From the Stinking Bishop to the Abbucciatto Aretino (and back)
Using corpora in the translation classroom

By Letizia Cirillo (Università di Siena, Italy)

Abstract & Keywords

English:
The present paper illustrates a short module of corpus-assisted translation conducted at the University of Siena for undergraduates
attending the BA in Languages and Intercultural Communication. Students were shown how to design and compile their own specialized
corpora, which they then used to produce a) translations into Italian of an English text on the traditional "Stinking Bishop"
cheese; b) translations into one of their foreign languages (English, French, German or Spanish) of a similar Italian text on the
typical "Abbucciatto Aretino" cheese; and c) presentations on relevant issues of corpus compilation and consultation. The projects thus obtained
show a strong linguistic-translational competence and considerable metalinguistic-metatranslational awareness, characteristics which
support the case for data-driven activities as a powerful tool enhancing translation students' observation and reflection skills and
encouraging their learning autonomy.

Keywords: specialized language, corpus-assisted translation, DIY comparable corpora

1. Introduction: corpora in translator education

In the past two decades, i.e. collections of "machine-readable authentic texts (...) sampled to be representative of a particular
language or language variety." (McEnery et al., 2006: 5), have had a growing impact on the practice and teaching of translation, and
corpus-based translation studies have become a sub-discipline of arguably both translation studies and corpus linguistics (see Granger &
Petch-Tyson, 2003; Olohan, 2004; Beeby et al., 2009; Zanettini, 2012; Fantinuoli & Zanettini, 2013, among others). It is undeniable
that being able to use state-of-the-art resources, such as digital reference and analytical tools, has become a crucial skill for (prospective)
translators (and, more generally, language service providers) to successfully respond to the demands of the job market. Corpora,
together with corpora compilation and consultation tools, are part and parcel of these resources, which, however, should not be seen just
as aids in training (technical) translators, but also, and above all, as means of educating reflective students, and thus future well-
rounded professionals who can use personal judgment in the social construction process of making situated decisions to solve situated
problems (see Kiraly, 2009). The pedagogical value of corpora is highlighted by Anton (2001: 5; my emphasis), who points out that
corpus-driven learning activities can:

improve competence, increasing learners’ knowledge of the language and the culture, and their awareness of how the former is used in
the latter;

engage capacity, helping learners to develop their ability to use the language as a means of communication, both in reception and in
production;

increase autonomy, providing learners with learning instruments which they can exploit independently, and developing their ability to
do so.

Within the translation classroom, different types of corpora can be employed to achieve different, though closely related, objectives
(Bernardini et al., 2003: 6). For instance, monolingual corpora in the target language (either reference or specialized corpora depending
on the translation task being performed) can help translators choose translation options that are appropriate to the target communicative
function(s) and addresses. On the other hand, comparable bilingual corpora (i.e. collections of texts originally produced in the
respective languages which are similar in that they share some common features, including text type/genre, domain, topic, and
publication date/span; see Bowker & Pearson, 2002: 93) can help translators gain a better understanding of both the source and the
target text (ST and TT; see § 2 and 3 below), and parallel corpora (i.e. collections of original texts aligned to corresponding
translations; see Bowker & Pearson, 2002: 92) can make it possible for them to observe the strategies other translators have opted for.

Within the translation process itself, corpora of different kinds can support translators in the various stages of their endeavour, i.e.
before, during, and after the translation proper. As explained in Anton (2000: 22), in the preparatory stage, specialized monolingual
corpora (language (SL) and language of translation (TL)) may be used respectively to generate concept and idea corpora: [1]

Finally, in the editing stage, TL monolingual corpora may help (apprentice) translators improve the internal cohesion and coherence of
the TT (ibid.: 26).

