) were clearly preferred over definitions-form, definitions-click and translations-click in BabelNet. The number of actions for definitions-click and translations-click are the same, because when users clicked on one of the highlighted terms in the source segment of the editor, both kinds of information were deployed in the EcoLexicon and BabelNet boxes at the same time.

In experiment 1 in EcoLexicon, the subjects preferred to consult definitions and translations through the form in the box rather than the right-click menu access, whereas in BabelNet the opposite occurred. In experiment 2, where new events were added for new functionalities (e.g. semantic relations), in EcoLexicon subjects clearly preferred the definitions-menu option over the definitions-form option. Images were rarely consulted in either resource in both experiments. The open-menu option of EcoLexicon was used only once in experiment 1. From the EcoLexicon right-click menu, users have the possibility of opening EcoLexicon in a browser for a more detailed view of the conceptual networks. After experiment 1, the decision was made to add related concepts as a new data category in the EcoLexicon box to encourage users to explore the semantics contained in EcoLexicon. The relations-menu option

was used seven times in experiment 2, and the relations-form option was used four times.

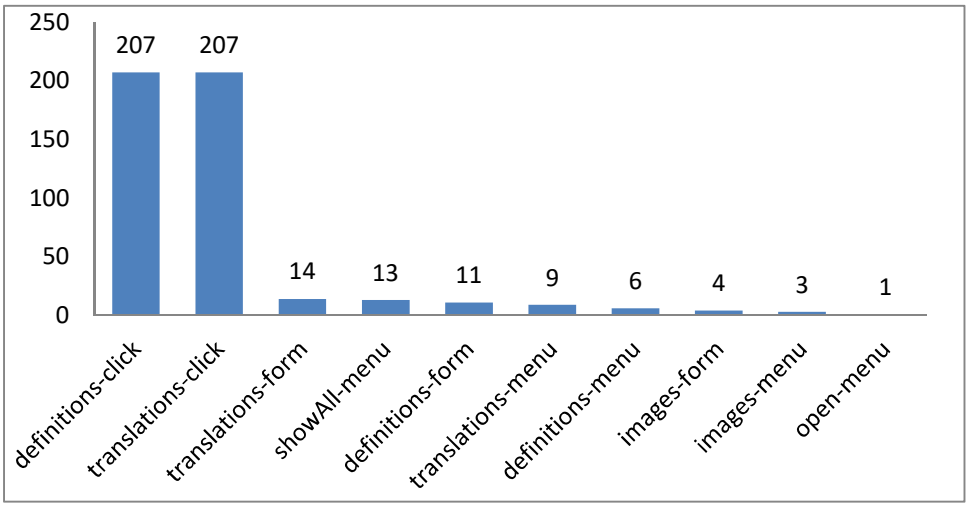

Figure 7: Actions performed within EcoLexicon – experiment 1.

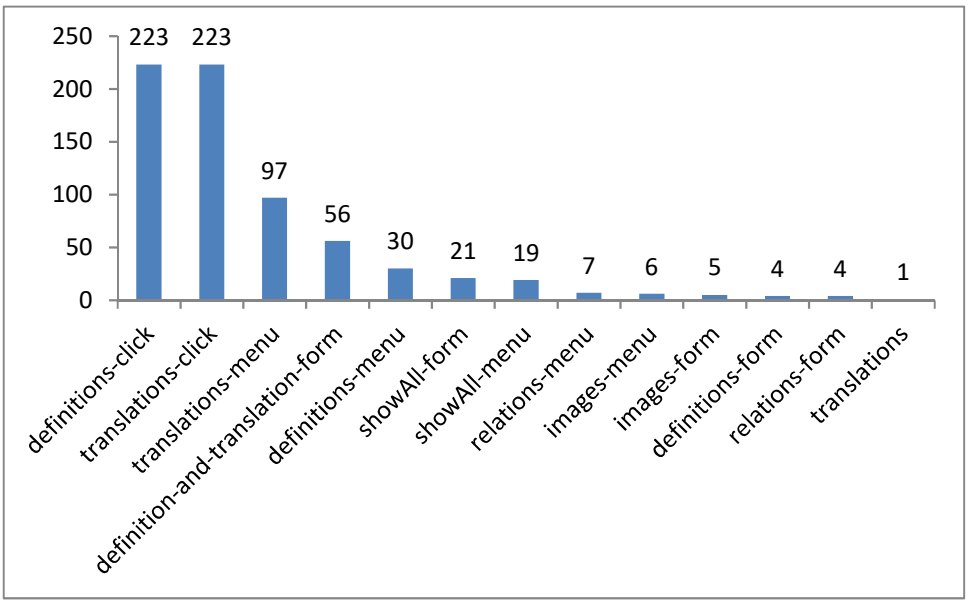

Figure 8: Actions performed within EcoLexicon – experiment 2.

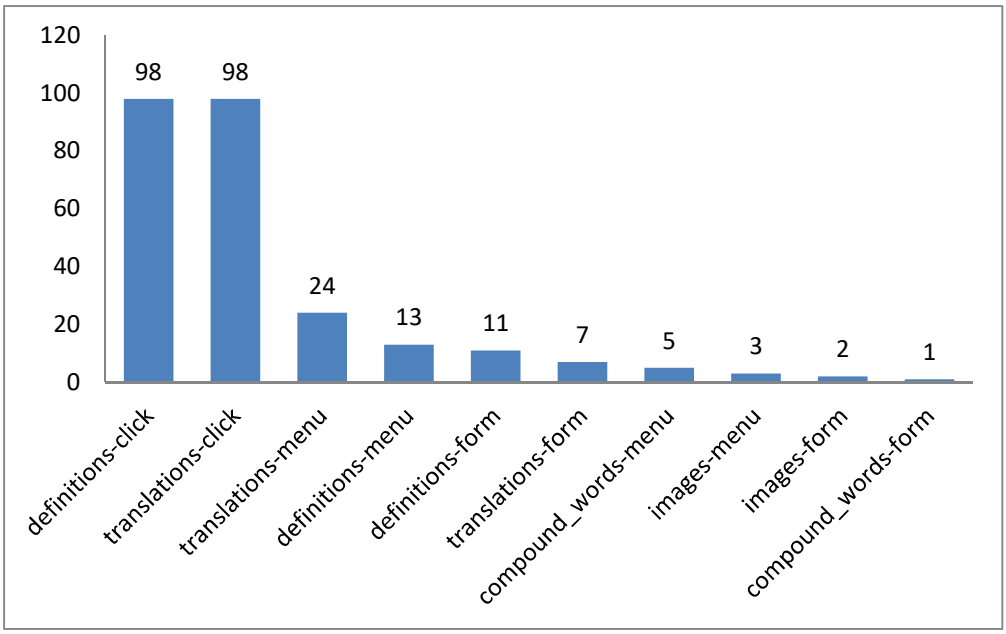

Figure 9: Actions performed within BabelNet – experiment 1.

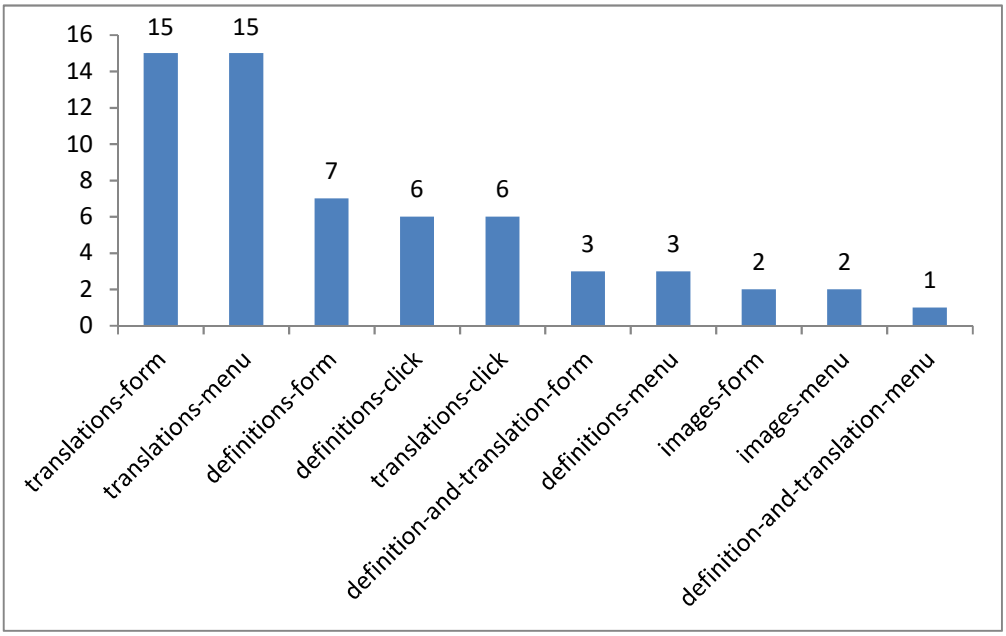

Figure 10: Actions performed within BabelNet – experiment 2.

The low number of actions carried out in Sketch Engine (Figure 11) in experiment 1 shows that the subjects were either not aware of the kind of information that can be extracted from a corpus, or did not know how to build meaningful queries. The latter is shown by the fact that seven of the 10 actions were simple concordance searches from the menu, where only the term needs to be selected in the editor. The subjects did not seem to be familiar with the basic syntax for more complex searches that would have provided more useful information, and they did not use the more advanced functionalities of corpus analysis, such as word sketches.

