

**INTERMODAL CORPORA
AND THE TRANSLATION CLASSROOM:
WHAT CAN TRANSLATION TRAINERS AND TRAINEES
LEARN FROM INTERPRETING?**

ADRIANO FERRARESI*

University of Bologna, Italy

Abstract

This contribution focuses on didactic applications of intermodal corpora, i.e. corpora featuring interpreted and translated language. It relies on EPTIC, a multiple-translation and intermodal parallel corpus containing EU Parliament plenary speeches in Italian and English. The peculiar nature of EPTIC allows the investigation of a set of translational alternatives which are distinguished by modality- and task-based constraints (written vs. oral, translated vs. interpreted). To exemplify the potential of such corpus evidence, teaching activities focusing on collocations are proposed that encourage students to reflect on the decision-making processes involved in the slower-paced, reflective task of translation, vs. the faster, more automatic one of interpreting. A method is also described that can facilitate the selection of relevant didactic examples.

Keywords: *multiple translation corpora, intermodal corpora, collocations, corpus-based teaching materials*

1. Introduction

Exploring the potential of different types of corpus setups—“from large established monolingual mixed reference corpora [...] to corpora of learner translations” (Aston, “Foreword” IX)—has long been a cornerstone of research on corpus-informed translation pedagogy. Among corpus setups, Multiple-Translation Corpora (MTC) have emerged as an appealing resource in the hands of translation scholars, trainers and students. By bringing together multiple target texts (TTs) of a single source text (ST), MTC make it possible to observe and compare the strategies adopted by different

* adriano.ferraresi@unibo.it

translators, be they professionals or trainees, when faced with the same input. Depending on one's research and/or pedagogical aims, they can provide insights into regularities and variation in the product of translation and prompt reflections on issues such as motivations underlying similar or divergent translation choices (Malmkjaer "Love Thy Neighbour"; Munday), or their contextual appropriateness (Castagnoli "Exploring Variation").

Along these lines, this contribution presents applications of the recently created European Parliament Interpreting and Translation Corpus (EPTIC; Bernardini, Ferraresi, Miličević). Conceived as an extension of EPIC (Russo *et al.*), EPTIC is quite unique in featuring English and Italian plenary speeches by members of the European Parliament in four *pseudo-parallel* aligned versions (see Section 3): transcriptions of interpreted speeches and their independently produced translated version, as well as the corresponding oral and written STs. The presence of spoken and written components and their alignment to each other make EPTIC an *intermodal* parallel MTC. The translational alternatives observable in EPTIC are not only produced by different language professionals, but are also distinguished by modality- and task-based constraints (written vs. oral language production, self-paced translation vs. real-time interpretation).

To illustrate the potential of corpora like EPTIC as pedagogical tools, a case study will be presented in which this corpus is used as a basis to design teaching materials focusing on a specific type of phraseological items, i.e. collocations. As well as being crucial to language learning and use in general (Jones and Sinclair), these have also been focused upon as an area of difficulty in both specialized and non-specialized translation (Biel; Mauranen).

A method will be described that allows teachers/researchers to isolate instances in which the lexical choices made by translators and interpreters are most divergent in terms of frequency and idiomaticity, signalling potentially problematic ST passages. Pedagogical applications of this type of corpus evidence will then be presented in the form of teaching activities that encourage students to reflect on the decision-making processes involved in the slow-paced, reflective task of translation, vs. the faster, more automatic one of interpreting, and on the available (and preferred) alternatives selected by experienced professionals engaged in these related yet different modes of language mediation.

The contribution is structured as follows. Section 2 reviews previous work on corpus-based studies of variation and invariance using MTC, and discusses the relevance of intermodal corpora for related research and teaching purposes. Section 3 presents the EPTIC corpus, and Section 4 the case study on phraseological variation in translation and interpreting, as well as two learning tasks based on intermodal concordance lines. Section 5

concludes by evaluating the advantages and limitations of using EPTIC in the translation classroom, and suggesting further ways in which it can be exploited.

2. Background: MTCS and Intermodal Corpora in Translation Research and Teaching

In his “minimalist definition” of translation competence, Pym claims that the two fundamental abilities that trainee translators should develop are “the ability to generate a series of more than one viable target text [...] for a pertinent source text” and “the ability to select only one viable TT from this series, quickly and with justified confidence” (“Redefining competence” 489). Similar claims, highlighting that variation lies at the very core of the translation process, are pervasive in the literature on MTC. The added value of MTC with respect to better-established “translation-driven” corpora, like comparable or parallel ones (Zanettin), resides precisely in the possibility they afford to assess the extent of such TT variation, and to identify the ST triggers that are likely to generate it.

Malmjkaer (“Pseudosubversive Use”) exemplifies the point by exploiting a MTC in the framework of what she calls “pseudosubversive” uses of corpora. According to the author, the focus on trends and regularities that is inherent in traditional corpus-based analyses can be a hindrance “to act[ing] creatively”, and “yet much translational activity requires creativity” (Malmjkaer, “Pseudosubversive Use” 121). Based on a corpus of published English translations of Hans Christian Andersen’s *The princess on the Pea*, she argues that observation of multiple equivalents helps to draw the students’ attention to a range of different solutions, rather than the most obvious one(s) only. These solutions stand in need of corroboration in other reference materials (in Malmjkaer’s example, a corpus of the complete work of H.C. Andersen), but the availability of a range of translation equivalents constitutes a valuable backdrop against which the appropriateness of individual options, however unique and creative, can be better evaluated.

A critical problem with MTC, which also represents one of the main limitations to their spread as a research and pedagogic tool, is that they are difficult to assemble outside the domain of literature (Bernardini). Unlike literary texts, which may be retranslated in time (see e.g. Ji and Oakes for further examples of literary MTC), technical, scientific and otherwise specialized texts are rarely translated more than once in real-world professional settings, and even when they are, translations are unlikely to be easily accessible. The work of Johansson and Castagnoli epitomizes two ways in which this problem can be tackled.