Today, translators are in an incredibly privileged, and at the same time tricky, position, in that most of the resources they need are
available through the Web. The Web itself can be considered a huge corpus, or rather a “corpus shop” (Bernardini et al., 2008: 10), in
which an astounding amount of readily available (but not necessarily relevant or even reliable) information can be found. To find their
way through this jungle of facts and texts, students need guidance and support, so that they can fully realize what they can – and
cannot – do with the Web, especially for translation purposes. To this end, involving students in the activities of corpus design and
compilation may be a way of offering them a bottom-up perspective on both the potential and limitations of corpus data and methods
(Bernardini et al., 2003: 11). As argued by Fantinuoli (2003: 56), the challenge seems to be a twofold: first collecting the “right material”
for the translation purposes at hand; second, knowing how to use it, which requires translators to be “highly competent in textual and
stylistic analysis and, in addition, to be computer- and software literate at a fairly advanced level” (ibid.).

In this article I illustrate a series of classroom activities which were part of a module of corpus-assisted translation conducted at the
University of Siena and were aimed at helping students familiarize themselves with both text (type) analysis and corpus-based tools. In
§ 2, I sketch the contents, underlying rationale, structure, and methods involved, while in § 3 I focus on the outcomes of the module,
describing specifically with the assignments students were asked to complete, and discussing examples taken from the submitted projects.
Finally, in § 4, I offer some concluding remarks on the added value of corpora in translator education.

2. A corpus-assisted translation module: overview and first guided activities

The course illustrated in this section is a module of corpus-assisted translation conducted between 2010 and 2011 at the Arezzo campus
of the University of Siena. The module was offered as part of a one-year course in English language and translation addressed to third-
year Italian undergraduates of the BA programme in Languages and Intercultural Communication, who were studying two foreign
languages (classical and modern) and one of the following foreign languages (German, English, French, or Spanish). The lessons (24 hours over six weeks) were mainly designed to give
the circa 20 attending students the opportunity to gain some hands-on experience of corpus-aided translation. As most students were not
familiar with corpus methods and resources, they were first guided through ad-hoc corpus searches using English and Italian corpora
from the Siena University corpus collection, with the aim of clarifying the key concepts of concordance and collocation and introducing
them to general reference language corpora. As a second step, students, working in small groups, performed searches of technical words
and phrases in parallel English-Italian corpora and of linking expressions in a monolingual English corpus of academic discourse, to
familiarize themselves with the language used to talk/write about a specific subject area, i.e. language for special purposes (LSP). Finally, based on a set of assignments they had to complete as part of their final assessment, students were shown how to design and compile their own specialized comparable corpora (semi)automatically from the Web (see § 3.1 and § 3.2 below).[2] In what follows I describe, lesson by lesson, the activities conducted in class with the students, which in my view generated significant insight into, and significant questions about, language use, and which should be easily replicable with other undergraduate students.

In lesson 1, after introducing the students to corpora and their main features, some initial guided searches were performed using the Leeds collection of Internet corpora, i.e. large general-purpose corpora created from the Internet starting from automated search engine queries (Sharoff, 2006).[3] Students were asked to click on the “Italian” radio button in the main query window, type in the word “polemica” and click on the “Concordance” radio button without changing the default parameters. They were then asked to observe the resulting concordance lines to help them familiarize themselves with the KWIC (key-word in context) format and were shown how to sort the results by left/right to see which words tend to co-occur, i.e. collocate, with “polemica”. By looking at the immediately preceding and following co-text, students soon realized that the examples listed included occurrences of “polemica” as both a singular feminine noun (meaning “controversy” or “debate”) and a singular feminine adjective (meaning “polemical”). They therefore asked if there were ways of refining the initial query to discriminate between word classes, which led to an explanation of how to query the corpus using the appropriate search syntax, thus producing separate concordance outputs when looking for “polemica” as an adjective or a noun (fig. 1). Students were then shown how to use the “Collocations” function of the Leeds corpora interface to obtain a list of the most frequent collocates of the noun “polemica” and were encouraged to observe how results changed when the range of left and right context was adjusted.