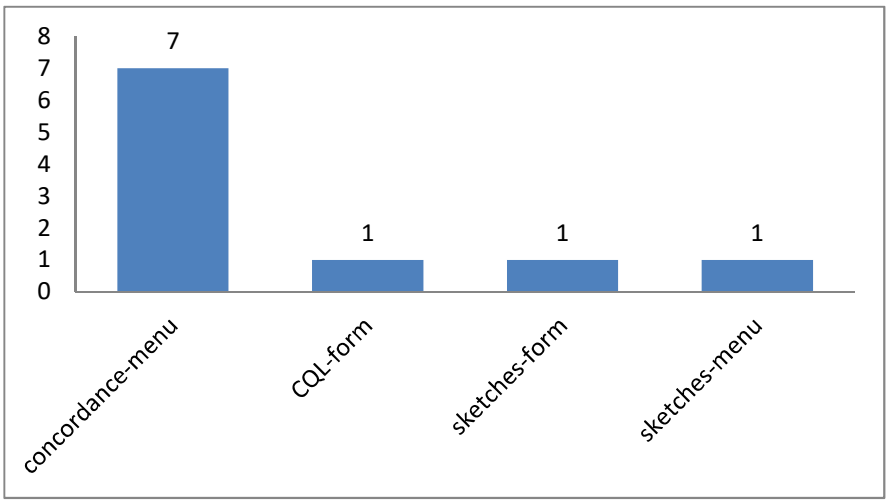

Figure 11: Actions performed within Sketch Engine – experiment 1.

In experiment 2, the subjects used Sketch Engine a great deal more (Figure 12), even if we take into account that the number of subjects working with Sketch Engine in EcoLexiCAT doubled in comparison with experiment 1. The concordance-menu was still clearly the preferred search option, but the other options were used as well, especially the word sketches, as opposed to the behaviour in experiment 1. The subjects in experiment 2 seem to be better versed in corpus analysis than those of experiment 1, although the more advanced option of CQL was only used twice.

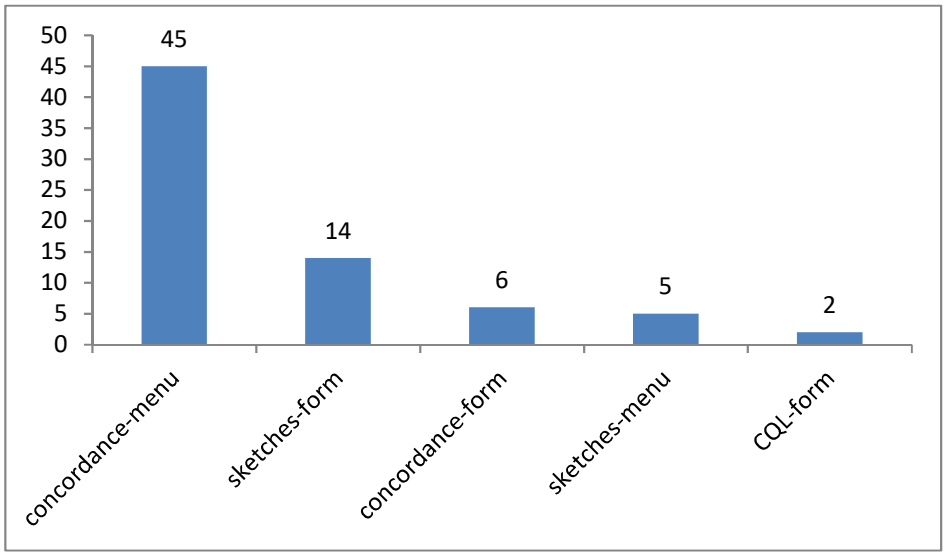

Figure 12: Actions performed within Sketch Engine – experiment 2.

The subjects in both experiments indicated the importance of having access to corpora in CAT tools, as most of them chose the essential or desirable options in the initial questionnaire (Section 3). In all likelihood, students are taught in their classes that corpus analysis is essential in the translation process, but not enough time is devoted to showing them how to actually obtain such information. In a study by Durán Muñoz

(2012), professional translators did not include access to corpora in their preferences when asked about terminological resources, probably because of lack of skills in corpus analysis and user-unfriendly search engines. Therefore, a user manual for EcoLexiCAT would have to provide easy-to-follow instructions on how to use the corpus options.

In IATE (Figures 13-14), 27 and 41 actions were carried out in experiments 1 and 2, respectively, with a slight preference for the right-click menu over the use of the form in the box in both experiments.

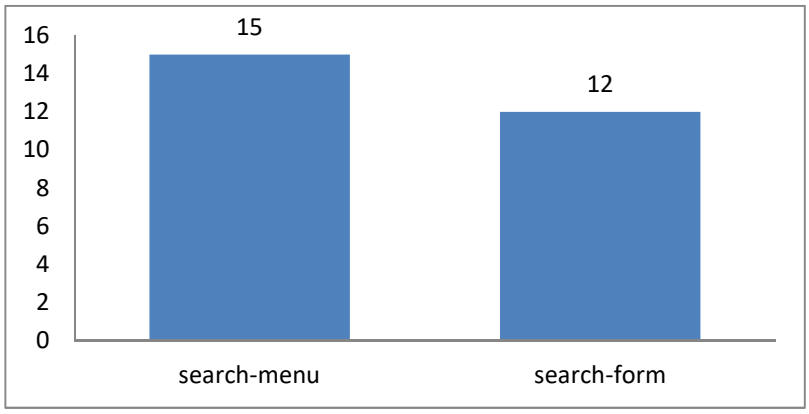

Figure 13: Actions performed within IATE – experiment 1.

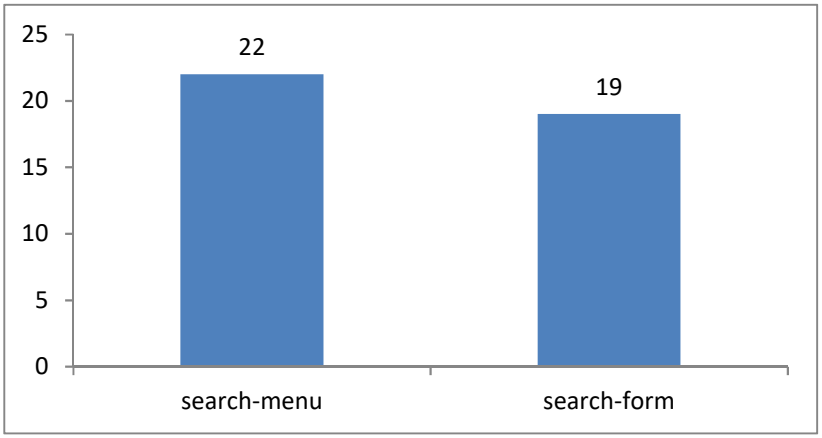

Figure 14: Actions performed in IATE – experiment 2.

With regard to other resources (Figures 15-16), the subjects in both experiments mostly used Linguee to find translation equivalents and terms in context, primarily during the first EN-ES translation task.

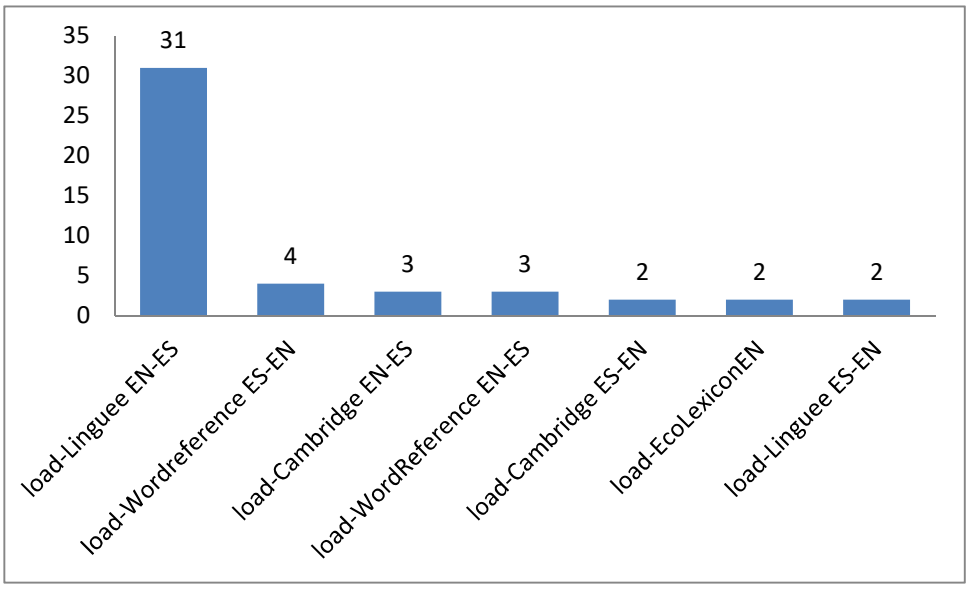

Figure 15: Actions performed within other resources – experiment 1.

However, in experiment 2 the subjects used more resources such as Wikipedia, TermiumPlus and Metaglossary, some of which were new resources added after experiment 1.