Johansson (“Why change subject”) describes a MTC featuring two English STs (a short story and a scientific article), and 10 Norwegian TTs of each, which were commissioned *ad hoc* to as many translation professionals. Turning to experimental data of this kind may raise doubts concerning their status as authentic, “naturally occurring language” (McEnery, Xiao, Tono 4). However, this strategy has emerged as one of the few viable options to build specialized MTC, and one which is adopted in other research settings where (readily) available non-experimental data are scarce, as is also the case with intermodal corpora (see *infra*). In terms of applications, the corpus setup allows Johansson to observe syntactic changes in sentence subjects in translation from English to Norwegian, telling apart shifts motivated by systemic lexical or syntactic differences between the two languages, through to shifts that can be explained by stylistic preferences of individual translators.

The use of multiple learner translations is suggested by Castagnoli (“Exploring Variation”), among other scholars, as an alternative to resorting to experimental data (see also Bowker and Bennison; Munday). Castagnoli’s *learner* MTC is used as a basis to investigate explicitation and ST interference patterns focusing on interclausal connectives. Like in Johansson’s study, regularities and variation in translation equivalents are interpreted in the light of obligatory (language-induced) vs. non-obligatory (translation-induced) shifts. Furthermore, the fact that TTs are produced by learners makes this corpus setup a promising one to assess appropriateness of TTs in relation to translation competence: besides being of descriptive interest in themselves, similar observations make the ideal basis for didactic activities, e.g. to encourage students to “practic[e] their evaluation and revision skills” (Castagnoli, “Investigating Competence” 14).

Despite methodological hurdles in their construction, MTC have thus proved rewarding for descriptive and applied studies of variation and invariance in translation. In what follows it is argued that intermodal corpora, i.e. corpora including the output of written translation and oral interpreting, can add a new dimension to this line of research.

Pioneered by Shlesinger (“Corpus-based Interpreting”), the intermodal corpus setup can be viewed as a result of a wider tendency to bridge the divide between translation and interpreting studies.¹ The underlying idea, as remarked by Gile (23), is that

¹ In a wider perspective, due to sharing features with, e.g., second language production and edited language, translation and interpreting are increasingly seen as instances of *constrained* varieties, i.e. types of language use that are subject to similar cognitive and/or social restrictions (Kruger).

since translation and interpreting share so much, the differences between them can help shed light on each, so that besides the autonomous investigation of their respective features, each step in the investigation of one can contribute valuable input towards investigation of the other.

Miriam Shlesinger's work has been pivotal in shaping the research agenda of corpus-based explorations at the interface between translation and interpreting. In particular, she first put forward the idea of extending monolingual comparable corpora of interpreted and original speeches by adding translations, arguing that this would allow one to discern "the characteristics of interpreting qua interpreting" (Shlesinger, "Corpus-based Interpreting" 488). In later work she acknowledged that this would also allow one to observe "differences between the oral and written modalities of translation, [and] to observe the effects of the ontology variable (original vs. translated) as well" (Shlesinger and Ordan 47), hence redressing the balance between interpreting- and translation-centred perspectives.

To the best of my knowledge, Shlesinger ("Towards a Definition") describes the very first *intermodal MTC*: this features the translational and interpretational output produced in experimental settings by six Hebrew translators/interpreters, who first rendered an English ST orally, and then, three years later, rendered the same ST in writing. In an attempt to overcome the limitations of this corpus, namely its small size (16,000 tokens), and the experimental nature of its data, Shlesinger and Ordan ("More Spoken") collect a larger intermodal corpus containing Hebrew interpreted and translated TTs produced in authentic professional conditions, as well as original Hebrew speeches. Unlike in the earlier corpus, however, translations and interpretations do not share the same STs, i.e. they are instances of comparable, rather than parallel texts. As was pointed out for MTC in general, obtaining multiple (intermodal) TTs sharing the same ST *and* produced in authentic conditions has emerged as a hard-to-achieve goal.

The two corpora just described are used by Shlesinger and Ordan to carry out intermodal and monolingual comparable analyses focusing on different measures of lexical variety, part-of-speech distributions and lexical features, e.g. the use of formal vs. colloquial terms. Based on these mainly quantitative comparisons, the authors conclude that interpreting can be seen as "an extreme case of translation, one in which those features that have been found to distinguish between translated and original texts [...]—e.g. simplification, lower type-token ratio, leveling—are found to be all the more salient" (Schlesinger and Ordan 54).

An intermodal corpus is also built by Kajzer-Wietrzny (*Interpreting Universals*) drawing on oral transcripts and written reports from plenary sessions of the European Union (EU) Parliament. The corpus features interpreted and translated texts into English from different languages (French, Spanish, German and Dutch), as well as original oral texts in English. Even though this setup would have allowed intermodal comparisons, the author restricts her analysis to the oral sub-corpus: her results are therefore of limited relevance in this context.

Finally, Bernardini, Ferraresi and Miličević (“From EPIC to EPTIC”) also exploit EU plenary session speeches for the construction of EPTIC. As mentioned in Section 1, EPTIC is at the same time a parallel MTC (featuring translated and interpreted TTs for the same ST), an intermodal corpus (featuring oral and written texts), and a comparable one (featuring mediated and original texts; see also Section 3). Replicating the well-known methodology proposed by Laviosa (“Core patterns”) to study lexical simplification, the authors focus both on the monolingual comparable level, finding that the mediation process reduces complexity in both interpreting and translation and in both language directions, and at the intermodal level, suggesting that interpreters simplify more than translators do. In two follow-up studies, Ferraresi, Bernardini and Miličević (“Collocations across languages”) and Bernardini (“Intermodal corpora”) broaden the perspective by adding parallel analyses of strategies underlying translators’ and interpreters’ production of collocations. Like Schlesinger and Ordan’s work, these studies have a mainly quantitative focus, and do not include a multi-parallel dimension linking STs to the corresponding TTs, and interpreted and translated TTs to each other (but see Bernardini 144-145). The case study presented in Section 4 represents one of the first attempts to explore this perspective afforded by EPTIC, and to make suggestions about its applications. Before that, a more thorough description of EPTIC elucidating its peculiar setup is in order: this is provided in Section 3.