When asked to look for the word “opportunità” (“opportunity”) and its most frequent collocates, students found out that a very common phrase is “pari opportunità” (i.e. “equal opportunities”). When they were invited to suggest synonyms for “opportunità” and came up with “possibilità” (“possibility”) and “occasione” (“chance”), they pointed out that phrases like “pari possibilità” or “pari occasioni” cannot in any way be considered equivalent to “pari opportunità”. This observation provided an opportunity to reflect on what must be taken as a unit of meaning, or lexical item (Sinclair, 1998): as the students observed, if it is true that the meaning of this specific phrase can be predicted from the meaning of its parts, i.e. it is compositional (see Manning & Schütze, 1999: 184), the phrase is not freely modifiable, and, above all, its components are not substitutable (ibid.).

Lesson 2 provided the opportunity to illustrate in more detail concepts related to collocation, i.e. colligation, semantic prosody and lexical bundles.[4] While collocation was exemplified by looking at the relationship between a search word (or node) and individual words occurring in its proximity, colligation was defined as the representation of collocates by grammatical categories and the relationships between these categories. Colligation was observed by looking at the concordance lines of a rather formal Italian verb, namely “ottomperare” (to obey/to comply with), which also provided an opportunity to show how wild cards (the symbol “%” in the case of the Leeds corpora) can be employed to retrieve all forms of a lemma (see fig. 2). Students were quick both at noticing that the verb “ottomperare” is used overwhelmingly within a specific syntactic pattern and at determining what this pattern is, i.e. SUBJ+VERB+prep (+a) +OBJ.
Semantic prosody expresses the speaker’s stance towards what s/he is talking/writing about (Hunston & Thompson, 2000: 5), thus being at a further level of abstraction compared to collocation and colligation and affecting longer stretches of discourse (Partington, 2004: 151; Sinclair, 2004: 178). In class, semantic prosody was discussed by observing concordance lines (and the list of collocates) generated by searching for the adjective “dilagante” (“rampant”). The results produced prompted the students to highlight that the Italian adjective (like its English counterpart) tends to co-occur with abstract nouns or noun phrases that define social, economic or political phenomena, behaviour or trends (e.g. “corruzione”, “crisi economica”, “vulgare”, respectively “corruption”, “economic crisis” and “vulgarity” in English), and that it tends to have a strongly negative connotation. In fact, many co-occurring words were nouns ending in -ismo, a very productive Italian suffix used to refer to morally condemnable attitudes, habits or actions (see fig. 3).

The next step was to guide students to the discovery that certain words occur together as a set, and that this set is open to the addition of other words to complete its meaning and grammar. They were asked to search for “grado” (“degree” or “level”) and look at its collocates both left and right. Some pointed out that “in grado di” (meaning able of) is a recurring lexical bundle, as shown by the joint frequency of “grado”’s most common collocates left (“in”) and right (“di”) (see fig. 4) and noted that “in” and “di” select the meaning of “ grado” and vice versa.
The short, “low-stake” activities described so far, while moving from teacher’s “isolated” input (words and phrases out of context) rather than from authentic texts (as is instead the case for the activities described in § 3.1 and § 3.2), proved conducive to enhancing students’ analytical skills, in that they triggered observations on language use starting from an unprung look at concordance lines (students were only asked to observe the words searched and their co-text left and right, without any further instruction). By the end of the second lesson, not only had students discovered that lexical items can correspond to more than just one (orthographic) word, but they had also realized that they include, besides their invariable part (or core), a that includes (collocation, coordination, and semantic prosody) that determine the items’ grammatical, semantic, and pragmatic realisation. Clearly, awareness of these features is vital for translators, who need to identify — in context — all the components that make up “functionally complete” units of meaning (Tognini Bonelli, 2000: 153) and then strategically define translational units (ibid. : 162).