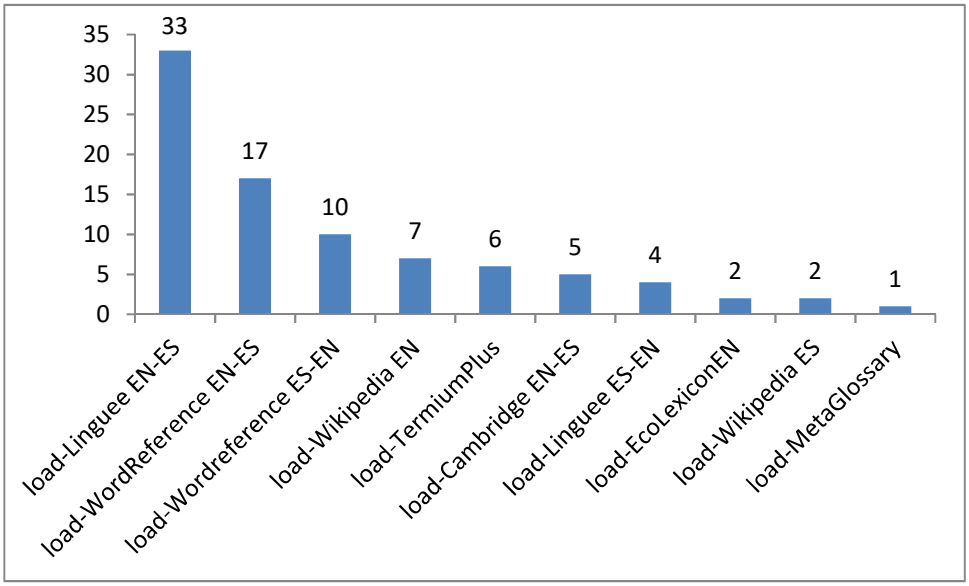

Figure 16: Actions performed within other resources – experiment 2.

From a qualitative point of view, the terms or text chains searched (labels) through each action within each resource were analysed and compared. In both experiments, there were many more searches with an English term or text chain as a starting point than a Spanish one. This is probably because the EN-ES task was performed first. Both tasks were on the same subject matter, and the subjects may have carried out most of the necessary research during the EN-ES task and already be familiar with many of the terms in both languages and the underlying domain knowledge.

Most users seem to have initially looked at the options provided in EcoLexicon for the terms marked in yellow in the source text (e.g. *detached breakwaters* and *hard coastal structures*). Then, when no option was given in EcoLexicon, subjects viewed the options marked in green in BabelNet, since the terms searched are clearly different at least in the most frequent searches. The order in this process was clearly influenced by the subject matter of the tasks as well as by the order and hierarchical structure of the terminological enhancement provided by EcoLexiCAT.

Regarding the kind of terms and chains searched for, multiword terms such *as hard coastal structure*, *detached breakwater*, and *artificial submerged reefs* were most extensively researched in nearly all resources. The search terms also matched the translation difficulties reported in the questionnaire that all subject groups filled in while translating in both experiments. Almost all difficulties reported were related to the lack of previous domain knowledge, which would impair the understanding of certain concepts, and to the lack of equivalences in the resources checked. Most of the resources that helped them solve their difficulties were the ones included in EcoLexiCAT, with the exception of some general language dictionaries and parallel texts found online. A few students also reported phraseological issues, which explains the queries of chains like *storm-induced*, *system* or *subject to*.

Curiously, EcoLexicon was searched for certain terms that initially seemed easy to translate, such as *erosion* (19 in experiment 1, 11 in experiment 2) and *cliffs* (10 and five, respectively). However, when working in a subject domain for the first time, researching more general terms and finding out how these concepts are related to others often helps to construct an initial mental representation of the domain.

What seems strange is that in experiment 1 some students looked for general language expressions, such as *continuamente* (continuously) and *significantly* in specialized resources such as EcoLexicon or BabelNet, instead of using the other resources menu. This indicates that maybe these resources should also be included on the left-hand side of the screen as a fifth box instead of as a pop-up window. On the other hand, this did not happen in experiment 2, which may again indicate that these subjects were better translators. However, in experiment 2 some subjects used the definition-and-translation-form to search for *define* and *remedy*. Furthermore, in both experiments the subjects looked for specialized terms in general resources such as Cambridge dictionary (*estuary* and *storm-induced* in experiment 1, and *detached breakwaters* and *soft cliffs* in experiment 2). Apart from that, some subjects in experiment 1 and fewer in experiment 2 used the definitions box in EcoLexicon (action: definitions-form) to find terms already marked in yellow in the text editor, such as *coastal structure*. This apparently strange behaviour can be explained by the fact that the subjects in our study were students with hardly any professional experience, although most students had a previous or almost finished translation degree, were students of a master's degree in translation, or both.

5. User performance

All target texts were evaluated by one reviser to ensure that the same criteria were applied in all cases. To assess the quality of the target texts of all groups, ten translation problems were identified for both the EN-ES and ES-EN assignments. The problems identified were based on those that the subjects mentioned repeatedly and on the reviser's expertise in the text type and domain. Depending on how well the subjects solved these problems, they could obtain up to 10 translation points. On the other hand, the language errors in both Spanish and English were subtracted 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.

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*.

In experiment 1 (Figure 17), the EcoLexiCAT translators outperformed the MateCat translators in both directionalities, although only slightly in the ES-EN assignment. The average quality of the target texts of both groups was not very high. This is understandable because most subjects in both groups did not have any professional translation experience or previous knowledge of the environmental domain. The results were promising though, as EcoLexiCAT helped to obtain a better target text in less time.

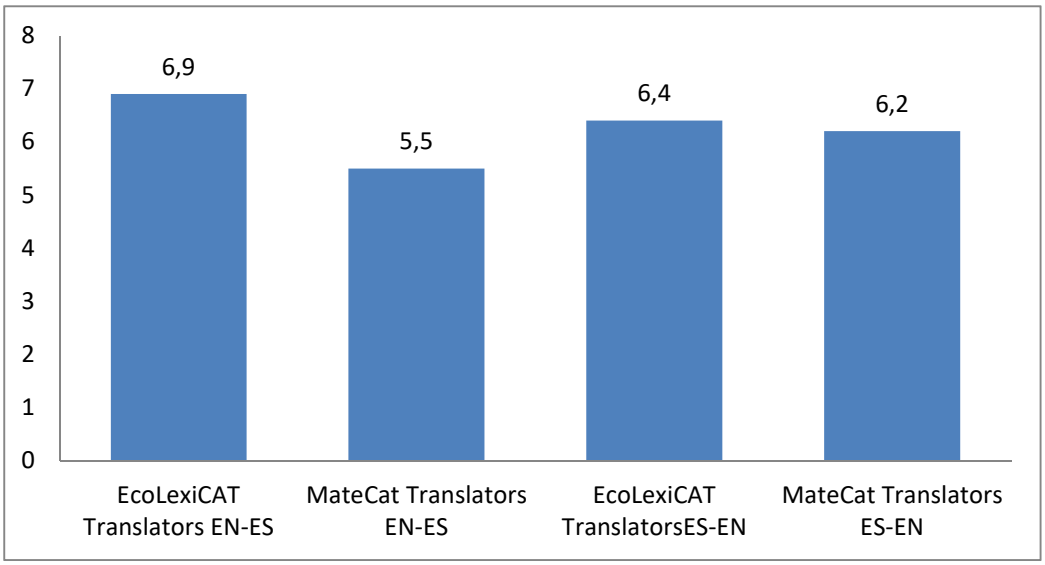

Figure 17: User performance in experiment 1 – quality.

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

In terms of the time invested (Figure 18), in both directionalities EcoLexiCAT translators outperformed the control group. Surprisingly, the EcoLexiCAT group took longer in the ES-EN assignment than in the EN-ES one, whereas the control group took longer in the EN-ES assignment. This is striking because even though it was a shorter source text, the assignment involved translating into a non-mother tongue of most of the subjects.

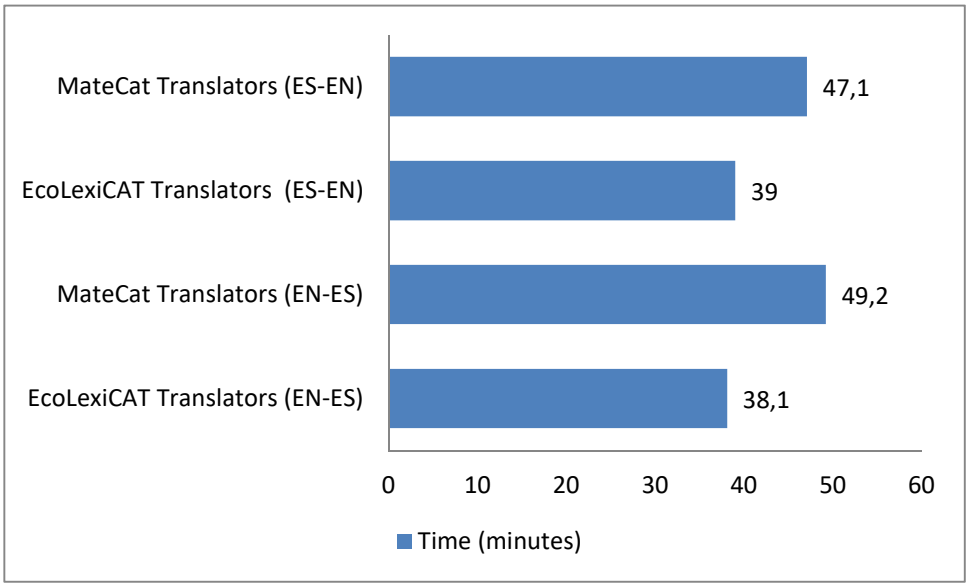

Figure 18: User performance in experiment 1 – time invested.