3. The EPTIC Corpus

The EU Parliament materials have been widely used in corpus and computational linguistics due to the manifold advantages they offer for research, including their authenticity, availability, and the number of featured languages (Tiedemann). Importantly for translation and interpreting studies, the institutional setting also guarantees that translators and interpreters involved are accredited professionals.

EPTIC builds on another well-known, EU Parliament-derived corpus, namely EPIC (European Parliament Interpreting Corpus; Russo *et*

al.), which was created by transcribing a number of original plenary speeches and their interpretations into selected languages (the combinations represented in EPIC are English↔Italian↔Spanish). During the creation of EPTIC, EPIC's transcripts of interpreted speeches and their STs were paired with the corresponding translated versions and respective STs. This was made possible by the fact that, for each plenary session, the EU Parliament publishes so-called "verbatim reports" of proceedings consisting of transcripts of the speeches and their translations into all EU official languages.²

Despite being called verbatim, the reports usually undergo an editing process: hence the claim that texts in EPTIC are in fact *pseudo*-parallel. The editing process may include addition of punctuation and removal of context-related comments, and correction of mistakes such as false starts, unfinished sentences or mispronunciations. Table 1 presents an example in which the original speech was delivered impromptu and underwent rather substantial changes. In the case of read-out speeches, which constitute the large majority in EPTIC, these editorial interventions are much more limited (Bernardini, Ferraresi, Miličević).

Table 1. First lines of EPTIC text t23 (Speaker: M. McDowell, Date: 11/02/2004): transcribed speech and corresponding verbatim report.

Transcript of speech	Verbatim report
<i>Ehm can I ask the pa- the Parliament also there are two outstanding opinions which ehm for the purposes of our presidential agenda we look forward to receiving from the Parliament ehm before ehm in in relation to ehm getting through our particular agenda //</i>	<i>There are two outstanding opinions which, for the purposes of getting through our presidency agenda, we look forward to receiving from Parliament.</i>

Concerning the relationship between STs and TTs, translation of the proceedings is based on the verbatim reports, and the process is performed without any reference to the interpreters' outputs, as confirmed by several EU officials consulted on this issue.

The language combination in EPTIC is English-Italian, including translations/interpretations in both directions. Considering all its sub-corpora, comprising simultaneous interpretations paired with their STs, plus

² Verbatim reports are available for plenary sittings held before 2012, when the EU discontinued this service. See <http://www.europarl.europa.eu/RegistreWeb/search/typedoc.htm?language=EN>

the corresponding translations and STs, EPTIC has a total of eight components. Its rather complex structure is shown in Figure 1: the *st-* and *tt-* prefixes indicate source and target texts, the *-in-* and *-tr-* affixes interpretations and translations, and the *-en* and *-it* suffixes the language the texts are in—English and Italian respectively.

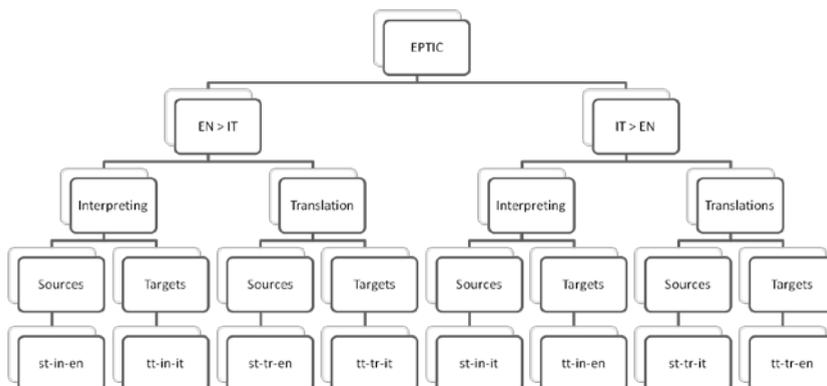


Figure 1. EPTIC structure

The construction of EPTIC is an ongoing project. At the moment, the corpus as a whole contains 568 individual texts, for a total of 250,093 words (disregarding truncated words in interpreted texts). The bigger, English=>Italian, portion contains four versions of 81 texts, while the smaller Italian=>English one has four versions of 61 texts. The majority of them are directly based on EPIC, i.e. they derive from the EU Parliament speeches delivered in February 2004, except for 44 texts coming from the March, April and July plenary sessions of the same year. Work is under way to add more recent texts from 2011, as well as a French component, which will turn EPTIC into a trilingual, intermodal parallel MTC.

In the present study, only the Italian translated and interpreted TTs and the respective English STs are used, amounting to a total of 119,548 words; the sizes of the individual sub-corpora are shown in Table 2.

Table 2. Size information for the Italian (TTs only) and English (STs only) sub-corpora

Sub-corpus	N. of texts	Word count
tt-in-it	81	33,675
tt-tr-it	81	36,876
Total	162	70,551
st-in-en	81	41,869
st-tr-en	81	36,685
Total	162	78,554
TOTAL	324	149,105

EPTIC is part-of-speech tagged and lemmatized using the TreeTagger,³ and indexed with the NoSketch Engine (Rychlý).⁴ Crucially, each text within the corpus is aligned at sentence level with its ST/TT, and with the corresponding text in the other modality (oral/written). Rich metadata are also encoded in the corpus and can be used to perform complex queries based on specific characteristics of texts and/or the speakers who delivered them. Available metadata, part of which were inherited from EPIC, include speech duration, delivery speed, delivery type (read, impromptu or mixed), text topic and length, as well as speaker and interpreter details (e.g. their gender and native language).

4. Using EPTIC to Design Translation Teaching Materials Focusing on Collocations

This Section presents a case study that illustrates the potential of EPTIC as a source of classroom materials focusing on Italian collocations. Section 4.1 describes a method for the principled selection of texts and concordance lines displaying variation at the phraseological level; a categorization of the insights that can be derived from corpus evidence thus extracted is presented in Section 4.2, and its didactic applications in Section 4.3.

4.1. A method for spotting phraseological variation

The proposed method exploits the nature of EPTIC as a MTC to identify cases in which translators and interpreters make phraseological choices characterized by different levels of conventionality with regards to target language norms. Divergences in MTC, as suggested by the body of literature reviewed in Section 2, may signal ST items that do not have readily available phraseological equivalents in the target language, or TT passages where idiosyncratic equivalents were selected. These cases, it is argued in the next two Sections, constitute interesting material to use when creating translation teaching activities.