The notion of translational unit gave me the opportunity to introduce parallel corpora in lesson 3. Here the aim was to show how this type of resources can provide not only translation candidates at lexical level, but also a repertoire of strategies adopted by actual translators to deal with non-equivalence at word level. To do this, we used the Opus corpus, to search, for instance, for “grado” and compare Italian concordance lines containing “in grado di” with their aligned English functionally equivalent units, finding that the Italian phrase can be translated either with the modal “can” or the phrasal modal “be able to”.[6]

The rest of lesson 3 was devoted to giving an overview of resources for translation that are available online for free, including monolingual and bilingual dictionaries of students’ working languages, as well as terminological databases (e.g. IATE),[7] which students later reported as helpful to complete course assignments. Among other things, students had the chance to familiarize themselves with corpora other than the Leeds collection of Internet corpora (for instance the BNC and the COCA) and with the freeware concordancer AntConc, which they employed for offline corpus consultation.[8] AntConc was appreciated by students for its user-friendly interface, particularly for the possibility of focusing the results on collocates in different positions left and right of the node in different colours, and for the word list function, a powerful tool listing lexical bundles based on their frequency. AntConc was used to perform queries on an offline subset of ITWAC (see note 3) and on a corpus of academic English.[9] The queries run in class on this corpus, while not immediately relevant to coursework, provided a smooth transition into the realm of language for special purposes and specialized language corpora, which students were asked to compile and consult for their translation projects (§ 3), and triggered some interesting observations about the features of LSP. As pointed out by Bowker & Pearson (2002: 26-27), although specialized vocabulary is probably LSP’s most striking element, a specialized language is also characterized by special collocations and specific stylistic features, such as the ways in which information is arranged. For instance, students queried the Groom corpus to investigate the use of certain linking expressions in academic discourse. One of them noted that the gerund “speaking” is rather productive in combination with verbs like “commonly”, “generally” and “strictly” and argued that the resulting expressions, which tend to occur in sentence-initial position, are employed to highlight transitions in the writer’s line of argument.

3. A corpus-assisted translation module: preparation of assignments and discussion of final projects

Lessons 4-6 of the module were devoted to designing and building bilingual specialized comparable corpora which were then used by the students to a) translate (in groups) from English into Italian; b) translate (individually) from Italian into one of their foreign languages; and c) prepare group presentations on relevant issues of corpus compilation and consultation for translation purposes. The English ST was about the traditional “Stinking Bishop” cheese from Gloucestershire, and students were also asked to write an accompanying translation commentary, while the Italian ST was about the similarly typical “Abbucciato” cheese from the Arezzo area. Foreign language instructors were involved in the correction and feedback regarding the translations into the students’ foreign languages. In lesson 7, students started working on the group translations, for which feedback was provided during lesson 9. In lesson 8, students gave the oral presentations mentioned in c), focusing on the individual translations.

The core of the module thus consisted in guiding students through the various steps needed to design and compile a corpus. As mentioned, the DIY (aka ad hoc or disposable) corpora built by students were specialized comparable corpora. The value of comparable corpora resides in making a vast number of language-use examples occurring in multiple contexts readily available for analysis in both the SL and the TL. The bilingual specialized comparable corpora were the English ST for the translation assignment described in a) and the Italian ST for the translation assignment described in b) above. The former, a 429-word description of the Stinking Bishop cheese (including its history and production process), was taken from the website of an online shop specializing in these produced in small forms using traditional methods: https://www.teddingtoncheese.co.uk/acatalog/de339.htm. The latter, a 593-word text on the “Pecorino Abbucciato Aretino” (including its history, a rather technical description of its production process, and its nutritional properties), was formerly available at: http://promozione.provincia.aretzo.it/materiale/ABBUCIATOT.pdf.

In § 3.1 some of the issues that emerged during corpus design and compilation will be discussed, while in § 3.2 the students’ projects will be selectively reviewed.