In experiment 2, however, the average quality of the target texts of both groups was higher than the average quality of both groups in experiment 1. This is surprising, as half of the subjects were undergraduate students in experiment 2, whereas in experiment 1 all of them were master's students. If we look at the translators group in experiment 2 (Figure 19, EN-ES: 8 and ES-EN: 7.3) and the group that translated with EcoLexiCAT in experiment 1 (Figure 17: EN-ES: 6.9 and ES-EN: 6.4), the improvement is clear, approximately one point more in both cases. In fact, in the ES-EN assignment, there was an average 9.3-minute time gain (Figure 20). This may be due to the fact that in experiment 2 better students were recruited, or that the improvements in EcoLexiCAT after experiment 1 had an impact on user performance.

As for the comparison between translators and post-editors in experiment 2, in terms of quality (Figure 19) the translators outperformed the post-editors in the EN-ES task, whereas in the ES-EN assignment the opposite occurred. In terms of the time invested (Figure 20), in both assignments post-editors outperformed translators, with a difference of 14.4 minutes for the EN-ES assignment and 10.6 in the ES-EN task.

This means that post-editing definitely reduces the average time spent on translation tasks, but it does not necessarily entail any improvement in quality.

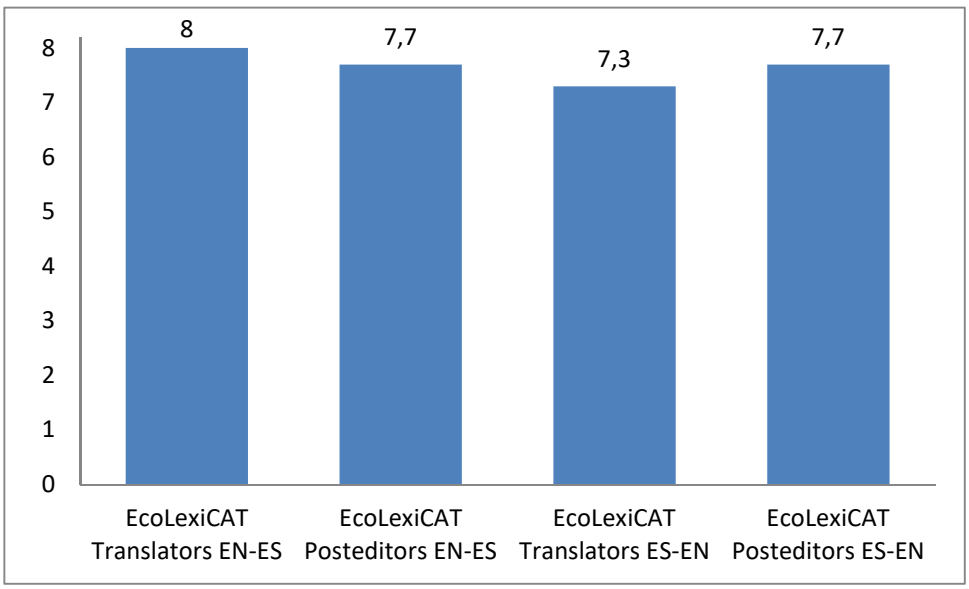

Figure 19: User performance in experiment 2 – quality.

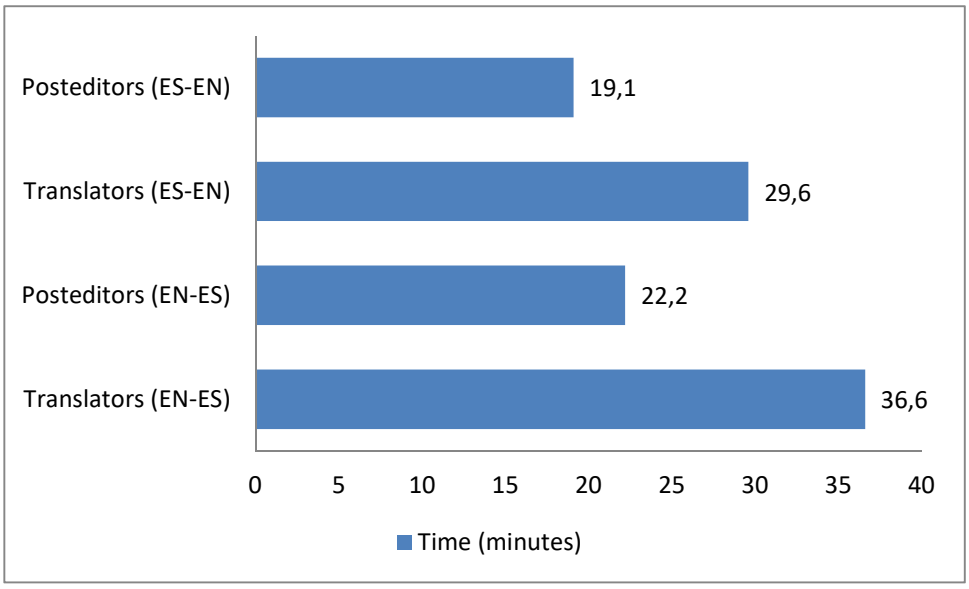

Figure 20: User performance in experiment 2 – time invested.

A qualitative comparison can be made when looking at the errors of the three groups (EcoLexiCAT translators in experiment 1; EcoLexiCAT translators in experiment 2; and EcoLexiCAT post-editors in experiment 2). Tables 2 and 3 show a collection of some of these errors accompanied by their frequency in all three subject groups.

When comparing the type of error made in the ES-EN task (Table 3), it seems that post-editing greatly reduced basic grammar and spelling problems in English, as all

subjects were translating into a non-mother tongue. However, when post-editing into their mother tongue (Table 2), subjects seemed to be more indulgent with the MT outputs.

In the EN-ES task, post-editors sometimes agreed too easily with MT options, for example when giving very literal translations of *are no remedy* and when translating *soft cliff* as *acantilado suave*, which in this context should be *blando*, since *soft* in this context refers to easily eroded cliff material. *Soft cliff* was a problem for all three groups, but the translators at least avoided the *suave* option. On the other hand, MT seemed helpful for the translation of the terms *inner surfzone* and *beach fills*. MT was also very helpful with the construction *to stabilize relatively deep tidal channels*, as the translators did not seem to understand that *relatively* affected the adjective *deep* and not the verb *stabilize*.

Translation problem EN-ES	Experiment 1 Translation	Experiment 2 Translation	Experiment 2 Post-editing
Term: hard coastal structure	1	2	1
Term: groyne, breakwater	2	2	4
Exp: are no remedy	5	1	6
Term: soft cliff	6	9	8
Term: high surge levels	2	5	4
Term: surf zone, inner surf zone	6	6	2
Comp: "Groyne…inner surf zone…"	6	2	1
Term: beach fills	3	5	2
Term: tidal, longshore and rip currents	2	0	1
Term: straight groynes, T-head, L-shaped and Y-shaped groynes	3	2	0
MT: at a more offshore position	1	2	1
MT: artificial submerged reefs	0	4	0
MT: mean sea level	0	4	3
MT: to stabilize relatively deep tidal channels	6	5	0
MT: diminish the generation of rip currents	0	1	0
MT: storm-induced erosion of sandy dunes and soft cliffs	3	0	4
MT: near the groyne heads	1	1	2

Table 2. Translation problems in EN-ES assignment.

However, because of the complicated word order of the sentence *storm-induced erosion of sandy dunes and soft cliffs during conditions with relatively high surge levels*, MT was not helpful in this case, whereas the translators in experiment 2 were capable of understanding the content. There are various indicators that the students of experiment 2 were generally better than those of experiment 1. For example, there were comprehension problems with the sentence: *Groynes are long, narrow structures perpendicular or slightly oblique to the shoreline extending into the surf zone (generally slightly beyond the low water line)*. Nevertheless, the subjects in experiment 2 tended

to be less precise, since they omitted *submerged* in the phrase *artificial submerged reefs*, and *mean* in *mean sea level* in their translations. In the first case, the post-editors did not show this problem, which was probably solved by the MT option.