For the purposes of this exemplificatory study, the notion of phraseology is narrowed down to a specific type of phraseological items, i.e. collocations, defined as combinations of lexical words that occur together

³ <http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

⁴ The corpus will be made freely available for consultation from the page: <http://corpora.dipintra.it>. This page also makes available the general language corpus of Italian *itWaC* (see Section 4), and other corpus resources developed at the Department of Interpreting and Translation of the University of Bologna, Italy.

more often than predicted by chance within a pre-defined part-of-speech (POS) pattern (Evert). While very limited with regards to the various types of phraseological items which should be brought to students' attention (including idioms, culture-bound set phrases, etc.; see Colson), this definition offers the advantage of providing objective, frequency-based and structural criteria for their identification in corpora. In addition, these criteria can be explicitly mentioned to students as guiding principles to tell apart standard vs. less standard phrases in a given language, along the lines, e.g., of Stewart ("Safeguarding").

The collocations focused upon are instances of pre- and post-modification of noun phrases in Italian translated and interpreted texts; the methodology, however, could easily be replicated for English TTs, and for other phraseologically productive syntactic structures (Evert). Specifically, only word sequences occurring in *adjective+noun*, *noun+adjective* and *noun+[preposition+]noun* patterns are considered. Table 3 reports examples of these word sequences; word-for-word English equivalents of Italian expressions are provided in italics.

For each word sequence extracted from the EPTIC Italian TTs, frequency data are obtained from an external reference corpus, i.e. itWaC, one of the largest available corpora for Italian (Baroni *et al.*). Resorting to a benchmark corpus like itWaC makes it possible to obtain more reliable statistics on collocation status, as well as to classify word sequences based on their frequency in Italian *in general*, and thus relate "observations to the target language usage norm, as approximated by a large reference corpus" (Bernardini 139).

Each EPTIC word sequence is classified according to two parameters: rarity/unattestedness (frequency lower than 2 in itWaC, vs. frequency equal or higher than 2), and strength (Mutual Information (MI) equal or higher than 5 in itWaC, vs. MI lower than 5). The selected frequency threshold filters out word pairs with a frequency of 0 or 1 in the reference corpus, i.e. word pairs that are either not attested or that would not count as collocations within a frequency-based paradigm—it will be remembered that, conventionally, a minimum frequency of 2 is required for a word pair to count as a collocation, i.e. a recurrent sequence of words (Jones and Sinclair). The MI threshold distinguishing between strong and weak collocations is based on the median of the MI scores obtained for the EPTIC-derived Italian collocations (see Ferraresi *et al.* 2016).

Table 3. Examples of Italian TT phrases extracted from EPTIC and classified as either rare/unattested sequences or strong collocations

Category	POS pattern	Example
Rare/unattested	Adjective-Noun	Esteso ripristino (<i>widespread rehabilitation</i>)
Rare/unattested	Noun-Adjective	Forze manichee (<i>Manichean forces</i>)
Rare/unattested	Noun-Noun	Ritiro (delle) risposte (<i>withdrawal [of] responses</i>)
Strong collocation	Adjective-Noun	Alto livello (<i>high level</i>)
Strong collocation	Noun-Adjective	Governi locali (<i>local governments</i>)
Strong collocation	Noun-Noun	Piani [di] emergenza (<i>emergency plans</i>)

Analyses of these two categories of phrases are meant to shed light on two complementary aspects of phraseological conventionality of TTs. The rare/unattested category accounts for cases of *deviation* from conventionality, pointing to extremely uncommon and potentially non-standard usages in Italian; these correspond to what Biel calls “possible but rare” pairs, or pairs that are “non-existent in the target language” (30). MI is used instead to zoom in on highly conventional, idiomatic collocations, i.e. “fixed sequences” that “are significant because of the collocational attraction between the lexical items” (Gerard 164).

After classifying all noun phrases in EPTIC, the number of rare/unattested pairs and of strong collocations is calculated for each translated and interpreted TT. For each pair of TTs (corresponding to a given ST), a measure of the relative difference (henceforth RD) in terms of number of phrase types is then calculated, separately for the two categories of rare/unattested word pairs and strong collocations. Thus, *positive* RD values indicate cases of overuse of a phrase category in translated texts compared to their interpreted counterpart, while *negative* values indicate an overuse in interpreted texts. Appendix 1 displays RD values for all pairs of EPTIC texts, for the two categories of phrases considered.

RD values are often used as a quantitative measure to identify variation in MTC-based descriptive studies (Castagnoli, “Investigating competence”). In an applied perspective, they can also be used to speed up the process of selecting relevant, corpus-based examples for the creation of teaching materials, a process which can be “maddeningly time-consuming” (Marco and Van Lawick 19). To give an example, the EPTIC Italian TT sub-corpora contain 3,234 strong collocations matching the parameters described above—hardly a manageable number for manual concordance analysis.

In what follows, multi-parallel, intermodal concordance lines are extracted from the 2 EPTIC texts displaying the greatest positive and the

greatest negative RD values for each phrase category, for a total of 8 texts.⁵ These correspond to instances of translations displaying greater phraseological conventionality than interpretations (i.e. featuring a higher number of strong collocations or fewer rare/unattested word pairs), and, vice versa, instances of interpretations displaying greater phraseological conventionality than translations.

4.2. Analysing multi-parallel intermodal concordance lines

Differences between translated and interpreted TTs in the EPTIC sample analyzed can be subsumed under three main categories: a) differences in register, b) differences in meaning, and c) differences in “collocationality”, i.e. related to the production of lexically conventional vs. less conventional phrases. This bottom-up categorization, emerging from data analysis, has no pretense of being exhaustive, but is provided as a starting point to discuss translators’ and interpreters’ phraseological choices, and to speculate on their possible motivation.

Example 1) illustrates a case of register shift in translation resulting from the use of a strong collocation (“svolgimento di indagini”, *initiation of an investigation*), a nominalized form which raises the register of the TT compared to the ST. By contrast, not only does the interpreter adopt a more literal verbal rendering (“ci sia”, *there is*), but s/he also opts for a rarer, possibly calqued, equivalent for “investigations”, i.e. “investigazioni”.