3.1. From corpus consultation to corpus compilation

The first bilingual specialized comparable corpus was constructed in lesson 4 using TextSTAT, a programme that includes tools for text analysis (similar to AntConc) as well as a Web crawler able to read files directly from the Internet and save them as a plain text file. First, students engaged in some preliminary activities that can be subsumed under the general rubric of corpus design.
Following the guidelines provided by Bowker & Pearson (2002: 45-54) they were invited to think about criteria for corpus compilation based on their specific needs and project goals, i.e. the translation of the Stinking Bishop text. It took students quite a long time to agree on all criteria (shown in tab. 1), which triggered an interesting debate on how difficult and time-consuming it may be to obtain a sample of specialized texts/language. Students were particularly keen to identify text communicative functions (informative vs promotional) and text addressees (tourists, consumers, etc.), as well as to argue for the need to include only full texts (as opposed to extracts) and to debate issues of source reliability. The latter topic aroused much discussion, and the class literally split into two opposing parties over the claim that texts should/should not necessarily be written by native speakers (although all students agreed that this may be rather difficult to ascertain).

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Suggestions agreed upon by students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Size</td>
<td>10,000 words (but expandable)</td>
</tr>
<tr>
<td>Number of texts</td>
<td>10-12 (5-6 per language) by different authors</td>
</tr>
<tr>
<td>Medium</td>
<td>Written</td>
</tr>
<tr>
<td>Subject</td>
<td>Cheese (manufacturing of)</td>
</tr>
<tr>
<td>Text type</td>
<td>Descriptive, semi-specialized (no recipes!) produced by experts for non-experts</td>
</tr>
<tr>
<td>Authorship</td>
<td>Online travel guides/company (cheese manufacturers) websites</td>
</tr>
<tr>
<td>Language</td>
<td>English and Italian originals</td>
</tr>
<tr>
<td>Publication date</td>
<td>Not relevant</td>
</tr>
</tbody>
</table>

Tab. 1: Wish list for the Cheese comparable corpus

Moving from theory to practice, students were divided into groups of three, each using a laptop, and were asked to download TextSTAT; they were then provided hard copies of the Stinking Bishop text to read individually and asked to jot down (again individually) five or six keywords, as well as their possible Italian equivalents resorting, if needed, to the online dictionaries and corpora they had used in previous lessons. The keyword lists thus obtained were then compared within each group, eventually resulting in a single word list. Among other words, students suggested the nouns “cheese”, “rind”, and “mould”, the adjectives “pungent”, “firm”, and “mature”, and the verb “pasteurise”. Each group googled various combinations of these words to find relevant English texts on the Web. Specifically, students were encouraged to look for five websites/pages, note down the corresponding URLs, read the contents of these sites/pages, and then describe the texts they found to the rest of the class, specifying which ones they would include/exclude in/from the corpus and why. This process was repeated for Italian texts starting from the students’ translation candidates of the ST’s keywords. Once everybody had the chance to speak and exchange keywords and URLs with her/his colleagues, each group proceeded to the actual compilation of the corpus through the very simple procedure supported by TextSTAT (see fig. 5).

In lesson 5, the TextSTAT concordancer was employed to run queries on the English and Italian subcorpora constructed during lesson 4. This time, unlike what had happened during the activities illustrated in § 2, students did not search out-of-context words suggested by their teacher, but autonomously selected words and phrases (e.g. the noun phrase “moisture content” and the adjective pair “soft and creamy”) from either the ST or the texts selected from the Web, which gave them a better sense of the patterns of meaning in the texts, including prosodies, as well as cohesion and coherence. Based on the observation of concordance lines, some students suggested revising the selection of webpages to harmonize the corpus in terms of text communicative function and style, and did so accordingly. The corpus compilation activity also prompted some questions and comments on very practical (technical) issues like preferred file format. Finally, a comparison between the searches performed on their DIY plain text corpora and the POS-based searches performed on the Leeds collection of Internet corpora triggered an interesting discussion on the added value of annotation, above all on the role of analytic metadata (Burnard, 2004) in making built-in linguistic information retrievable.