In the ES-EN task the differences are not as clear, possibly because all the students were translating into a foreign language. However, MT again led to a more literal translation (e.g. in *remodelados*). In addition, in all cases where post-editors had problems with the term *ambientes mesomareales*, this was due to the fact that MT omitted *mesomareal,* and the post-editors did not correct this. Some results again show that the subjects of experiment 1 did not perform as well as those of experiment 2, as they had problems with expressions such as *están sujetos a* and comprehension problems with the sentence *para los casos de desembocaduras sin diques y con diques de encauzamiento.*

Translation problem ES-EN	Experiment 1 Translation	Experiment 2 Translation	Experiment 2 Post-editing
Term: sistemas abiertos	1	1	0
Exp: remodelados	4	4	9
Exp: están sujetos a	5	0	0
Term: acreción	2	1	1
Term: ambientes mesomareales	1	3	6
Term: un canal formado por una flecha	3	4	5
Term: bocana, desembocadura	9	9	8
Comp: relación existente… bocana	2	1	0
Term: diques de encauzamiento	9	2	1
Comp: Los procesos sedimentarios…barras.	4	0	1
MT exp: se aceleran	5	3	6
MT corrientes inducidas por el oleaje	1	1	0
MT condiciones de cierre	2	1	0
MT estuarios mareales	1	1	0
MT con una energía moderada del oleaje	0	1	0

Table 3. Translation problems in ES-EN assignment.

6. User satisfaction

User satisfaction was measured in three subject groups of 10 members each: EcoLexiCAT translators in experiment 1; EcoLexiCAT translators in experiment 2; post-editors in experiment 2. When asked about the general usefulness of the tool for the translation of environmental texts, the subjects in the first experiment said that the tool was very useful (60%) or useful (40%). Likewise, in the second experiment the subjects stated that the tool was very useful (70% EcoLexiCAT translators and 80% EcoLexiCAT post-editors) or useful (30% EcoLexiCAT translators and 20% EcoLexiCAT post-editors). No subjects answered “not very useful” or “useless”. These figures indicate that the tool had improved from the first to the second experiment,

and also that post-editors found it even more useful than the translators.

The parameters of functionality, usability and efficiency were evaluated, based on the rating of different items on a 1-to-5 Likert scale, where 1 was the lowest rating and 5 the highest. After that, subjects could fill out a free-text field to report problems, make suggestions for improvement, and/or note the tool's strengths.

Regarding functionality (Figures 21-23), the subjects were asked whether the tool contained suitable features for: (1) the translation of environmental texts; (2) the comprehension phase of an environmental text; and (3) the production phase of an environmental text.

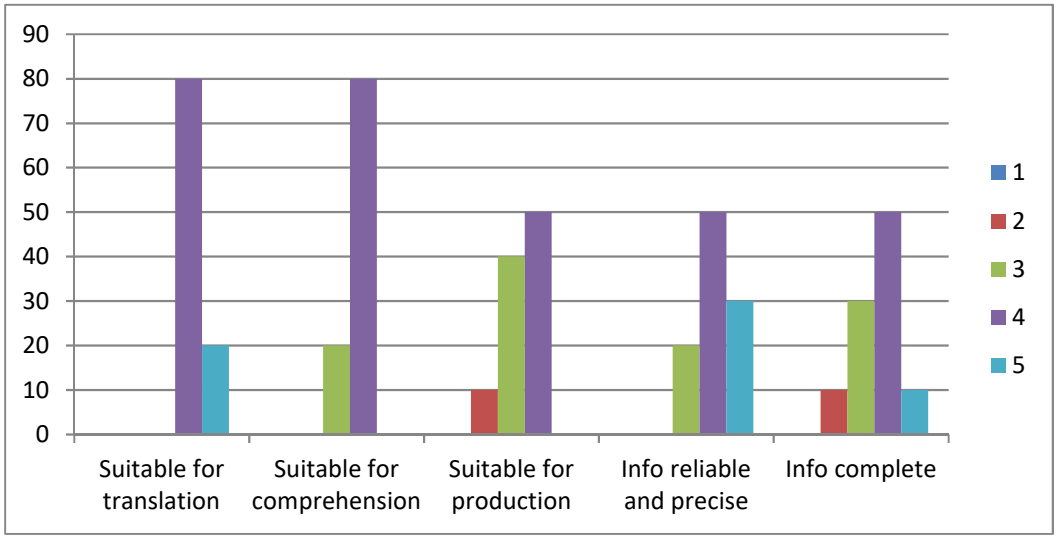

Figure 21: Functionality of EcoLexiCAT – translators in experiment 1.

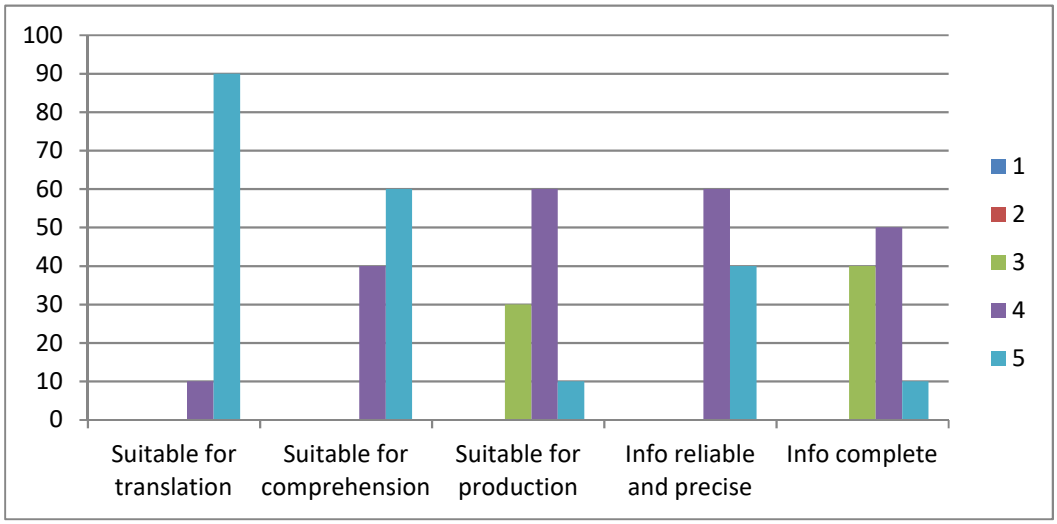

Figure 22: Functionality of EcoLexiCAT – translators in experiment 2.

Generally speaking, from experiment 1 to experiment 2, the tool was rated better, since its suitability for translation was given a score of 5 by 20% of translators in experiment 1, as compared to 90% of translators and 80% of post-editors in experiment 2. Its suitability for the comprehension phase was rated better than that for the production phase in all three groups, though the upward trend continued from experiment 1 to experiment 2. Comprehension was rated with a score of 4 by 80% of translators in experiment 1, but with a 5 by 60% of translators and post-editors in experiment 2. Production received a somewhat lower score, which means that EcoLexiCAT is currently more comprehension-oriented, and that future improvements should focus on increasing assistance in production-oriented tasks. However, a slight upward trend was still evident from experiment 1 to experiment 2. The minimum score in experiment 2 is 3, and the percentage of 4 rose from 50% in experiment 1 to 60% in experiment 2. Not surprisingly, 40% of the post-editors rated production with a 5 and 60% with a 4, which is only natural, since in the case of obtaining highly reusable MT output the text production phase was obviously enhanced.

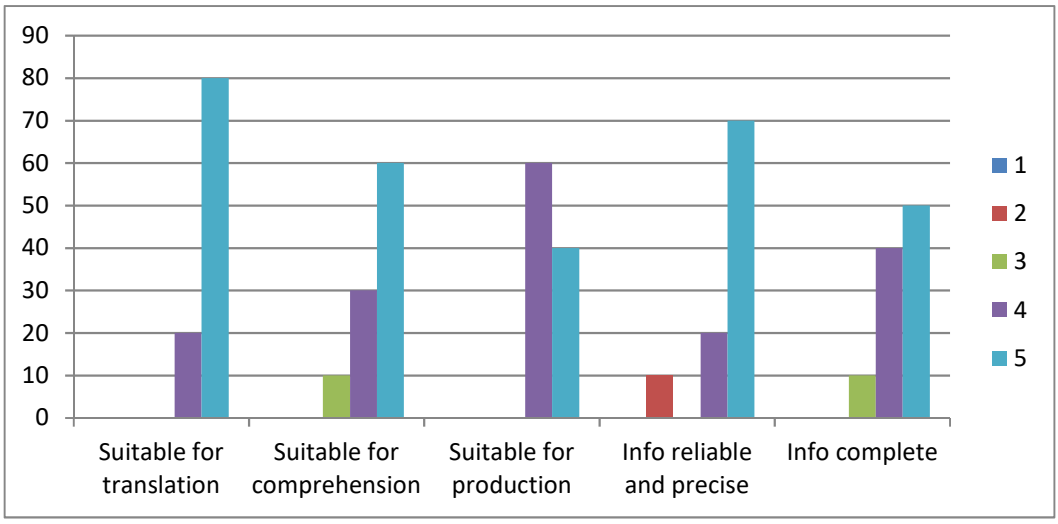

Figure 23: Functionality of EcoLexiCAT – post-editors in experiment 2.

Regarding the reliability, precision, and completeness of the information provided, the scores given by translators in both experiments were practically the same, whereas those given by post-editors were much higher. For instance, reliability and precision were given a 4 by 50% and 60% of translators but received a 5 from 70% of post-editors. These figures call for the continuous extension, improvement and maintenance of terminological resources. Similarly, in experiment 1 one of the subjects stated that the only improvement necessary was to expand the knowledge currently contained in EcoLexicon. Moreover, a translation difficulty reported by a few subjects was the fact that in all resources synonyms and term variants are listed with no clues on how to choose one or another.