- 1)
- a. Mi appello alla Commissione affin ché assicuri lo <svolgimento di indagini> adeguate ed esaurienti sulla questione. [TT-TR]
 - b. Mi rivolgo alla Commissione per assicurare che effettivamente ci sia un'indagine molto approfondita in merito. [TT-IN]
 - c. I call on the Commission to ensure that there is a proper and exhaustive investigation into this problem. [ST-TR+IN, ID: t79]⁶

Use of a collocation can also result in a lowering of text register. In example 2), the translated TT has the rather formal one-word equivalent “deplorable” (~ *deplorable*), used as an equivalent of “unfortunate”; the interpreted TT instead features the more informal, though highly idiomatic, collocation “vero peccato” (~ *real pity*).

⁵ Since RD results can be skewed by low frequencies, only texts with at least 5 instances of the phrase type of interest are considered.

⁶ ID refers to the EPTIC text ID (see Appendix 1). For space reasons, only one version of the ST is provided. It should be remarked, however, that the STs of the interpreted and translated texts discussed here display no substantial difference.

2)

- a. ... sarebbe stato un <vero peccato> se le autorità pakistane credessero veramente di poter risolvere ehm questo ca- chesta ehm caso della petroliera Tasman Spirit de tenendo semplicemente l'equipaggio.[TT-IN]
- b. Sarebbe deplorabile se le autorità pakistane credessero di poter risolvere il caso della Tasman Spirit tenendo in carcere l'equipaggio.[TT-TR]
- c. It would be most unfortunate if the Pakistani authorities were to believe that they can resolve the Tasman Spirit case by detaining the crew.[ST-IN/TR, ID: t50]

The second category of differences encompasses cases in which the translator or the interpreter operate a change of meaning in the TT, either by introducing information that was not explicitly provided in the ST, or omitting it. The former case is exemplified in 3), where both the translated and the interpreted TT feature a strong collocation. However, while the interpreter provides a close equivalent for the English “economic aspects” (“aspetti economici”), the translator reformulates it as “crescita economica” (*economic growth*), explicitating a meaning that was implicit in the vaguer ST expression.

3)

- a. ... è in condizione di adottare un approccio volto a favorire la sicurezza e la <crescita economica> e a promuovere la cooperazione regionale...[TT-TR]
- b. Aspetti economici e di sicurezza per promuovere la per promuovere la cooperazione.[TT-IN]
- c. This network is well placed to develop an approach that spans security and economic aspects and promotes regional cooperation...[ST-TR/IN, ID: t79]

The excerpts in 4) exemplify a case of omission. The ST features a figurative collocational expression, “gangster regime” (MI > 7 in ukWaC, the English counterpart of itWaC), which is rendered by the translator as “regime criminale”, an equally salient collocation in Italian; the collocation is not reproduced by the interpreter, who omits the second part of the ST sentence altogether. Lexicalized metaphors like this, and the immediately following “dustbins of history”, are known to pose serious challenges in simultaneous interpreting (Spinolo), and one can wonder whether the substantial omission in 4b) is at least partly motivated by difficulties in dealing with this metaphor-rich speech.⁷

⁷ A new, multimodal version of EPTIC is currently being set up that provides access to videos (with original and interpreted audio), from concordance lines. This feature should greatly facilitate analysis of interpreter performance.

4)

- a. E sorto in oltre la Russia a ritirare le sue truppe dalla Transnistria, a permettere una riunificazione pacifica della Moldavia e a consegnare il <regime criminale> di Tiraspol alle pattumiere della storia.
- b. Per quanto riguarda il Transnistria bisogna pre- permettere una riunificazione pacifica con la Moldavia.
- c. I also urge Russia to withdraw its troops from Transnistria, allow a peaceful reunification of Moldova and consign the Tiraspol gangster regime to the dustbins of history. [ST-TR/IN, ID: t76]

The last major type of differences observed in the data concerns instances in which the translator's and interpreter's choices of phrases differ mainly (but not necessarily only) along the axis of conventionality/attestedness. The examples that follow illustrate these differences with reference to a specialized term-like unit, and a "general language" phrase.

In example 5) the interpreted TT displays a more standard phrase than the translated one. Specifically, the ST features a multi-wordterm, "long-term immigrants", which has no readily available equivalent in Italian. To tackle the problem, the interpreter seems to pick the two most frequent equivalents for the noun and the modifier ("immigrati" and "a lungo termine"). For no obvious reason, the translator opts instead for "immigranti" and "di lungo periodo", whose frequencies are lower by at least an order of magnitude in itWaC than the options selected by the interpreter, making the resulting phrase less conventional, and arguably more obscure and technical-sounding.

5)

- a. Il Consiglio ha approvato una direttiva stentata sugli <immigranti di lungo periodo>... [TT-TR]
- b. Bisogna vedere gli immigrati a lungo termine... [TT-IN]
- c. The Council passed a grudging directive on long-term immigrants... [ST-TR+IN, ID: t19]

Example 6) also seems to defy attempts at getting at translators' and interpreters' decision-making processes for certain TT phraseological choices. Here, the interpreter produces the general language phrase "grave tristezza" for "great sadness", opting for a close equivalent of the ST noun ("tristezza"), and introducing a slight change of meaning through the adjective, which can be back-translated as *serious* (instead of "great"). One reason for this choice might be the intention to avoid the vague adjective "grande" (~ *great*), which however constitutes a much more frequent collocate of "tristezza". With an MI score of 5, "grande tristezza" can be

considered as a strong collocation (according to the parameters outlined in Section 4.1), while “grave tristezza” is attested only once in itWaC. The equivalent selected by the interpreter can therefore be qualified as an “untypical collocation” (Mauranen), i.e. one that is understandable but not conventional in the target language, and that is likely to result from “the fact that the right collocation was not prompted” (Biel 31). The most obvious equivalent collocation is not produced by the translator either, who, however, opts for “profondo rammarico” (MI > 10; ~ *deep regret*), a more salient (and higher register) target language collocation.