In lessons 5 and 6, students compiled bilingual specialized comparable corpora using the corpus-building tool known as WebBootCaT, which automatically downloads webpages starting from a list of “seed words”.[12] The two languages involved were Italian and one of the students’ foreign languages (English, French, German, and Spanish). While the students had to complete the translation from Italian at home, the corpus compilation activity preceding the task was conducted mainly in class in pairs and groups of three. Students were given hard copies of the Italian text of the Abbacciato Aretino, which took them a few minutes to read. As with the TextSTAT-based task illustrated above, they were invited to select five or six specific words from the text and look them up in the Italian corpus they
had built using TextSTAT as well as in a larger (reference) Italian corpus (see note 3), comparing concordance lines and collocations and noting down differences (if any). Further, they were asked to look up possible translations of the words chosen in online bilingual dictionaries and check these translation candidates in the foreign language in online monolingual dictionaries and corpora. These words were then used as seed words to create corpora in students’ foreign languages following the WebBootCaT procedure (fig. 6).

Once each group had obtained a bilingual comparable corpus, students were shown how to use word sketches in the Sketch Engine. A word sketch is a powerful tool that displays the search word’s colligations (fig. 7). In class, it was used to compare collocation patterns occurring in the disposable specialized corpora and in large general reference corpora such as the BNC.

Finally, I explained how to merge subcorpora with the Sketch Engine compiling function (which many of the students later did, after realizing their initial corpus was not specialized or large enough and thus needed to be adjusted/expanded) and how to share their corpora with the colleagues in their group, as well as with the module instructor.[13]

3.2. Translating with corpora
In this section I will discuss some of the projects submitted by students at the end of the module, specifically some of the commentaries accompanying the English-Italian translations of the Stinking Bishop text (assignment a), and some of the oral presentations describing the steps taken to complete the translation of the Abbucciato Aretino Italian text into the chosen foreign language (assignment b and subtask c). Although space constraints do not permit a thorough analysis of the projects, I will dwell on some of the issues raised by students, which show how they implemented the methods learned in class.

Starting from assignment a), in their commentary to the translation of the Stinking Bishop text, a first group of students dealt, among other things, with the translation of “washed rind” and “washing solution”, for which they found equivalents in the Italian section of the comparable corpus they had built with TextSTAT, where they observed the collocations of “lavata” and “lavaggio” (fig. 8). In an attempt to obtain further evidence, the students found the product information sheet of French “Époisses de Bourgogne”, which is
referred to in the ST as being similar to the Stinking Bishop. The information sheet on Epoisses provided additional evidence supporting the collocation “crosta lavata” as opposed to “buccia lavata” (“buccia” being another possible translation of the English “rind”), as shown by the highlighted extracts in figure 9.

In their commentary, a second group of students noted that the most difficult part of the translation of the Stinking Bishop text was choosing how to render “smells of old socks” in Italian. Having agreed on the fact that a word-for-word translation would not make the cheese very appealing to Italian-speaking potential buyers, they opted for “ha un odore particolarmente pungente” (has a particularly pungent smell), arguing that 1) pungent means strong without necessarily having a negative value, as shown by the collocations and word sketches of “pungent” from the BNC (fig. 7); and 2) the intensifying adverb “particularly” somehow compensates for the omission of the suggestive expression used in the ST.

Moving to assignment b), in the commentary to his English translation of the Abbucciato Aretino text, a student explained that one issue he had to face during the Italian-English translation was the term “struttura” referring to the physical consistency of the Abbucciato Aretino. Having realized that bilingual dictionaries were not of any help, he first turned to the IATE terminological
databank, where he narrowed his query to the “agriculture, forestry and fisheries” domain (fig. 10), and then checked what he believed to be the best translation option he had found, i.e. “body”, against the specialized comparable corpus he had compiled using the Sketch Engine WebBootCaT. Moving from the concordance lines of “body” therein, he traced a specialized English glossary on cheese (fig. 11), where he found the definition of the term “body”.