When asked to rate the usefulness of external resources during their assignments (Figures 24-26), EcoLexicon, Sketch Engine, Linguee and Wikipedia were rated best in experiment 1. However, this did not exactly correspond to user behaviour (Section 4),

since Sketch Engine was rarely consulted, and Wikipedia was not consulted at all. This shows how users' introspection cannot be the only method used to evaluate a tool. In experiment 2, where new resources were added as other resources (TermiumPlus, Metaglossary, OneLook, and Majstro), EcoLexicon was again the best rated resource (rated 5 by 90% of translators and post-editors), followed by Linguee, Wordreference, Cambridge, and Wikipedia. Again, these results do not exactly correspond to the figures reported in Section 4. For example, Sketch Engine was not reported among the best resources even though it was used more often than individual other resources. Among the worst rated resources (because they were not useful or were not needed), Termium Plus, Metaglossary, OneLook, and Majstro were mentioned. These resources were among those integrated after experiment 1.

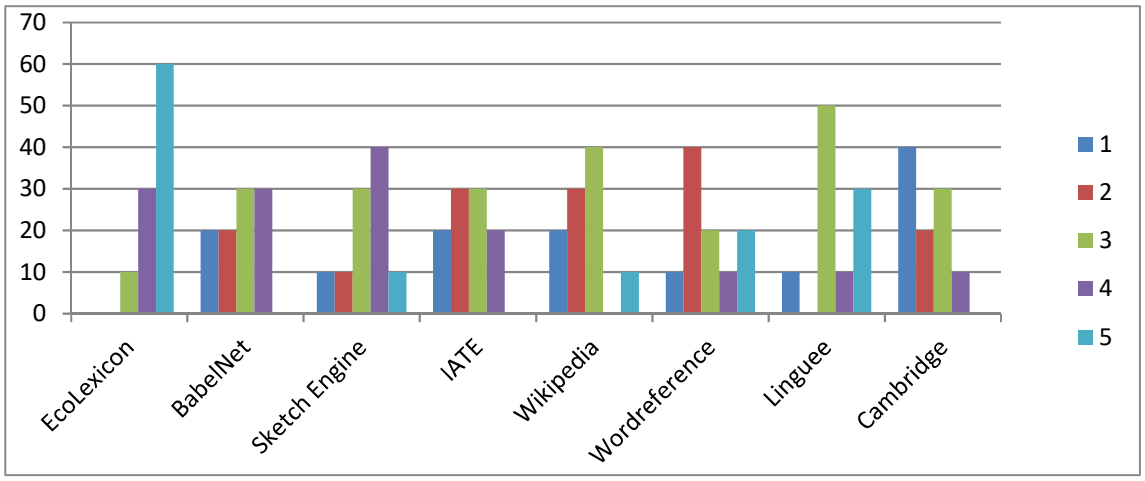

Figure 24: Usefulness of external resources – translators in experiment 1.

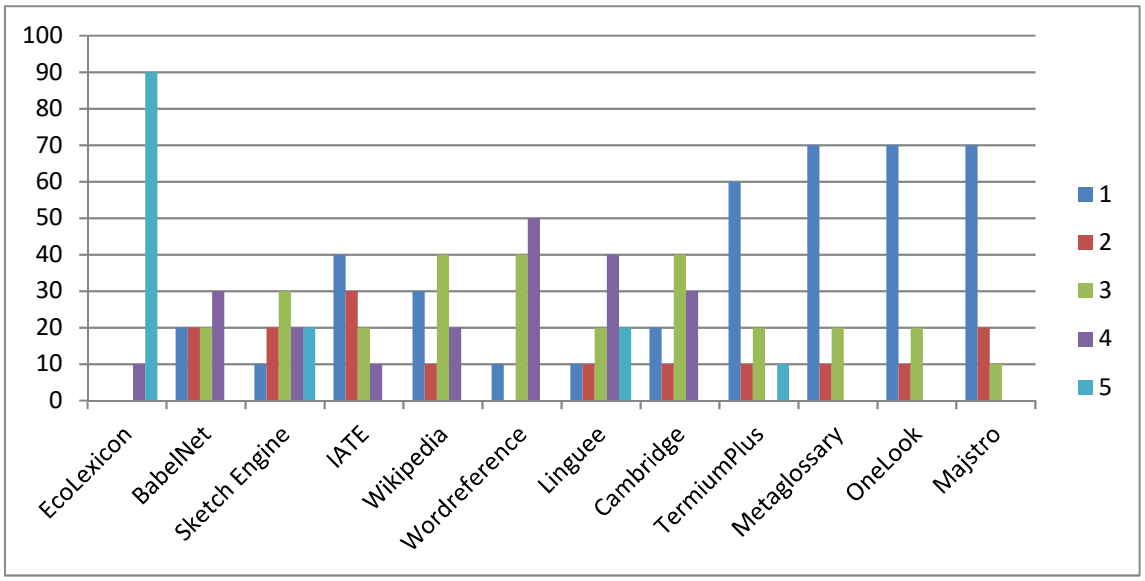

Figure 25: Usefulness of external resources – translators in experiment 2.

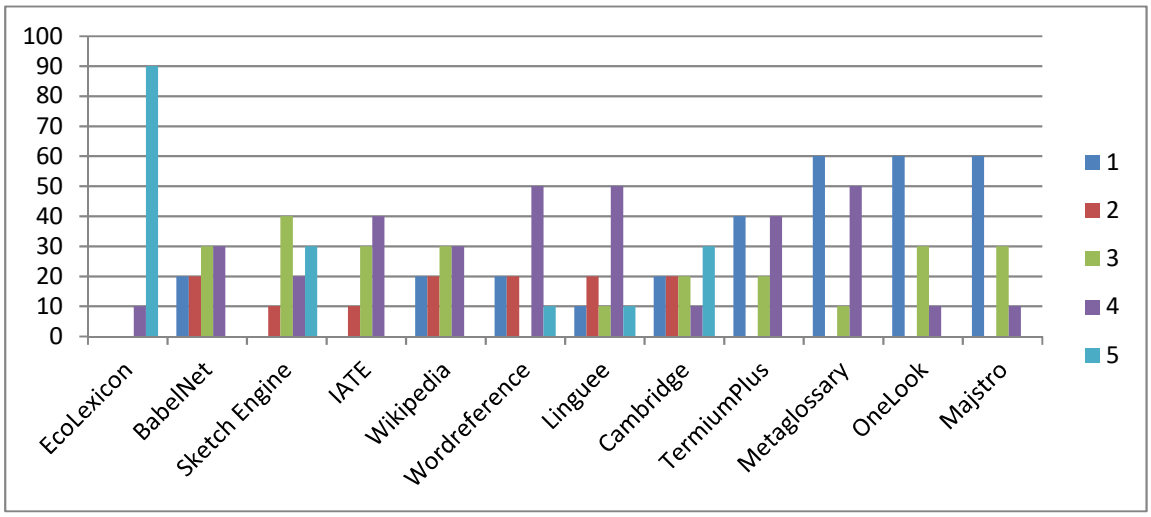

Figure 26: Usefulness of external resources – post-editors in experiment 2.

As for usability (Figures 27-29), the subjects were asked the following about EcoLexiCAT: (1) if it was intuitive and easy to use; (2) if it had a functional design; and (3) if it provided an adequate interaction with the layout (e.g. resizing of the windows).

In both experiments the interaction with the layout was rated the worst. Thus, future improvements should head in this direction, although some of them were already integrated after experiment 1. The score of the design remained stable in the translators groups (in both experiments 40% of the translators rated it with a 5, and 50% with a 4), although the post-editors rated it higher (70% with a 5 and 30% with a 4). Regarding ease of use, this was the parameter that improved the most, since 40% of translators in experiment 1, 70% of the translators in experiment 2, and 100% of the post-editors in experiment 2, rated it with 5.

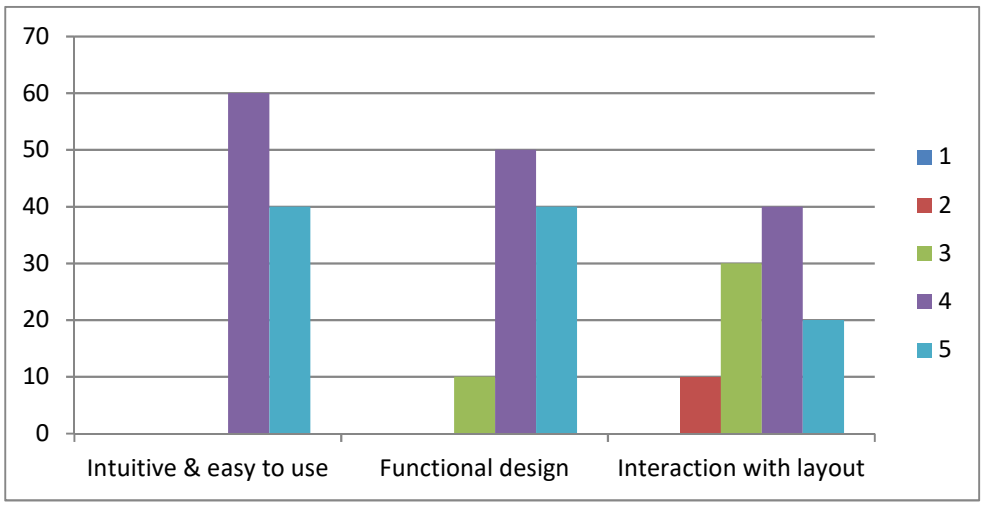

Figure 27: Usability of EcoLexiCAT – translators in experiment 1.