6)

- a. ... benché... Chiaramente co- siamo colpiti da una <grave tristezza> data la notizia della morte del presidente Trajkovski... [TT-IN]
- b. ... sebbene questo piacere sia chiaramente venato di profondo rammarico per la notizia della morte del Presidente Boris Trajkovski... [TT-TR]
- c. ... though this pleasure is clearly tinged with such great shared sadness at the news of the death of President Boris Trajkovski... [ST-TR+IN, ID: t81]

Phraseology is one of the crucial aspects contributing to what Biel calls the “textual fit”, or naturalness, of translated texts. Yet, assessing the extent to which phrases in translated (or interpreted) language may be considered as natural, or assessing motivation for any deviation from naturalness is far from trivial, especially when observing phraseological choices made by highly-skilled language professionals. Adopting a quantitative-oriented approach to the definition of collocations, and drawing on concordance lines from the Italian translated and interpreted sub-corpora of EPTIC, this Section has attempted to show how multi-parallel intermodal concordance lines can prompt reflections on similar issues. The next Section provides some suggestions as to how these reflections could be brought into the translation classroom.

4.3. EPTIC-based teaching materials on collocations: a proposal

Corpora have traditionally been used in translation teaching in two ways, i.e. a) as tools for autonomous learning on the part of trainees; or b) as a source of classroom materials, especially in task-based approaches, whereby a trainer selects samples for translation tasks and “controls their use with a view to achieving [...] pedagogic objectives” (Marco and van Lawick 10). The latter approach is followed here.

Appendixes A and B present two didactic activities on collocations which would fit, e.g., in a specialized English=>Italian translation course for students with at least basic knowledge of corpus-related techniques. Both activities are based on the examples discussed in Section 4.2. In designing

the activities, concordances for full sentences were extracted from EPTIC, and the interpreted TTs were slightly edited to remove false starts, filled pauses and hesitations, so as to let students focus on phraseological choices and not be distracted by transcription issues. No other intervention was carried out on the texts, which may thus include debatable choices not related to phraseology (but still worthy of classroom discussion), e.g. the use of verb tenses in “Sarebbe stato un vero peccato se [...] credessero” (instead of “avessero creduto”).

The first activity (see Appendix B) is one where students are asked to analyze and discuss intermodal concordance lines. These were selected among the instances of register and/or collocational shifts that do not seem to be motivated by ST triggers. Classroom discussion could be approached from different perspectives, including appropriateness of single phraseological choices with reference to the ST and target language conventions, and/or issues related to norms, strategies and constraints of translation and interpreting tasks in general. From a linguistic perspective, intuitions on the level of conventionality of phraseological items can first be (dis)confirmed by looking up their frequency/MI score in a general language corpus like itWaC (see footnote 4). Students could then be prompted to notice how translators seemingly use collocations to make the TT more formal or conventional than the ST (in cases *a* and *c*), and that a similar tendency is observed (in case *b*) in the choice of the word “deplorable”. These shifts should stand out all the more clearly when compared with the choices made by interpreters, who either opt for more literal renderings (in cases *a* and *c*), or even lower the degree of formality (in case *b*). Concerning higher-level, cognitive aspects underlying these choices, the corpus evidence provided could also be used to stimulate a broader debate on the norms governing translators’ and interpreters’ decisions, and especially the extent to which they may result from conscious strategies or from task-related automatisms.

The second activity (see Appendix C) focuses on ST phraseological items which were found to be problematic for interpreters and/or translators. In this activity, students are required to first sight-translate a short sentence, and then produce a written revised version based on their sight-translation. This should give them first-hand experience of the constraints involved in the fast, automatic process of interpreting, as opposed to the slower and more reflexive process of translation. After completion of the task, students could compare their decisions to those of the highly-skilled professionals featured in EPTIC, evaluating the relative merits and limitations of the different renderings. They could also be invited to consult itWaC in the search for alternative collocates of the words in focus, along the lines of what was done for “tristezza” in Section 4.2. This makes it possible to

enlarge the pool of potential phraseological equivalents for challenging ST phrases, e.g. “regime violento/fuorilegge” (~ *violent/outlaw regime*) for “gangster regime”, or “immigrati di lunga durata/data” (literally *long-date/long-duration immigrants*) for “long-term immigrants”, which in turn should emphasize the positive reinforcement role that different types of corpora can play in the translation process (Aston, “Corpus use”).

5. Conclusion

Intermodal corpora, i.e. corpora bringing together parallel or comparable translated and interpreted texts, are still relatively new within corpus-based descriptive approaches to translation (and interpreting) studies. Yet, in the words of Miriam Schlesinger, who first proposed this corpus setup, they open up fascinating new perspectives in the discipline(s), insofar as

translation scholars can learn about the process and product of (written) translation by finding out more about interpreting – and interpreting scholars can infer about this high-pressure form of translation by observing the slower, more readily observable process and product of (written) translation. (*Schlesinger and Ordan 44*)

Drawing on the recently created EPTIC corpus, and exploiting its multi-parallel nature, which makes it akin to better-established multiple translation corpora, this contribution has started to explore how intermodal observations could be integrated in task-based translation teaching materials focusing on collocations.

Despite its inherent limitations, including its small size, and the fact that it is a single-genre corpus of parliamentary proceedings (arguably far less attractive to students than, say, MTCs of literary texts), it is suggested that the potential and flexibility of EPTIC as a didactic tool are substantial. Applying the method described here, further studies could look at intermodal variation in terms of other areas of difficulty for translation and interpreting students, including connective explicitation (Castagnoli, “Investigating competence”), or the expression of verb subjects in translation between null- and non-null-subject languages (Mikolič).

If, as claimed by Pym (“On omission” 94), “current practices on the written side of business [*i.e. translation*] suggest that time constraints and multi-tasking are problems for all”, and not just for interpreters, the training of future language professionals could greatly benefit from similar explorations at the interface between modes of language mediation.