Fig. 10: English translations of “struttura” provided by IATE

Fig. 11: Collocations of “body” in one of the students’ DIY corpora

As we have seen, the strategies adopted by students to verify their translation candidates include not just consulting their comparable corpora, but, as a further step, examining additional comparable texts on cheese (see fig. 9 above). Some students also considered texts that were not comparable to their ST in terms of genre and intended readers but were nonetheless consulted and described in their commentaries as “authoritative sources of information as well as linguistic evidence”. Sometimes such texts were discovered by chance, as happened to one student, who, in looking for the best Spanish translation of “spino” (a long whisk used for stirring and separating curd from whey during the cheesemaking process) in various sources, including the cheese comparable corpus he compiled using WebBootCaT, bumped into the Spanish version of an official EU regulation on protected designation of origin, which referred specifically to one of the typical cheeses manufactured in Italy, and could thus confirm that the Spanish equivalent of “spino” is “cucharón”.

Another student, in explaining why she had not translated, or even thought of translating (e.g. by providing an explanatory periphrasis in a translator’s note), “Pecorino Abboccato Aretino” and other culture-specific terms into French, introduced the concept of realia – precisely ethnographic realia – quoting the work of some translation scholars she had had the chance to read (e.g. Osimo, 2004). She argued that “Pecorino Abboccato Aretino” is a culture-bound element for which no functional equivalent is available in French. Since French readers may not be familiar with this term and concept, she added, translating it would imply writing a periphrasis. However, she concluded, such a periphrasis is superfluous in a text that has the “Pecorino Abboccato Aretino” as its main topic and thus provides detailed information about it.

Lastly, a student explained how she had resorted to Google images to have a clearer idea of what “fuscelle” are (strainers used as cheese moulds allowing the cheese to drip) and, at a later stage, to make sure that the German “Käseformen” she had found in her comparable corpus referred to the same objects.
All students were very careful about identifying the translation problems encountered and describing the various steps of the progressive approximation process leading to their solution. What exceeded my expectations was the logical rigor, and at the same time the effectiveness, with which students integrated the various resources available to provide evidence supporting their translation choices. Finally, their procedural competence and metatranslational awareness clearly emerged in their accounts about using comparable corpora for translation purposes. In the following list, I have summarized the points made by students themselves in their presentations:

With comparable corpora, you can:

- learn something new about a specific domain/topic;
- understand the ST;
- look for equivalents, definitions and contexts of use in both the ST and TL;
- identify and reproduce the features of the specific genre/register in the TL;
- choose the “right” lexical items.[15]

4. Thought-provoking tools for problem-solving processes: concluding remarks on corpora and translation

Corpus-based learning activities have often been described as a journey of discovery (Bernardini, 2000) in which the teacher’s role is that of an facilitator (Aston, 2000: 42). The corpus work presented in this paper is no exception to this general statement, and the student’s projects illustrated in § 3.2 confirm the crucial role that corpus methods and tools play for the teaching and learning of translation. To be more precise, and in line with what was suggested by Varantola (2003: 59), DIY comparable corpora proved to be performance-enhancing tools that helped students to overcome the mismatch between competence and performance in translation, especially into their foreign languages. In this respect, one of the main benefits of comparable corpora can be considered that of “reassuring” novice translators about their choices (ibid.: 67). In other words, if it is true that translation is a problem-solving process par excellence, in which translators are called upon to make informed, accountable decisions, then comparable corpora can be said to strategically support these decisions. In fact, judging from the way in which students documented their decision-making process in the commentaries and presentations discussed in § 3.2, corpus resources, especially the DIY corpora they built, did more than just reassure them. On the one hand, these corpora were used to generate and test hypotheses as to the interpretation of the ST and as to appropriate translations in the TT; on the other hand, they were used to improve the quality of both text interpretation and translation, capturing subtleties of the ST and producing (or approximating) native-like translations.