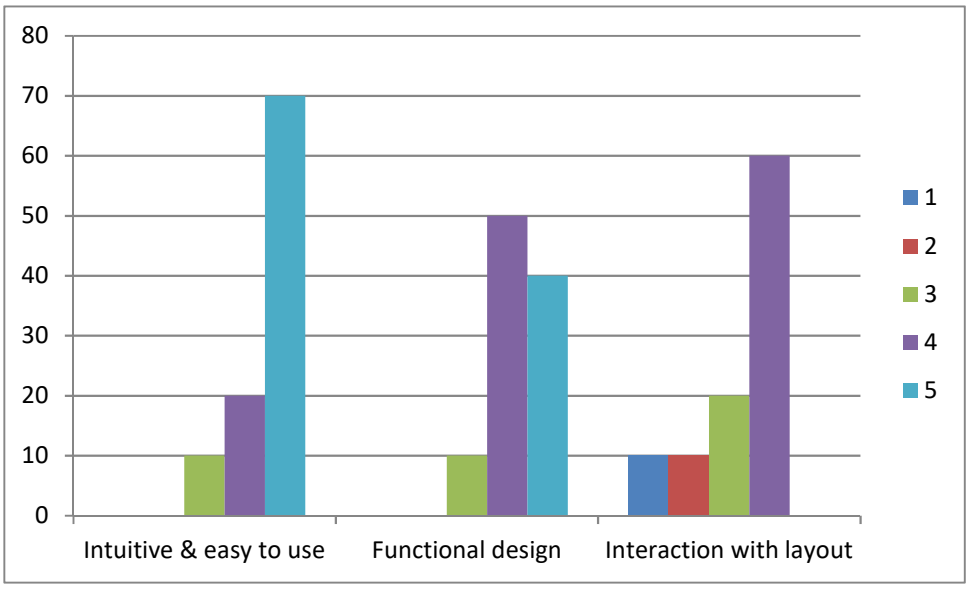

Figure 28: Usability of EcoLexiCAT – translators in experiment 2.

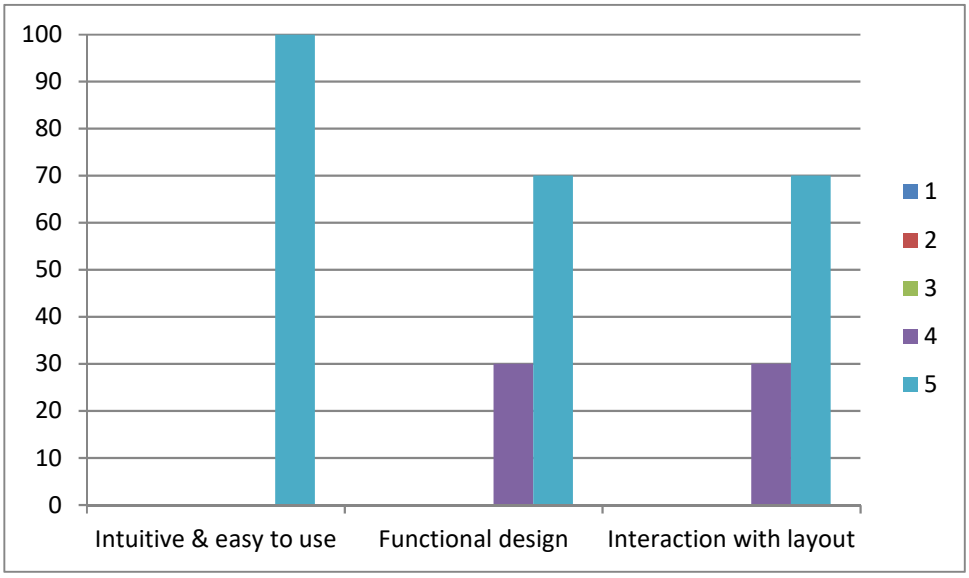

Figure 29: Usability of EcoLexiCAT – post-editors in experiment 2.

Finally, efficiency (Figures 30-32) was assessed based on whether the information loaded at the right speed and fluidity: (1) user interaction with the editor; (2) interaction of the editor with external resources; and (3) user interaction with external resources. All parameters improved from experiment 1 to 2. In experiment 1, they were mostly rated with a 4, whereas in experiment 2 they were mostly rated with a 5. In experiment 1, user-editor and user-resources interaction scored worse than information loading speed and editor-resources interaction. In experiment 2, user-editor interaction and information loading speed improved significantly, but user-resources and editor-resources interaction showed some flaws, even if the general trend was positive. Comparing translators' and post-editors' assessments in experiment 2, post-editors clearly gave higher scores to all parameters.

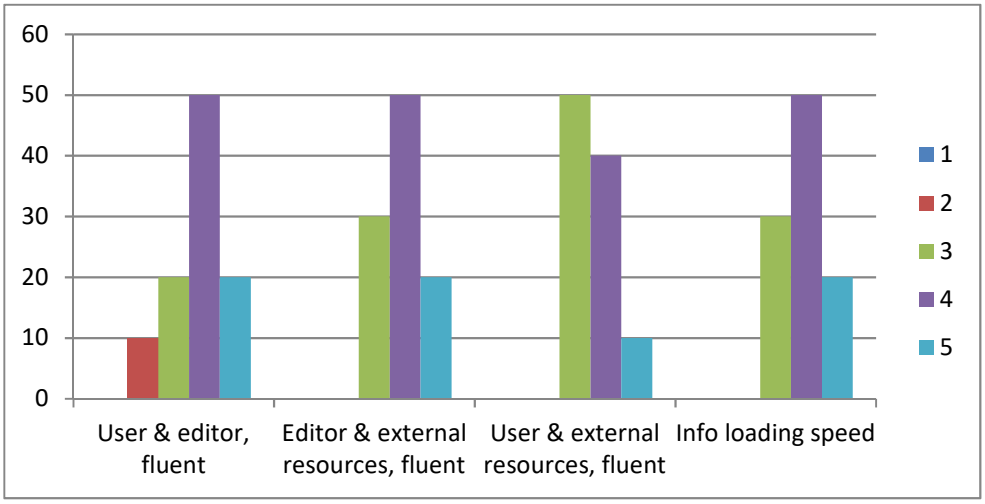

Figure 30: Efficiency of EcoLexiCAT – translators in experiment 1.

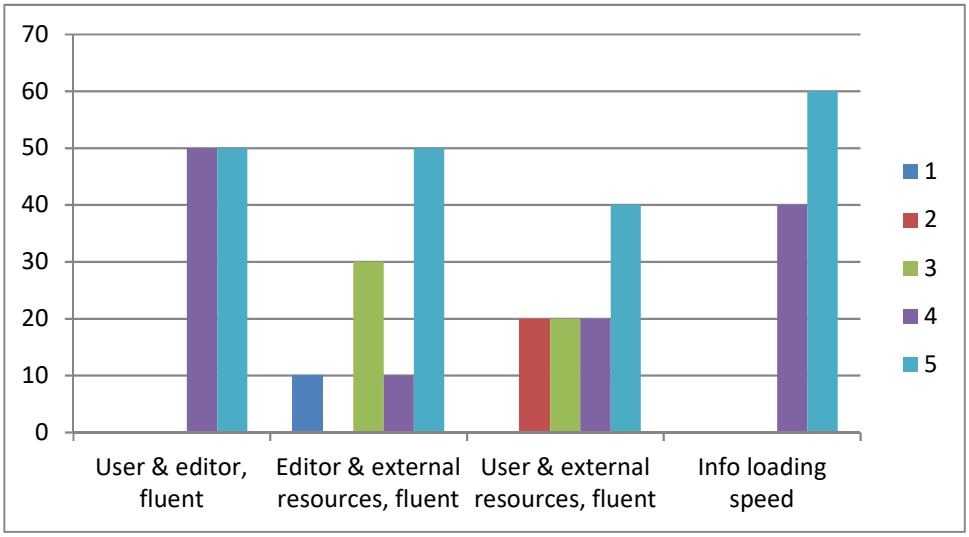

Figure 31: Efficiency of EcoLexiCAT – translators in experiment 2.