Works cited

- Aston, Guy. "Corpus Use and Learning to Translate". *Textus* 12 (1999): 289-314. Print.
- . "Foreword". *Corpus Use and Translating*. Eds. Allison Beeby, Patricia Inés Rodríguez, and Pilar Sánchez-Gijón. Amsterdam: Benjamins, 2009. IX–X. Print.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi, and Eros Zanchetta. "The Wacky Wide Web". *Language Resources and Evaluation* 43.3 (2009): 209–226. Print.
- Bernardini, Silvia. "Intermodal Corpora. A Novel Resource for Descriptive and Applied Translation Studies". *Corpus-based Approaches to Translation and Interpreting*. Eds. Gloria Corpas Pastor and Miriam Seghiri Domínguez. Bern: Peter Lang, 2016. 129–148. Print.
- Bernardini, Silvia, Adriano Ferraresi and Maja Miličević. "From EPIC to EPTIC—Exploring Simplification in Interpreting and Translation from an Intermodal Perspective". *Target* 28.1 (2016): 61–86. Print.
- Biel, Łucja. "The Textual Fit of Legal Translations". *Teaching Translation and Interpreting*. Ed. Łukasz Bogucki. Newcastle upon Tyne: Cambridge Scholars Publishing, 2010. 25–39. Print.
- Bowker, Lynne and Peter Bennison. "Student Translation Archive and Student Translation Tracking System". *Corpora in Translator Education*. Eds. Federico Zanettin, Silvia Bernardini, and Dominic Stewart. Manchester: St. Jerome. 2003. 103–118. Print.
- Castagnoli, Sara. "Exploring Variation and Regularities in Translation with Multiple Translation Corpora". *Rassegna Italiana di Linguistica Applicata* 43.1 (2011): 311–332. Print.
- Castagnoli, Sara. "Investigating Trainee Translators' Contrastive Pragmalinguistic Competence". *The Interpreter and Translator Trainer* 10.3 (2016): 343–363. Print.
- Colson, Jean-Pierre. "Cross-linguistic Phraseological Studies". *Phraseology: An Interdisciplinary Perspective*. Eds. Sylviane Granger and Fanny Meunier. Amsterdam: Benjamins, 2008. 191–206. Print.
- Evert, Stefan. "Corpora and Collocations". *Corpus Linguistics, vol. 2*. Eds. Anke Lüdeling and Merja Kytö. Berlin: Mouton de Gruyter, 2008. 1212–1248. Print.
- Ferraresi, Adriano, Bernardini, Silvia and Maja Miličević. "Collocations across Languages". *Proceedings of Corpus Linguistics 2015*. Eds. Federica Formato and Andrew Hardie. N.p., 2015. 106–109. Web: <http://ucrel.lancs.ac.uk/cl2015/doc/CL2015-AbstractBook.pdf>. 27 November 2016.
- Gerard, Jessica E. *The reading of Formulaic Sequences in a Native and Non-native Language: An Eye Movement Analysis*. PhD thesis, University of Arizona, 2007. Print.

- Gile, Daniel. "Translation Research Versus Interpreting Research". *Translation Research and Interpreting Research*. Ed. Christina Schäffner. Clevedon: Multilingual Matters, 2004. 10–34. Print.
- Ji, Meng and Michael P. Oakes. "A Corpus Study of Early English Translations of Cao Xueqin's Hongloumeng". *Quantitative Methods in Corpus-based Translation Studies*. Eds. Michael P. Oakes and Meng Ji. Amsterdam: Benjamins, 2012. 177–208. Print.
- Johansson, Stig. "Why Change the Subject? On Changes in Subject Selection in Translation from English into Norwegian". *Target* 16.1 (2004): 29–52. Print.
- Jones, Susan and John McHardy Sinclair. "English Lexical Collocations". *Sinclair on Lexis and Lexicography*. Ed. Joseph A. Foley. Singapore: UniPress, 1974/1996. 21–54. Print.
- Kajzer-Wietrzny, Marta. *Interpreting Universals and Interpreting Style*. PhD thesis, Adam Mickiewicz University, 2012. Print.
- Kruger, Haidee. "Language Change, Photoshopped Language and Constrained Communication". *EST Newsletter* 45 (2014): 8–10. Print.
- Laviosa, Sara. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose". *Meta* 43.4 (1998): 557–570. Print.
- Malmkjaer, Kirsten. "On a Pseudosubversive Use of Corpora in Translator Training". *Corpora in Translator Education*. Eds. Federico Zanettin, Silvia Bernardini, and Dominic Stewart. Manchester: St. Jerome. 2003. 119–134. Print.
- . "Love Thy Neighbour: Will Parallel Corpora Endear Linguists to Translators?". *Meta* 43.4 (1998): 534–541. Print.
- Marco, Josep and Heike van Lawick. "Using Corpora and Retrieval Software as a Source of Materials for the Translation Classroom". Eds. Allison Beeby, Patricia Inés Rodríguez, and Pilar Sánchez-Gijón. Amsterdam: Benjamins, 2009. 9–28. Print.
- Mauranen, Anna. "Translation Universals". *Encyclopedia of language & linguistics, 2nd ed.* Ed. Keith Brown. Amsterdam: Elsevier, 2006. 93–100. Print.
- McEnery, Tony, Richard Xiao and Yukio Tono. *Corpus-based Language Studies*. London and New York: Routledge, 2006. Print.
- Munday, Jeremy. *Evaluation in Translation*. London and New York: Routledge, 2012.
- Pym, Anthony. "Redefining Translation Competence in an Electronic Age". *Meta*, 48.4 (2003): 481–497. Print.
- . "On omission in Simultaneous Interpreting". *Efforts and Models in Interpreting and Translation Research*. Eds. Gyde Hansen, Andrew Chesterman, and Heidrun Gerzymisch-Arbogast. Amsterdam: Benjamins, 2008. 83–105. Print.
- Russo, Mariachiara, Claudio Bendazzoli, Annalisa Sandrelli, and Nicoletta Spinolo. "The European Parliament Interpreting Corpus (EPIC)". *Breaking Ground in Corpus-based Interpreting Studies*. Ed. Francesco Straniero Sergio and Caterina Falbo. Bern: Peter Lang, 2012. 35–90. Print.
- Rychlý, Pavel. "Manatee/Bonito—A Modular Corpus Manager". *Proceedings of the 1st Workshop on Recent Advances in Slavonic NLP*. N.p., 2007. 65–70. Print.