In both assignments, students proved able to combine a close qualitative analysis of the STs – at the level of language, discourse and the socio-cultural environments in which the texts were produced – and the “larger picture” of quantitative data provided by both specialized and general reference corpora. In that students were brave enough to move beyond the clear-cut distinction between grammatical and lexical categories, we may go as far as saying that they embraced a corpus-driven approach (see Tognini Bonelli, 2001: 84-85). In addition, the uses they made of corpus resources they had become familiar with or created themselves were not predetermined by their teacher but emerged during their decision-making processes in sometimes unexpected ways. Overall, students’ observations on their translation assignments denote a considerable metalinguistic and metatranslational awareness (which goes hand in hand with the high linguistic-translational competence shown in their TTs), as well as considerable critical thinking and an ability to apply acquired skills to new problems and to learn new skills (Bernardini, 2000: 85-86) – in other words, to be autonomous learners. The results obtained support the case for a translation pedagogy in which corpora establish a negotiation space between teachers and learners (Aston, 2001: 44) that increases learners’ autonomy, giving them a “sense of ownership” (Sinclair, 2000: 7) and responsibility. Ultimately, and from a very practical point of view, it is my conviction that teaching and learning experiences of the kind described in this paper are replicable with limited resources and even with a larger number of students. As claimed by Zanca (this volume), there is an incredible amount of “online stuff” that can be used (for free) both in the classroom and outside of the classroom and, as shown in § 3.2, it is amazing how much students can learn, both intentionally and incidentally, when using it.

Acknowledgments

The present paper could not have been written without the outstanding interest and motivation demonstrated by the students of the BA in Languages and Intercultural Communication who attended the corpus-assisted translation module the paper draws on. I am also indebted to Lauraie Anderson, Lucia Cacci, Daniele Corsi, and Barbara Innocenti for the time and effort they put in the correction of students’ translations into English, German, Spanish, and French respectively. Finally, I would like to thank the editors of this volume for giving me the chance to share such a rewarding teaching experience.

References


Notes

[1] Specialized corpora contain “texts dealing with a particular subject area and written by experts for a varied readership (experts to experts, experts to students, expert to laymen)” (Kübler, 2003: 29).

[2] The main references throughout the module were an introductory volume to corpus linguistics (Spina, 2001) and a practical guide to creating and using corpora for translation (Bowker & Pearson, 2002).

[3] Available at http://corpus.leeds.ac.uk/internet.html. The Italian section of the Leeds corpora (iWiC) was mainly used, together with “La Repubblica”, a 380 million token corpus of articles published between 1985 and 2000 in the homonymous Italian daily newspaper. I would like to thank Eros Zanchetta at the University of Bologna for providing access to the SSLMIT Dev (now CoLiTec) corpora, including “la Repubblica” (https://corpora.dipintra.it/), and to an offline subset of iWiC.

[4] Lexical bundles are also called n-grams, multi-word units, or clusters.

[5] On the evaluative prosodies of suffixes see Partington (this volume).


[8] Free search interfaces for the BNC (British National Corpus) and the COCA (Corpus of Contemporary American English) corpora are available respectively at: https://corpus.byu.edu/bnc/ and https://corpus.byu.edu/coca/; AntConc is downloadable from: http://www.laurenceanthony.net/

[9] The corpus of academic articles was compiled by Nicholas Groom (Centre for English Studies, University of Birmingham) and contains nearly thirteen million words taken from sixteen international journals of economics, history, and sociology published between 1999 and 2003.


[12] WebBootCaT is part of the Sketch Engine corpus manager software available at https://www.sketchengine.co.uk/. To avail themselves of this tool, the students opened a 30-day trial subscription to the Sketch Engine full package. A free version of the WebBootCaT, called BootCat front-end, is downloadable from http://bootcat.dipintra.it


[14] Although the reader is left to wonder whether and to what extent the original “smells of old socks” can be considered appealing to English-speaking buyers.

[15] When asked to clarify the adjective “right”, the student who used it said that it qualifies a lexical item as being appropriate to the context in which it is used, as shown by the occurrences of that specific lexical item in the specific context of use under investigation.