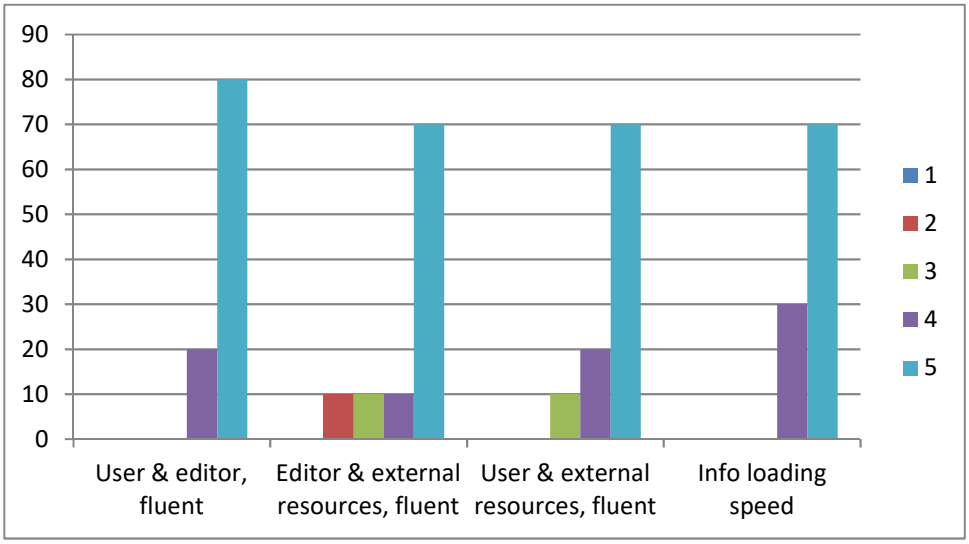

Figure 32: Efficiency of EcoLexiCAT – post-editors in experiment 2.

The post-editor group also answered a question regarding MT efficiency. They were asked to assess on a 1-5 Likert scale the frequency with which they encountered common issues in MT (i.e. inadequate terminology, literal translation, problems with numbers and figures, omissions, additions, etc., with the results shown in Figure 33). These results, together with those related to the time invested, show that the reusability of MT output was significantly high. Unintelligible segments were rare, as well as omissions, additions and issues related to spelling, gender and number, punctuation and capitalization, and words that should be kept in the source language. In contrast, word order, literal translations, and inadequate terminology were the issues that were most often encountered, and on which the post-editing process had to focus. Most users acknowledged that MT was of great help.

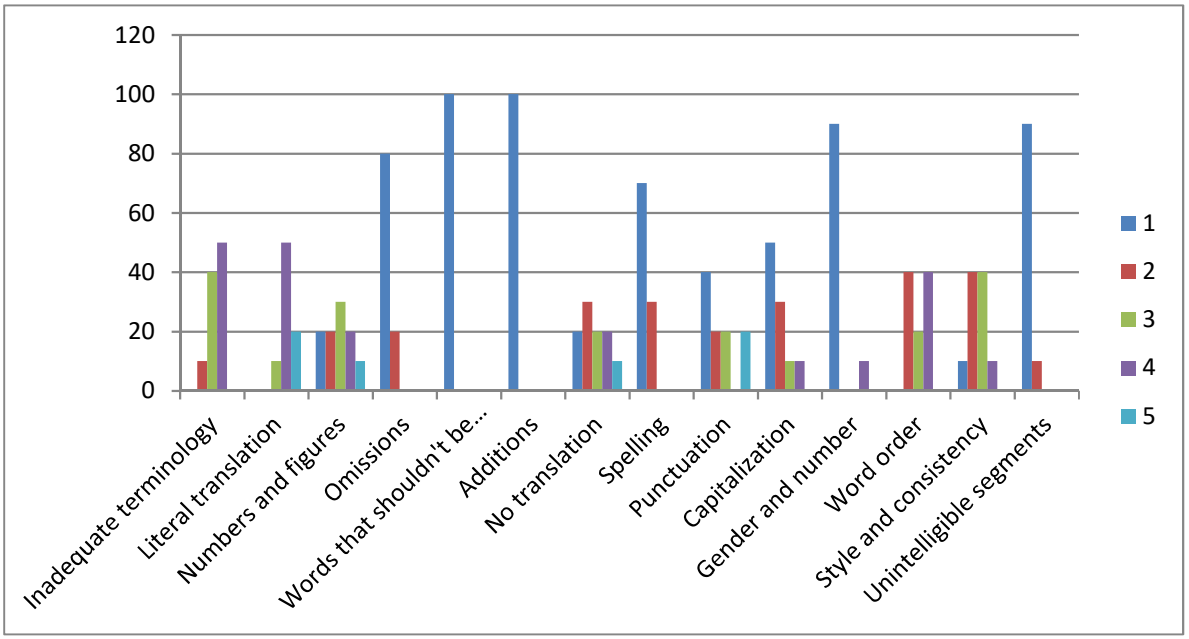

Figure 33: MT issues – post-editors in experiment 2.

When asked about the tool's flaws and possible improvements, in experiment 1 several subjects reported some bugs and efficiency issues regarding the other resources pop-up window – it could not be resized or moved, making things difficult to see – and the predictive typing feature in the target segment, which did not work well in the case of multiword terms. Moreover, certain plural multiword terms in Spanish were not lemmatized properly, and thus not recognized as terms in EcoLexicon. These issues were addressed before experiment 2, but again users reported other problems related to both issues: sometimes the other resources window would disappear until the browser was refreshed, and the predictive typing feature added a line break in the target segment.

One subject in experiment 1 suggested adding the other resources window to the left-hand side of the screen, as already inferred from the analysis of user behaviour (Section 4). However, in experiment 2 the users seemed to be happy having the general

language resources in that window, instead of placing them with the terminology resources.

Among other suggestions for improvement, the subjects in experiment 1 proposed the addition of the resources added before experiment 2. In experiment 2, the subjects proposed the inclusion of an environmental corpus in Spanish, part-of-speech information, style guides, reliability rates for terms usage, possibility of having shortcuts for the different searches, and a better integrated quality assessment tool.

When asked about the positive aspects of the tool, many subjects in both experiments pointed out that the quick and easy access to so many resources in the same interface, as well as the fact that the search terms do not need to be typed, is the main strength of the tool, which is the whole idea behind our concept of terminology enhancement. However, there were also several users that felt overwhelmed by the amount of information shown. They proposed making the layout more flexible so that users could customize the order, amount, and position of resource boxes. Users also highlighted the usefulness of Sketch Engine and EcoLexicon, especially its definitions, term equivalents, and images.

7. Conclusions and future work

Based on user expectations, EcoLexiCAT can be regarded as a tool specifically tailored to user needs and conceived in line with the augmented translation approach. According to user performance, the results of the experiments indicate that integrating terminology enhancement in the translation workflow in a stand-alone interface improves the quality of the translation and reduces the time spent on the task. MT post-editing, however, reduces the time spent on the task but does not necessarily raise the quality. With regard to user behaviour, we can conclude that the most useful resource in EcoLexiCAT is EcoLexicon, which is hardly surprising, since the tool is specifically conceived for environmental translation. The increased use of Sketch Engine was observed in experiment 2. Definitions and term equivalents were the data categories most often consulted in all resources. Users also showed a clear preference in the way they accessed information. In this sense, clicking was the preferred mode, followed by the right-click menu option, and finally by typing the search in the form.

Regarding user satisfaction, the three parameters point to a favourable evaluation of EcoLexicon, although efficiency will be the first aspect to be improved in the future. Post-editors tended to rate the tool better as a whole, since all parameters showed higher figures in this subject group. Comparing translators' general assessments in experiments 1 and 2, those in experiment 2 were slightly better. We can thus conclude that both improvements from experiment 1 to 2 and the MT feature had a positive impact on the evaluation of EcoLexiCAT.

Based on these studies, EcoLexiCAT thus seems to be on the right path. However, it

still needs to be assessed by more prospective users. Wider studies with larger samples, including professional translators, will be carried out in the future. Other features and resources will also need to be added to the tool, especially those related to text production tasks, such as phraseological information and access to the EcoLexicon Spanish corpus. All flaws and bugs reported will also be fixed. Moreover, if EcoLexiCAT were extensively used, it would be possible to draw meaningful conclusions about the kind of terms/concepts most researched through each of the resources and data categories. This would provide valuable insights into how to build and improve augmented translation tools.

8. Acknowledgements

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

9. References

De Palma, D.A. & Lommel, A. (2017). *Augmented Translation Powers up Language Services*. Common Sense Advisory.

Durán Muñoz, I. (2012). Meeting translators' needs: translation-oriented terminological management and applications. *The Journal of Specialised Translation*, 18, pp. 77–92.

Faber, P., León-Araúz, P. & Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In E. Bárcena, T. Read & J. Arús (eds.) *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, 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.

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.

ISO/IEC 9126-1 (2001) Software engineering -- Product quality. International Organization for Standardization, Geneva, Switzerland.

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

León-Araúz, P., Reimerink, A. & Faber, P. (2019). Translating environmental texts with EcoLexiCAT. In M. Ji (ed.) *Translating and communicating environmental cultures*. London: Routledge.

León-Araúz, P., San Martín, A. & Reimerink, A. (2018) The EcoLexicon English Corpus as an open corpus in Sketch Engine. In J. Čibej, V. Gorjanc, I. Kosem & S. Krek (eds.) *Proceedings of the 18th EURALEX International Congress*, Ljubljana: Faculty of Arts, pp. 893-901.

León-Araúz, P. & Reimerink, A. (2018). Evaluating EcoLexiCAT: a Terminology-Enhanced CAT Tool. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis & T. Tokunaga (eds.) *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France: European Language Resources Association (ELRA).

Lommel, A. (2018). Augmented translation: A new approach to combining human and machine capabilities. *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas* (Volume 2: User Papers), Boston, March 17 - 21, 2018.

Lommel, A. (2017). Augmented Translation. In Brown-Hoekstra (ed.) *The Language of Localization*. XML Press.

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.