- Shlesinger, Miriam. “Corpus-based Interpreting Studies as an Offshoot of Corpus-based Translation Studies”. *Meta* 43.4 (1998): 486–493. Print.
- . “Towards a Definition of Interpretese”. *Efforts and Models in Interpreting and Translation Research*. Eds. Gyde Hansen, Andrew Chesterman, and Heidrun Gerzymisch-Arbogast. Amsterdam: Benjamins, 2008. 237–253. Print.
- Shlesinger, Myriam and Noam Ordan. “More Spoken or More Translated? Exploring a Known Unknown of Simultaneous Interpreting”. *Target* 24.1 (2012): 43–60. Print.
- Spinolo, Nicoletta. “Afrontar el Lenguaje Figurado en Interpretación Simultánea”. *Intralinea* 15. Web: <http://www.intralinea.org/specials/article/2011>. 26 November 2016.
- Stewart, Dominic. “Safeguarding the Lexicogrammatical Environment”. *Corpus Use and Translating*. Eds. Allison Beeby, Patricia Inés Rodríguez, and Pilar Sánchez-Gijón. Amsterdam: Benjamins, 2009. 29–46. Print.
- Tiedemann, Jörg. “Parallel Data, Tools and Interfaces in OPUS”. *Proceedings of the 8th International Conference on Language Resources and evaluation*. N.p., 2012. Print.
- Zanettin, Federico. *Translation-driven Corpora. Corpus Resources for Descriptive and Applied Translation Studies*. Abingdon: Taylor & Francis, 2012. Print.

Appendix 1

Number of phrase types and RD values in EPTIC Italian target texts, split by category of phrase

Text ID	Strong collocations (MI \geq 5)			Rare/unattested phrases (FQ < 2)		
	N. of types in TR-TT	N. of types in IN-TT	RD value	N. of types in TR-TT	N. of types in IN-TT	RD value
t01	58	44	31.82	10	4	150.00
t02	21	12	75.00	7	3	133.33
t03	15	12	25.00	4	3	33.33
t04	12	9	33.33	6	6	0.00
t05	13	13	0.00	2	1	100.00
t06	21	10	110.00	4	2	100.00
t07	82	45	82.22	18	25	-28.00
t08	8	4	100.00	0	0	NA
t09	5	4	25.00	3	2	50.00
t10	3	3	0.00	0	0	NA
t11	21	15	40.00	2	3	-33.33
t12	19	18	5.56	3	4	-25.00
t13	23	8	187.50	1	1	0.00
t14	35	22	59.09	6	7	-14.29
t15	25	20	25.00	3	2	50.00
t16	24	23	4.35	0	0	NA
t17	6	2	200.00	0	1	-100.00
t18	118	87	35.63	4	22	-81.82
t19	37	18	105.56	9	5	80.00
t20	44	23	91.30	4	3	33.33
t21	18	14	28.57	1	4	-75.00
t22	22	13	69.23	2	4	-50.00
t23	61	37	64.86	8	11	-27.27
t24	4	3	33.33	1	2	-50.00
t25	53	50	6.00	3	18	-83.33
t26	22	10	120.00	2	2	0.00
t27	56	40	40.00	7	11	-36.36
t28	62	43	44.19	15	13	15.38

Text ID	Strong collocations (MI \geq 5)			Rare/unattested phrases (FQ < 2)		
	N. of types in TR-TT	N. of types in IN-TT	RD value	N. of types in TR-TT	N. of types in IN-TT	RD value
t29	22	7	214.29	3	2	5000
t30	13	11	18.18	1	0	NA
t31	62	34	82.35	8	19	-57.89
t32	19	10	90.00	1	0	NA
t33	16	11	45.45	2	5	-60.00
t34	23	11	109.09	3	3	0.00
t35	24	7	242.86	2	2	00
t36	33	16	106.25	6	1	500.00
t37	16	12	33.33	5	2	150.00
t38	59	43	37.21	9	11	-18.18
t39	22	14	57.14	3	4	-25.00
t40	25	20	25.00	9	5	80.00
t41	27	15	80.00	1	1	0.00
t42	24	13	84.62	13	13	0.00
t43	27	21	28.57	11	8	37.50
t44	10	12	-16.67	1	2	-50.00
t45	17	8	112.50	4	5	-20.00
t46	11	8	37.50	4	1	300.00
t47	65	40	62.50	12	10	20.00
t48	23	10	130.00	8	1	700.00
t49	12	12	0.00	2	3	-33.33
t50	9	15	-40.00	4	6	-33.33
t51	4	6	-33.33	2	1	100.00
t52	7	4	75.00	0	3	-100.00
t53	2	3	-33.33	0	0	NA
t54	13	13	0.00	4	4	0.00
t56	5	3	66.67	0	2	NA
t57	0	0	NA	0	0	NA
t58	3	1	200.00	0	1	NA
t59	3	2	50.00	0	0	NA
t60	0	2	NA	0	1	NA

Text ID	Strong collocations (MI \geq 5)			Rare/unattested phrases (FQ < 2)		
	N. of types in TR-TT	N. of types in IN-TT	RD value	N. of types in TR-TT	N. of types in IN-TT	RD value
t61	1	0	NA	0	0	NA
t62	10	0	NA	0	0	NA
t63	0	0	NA	1	0	NA
t64	1	6	-83.33	0	0	NA
t65	90	70	28.57	6	8	-25.00
t66	41	19	115.79	8	7	14.29
t68	42	36	16.67	2	6	-66.67
t69	12	7	71.43	1	2	-50.00
t70	9	9	0.00	2	1	100.00
t71	15	10	50.00	2	1	100.00
t72	21	9	133.33	0	1	NA
t73	15	6	150.00	4	2	100.00
t74	24	20	20.00	5	4	25.00
t75	52	38	36.84	5	4	25.00
t76	29	6	383.33	7	2	250.00
t77	37	25	48.00	2	4	-50.00
t78	13	2	550.00	3	0	NA
t79	29	5	480.00	4	4	0.00
t80	10	6	66.67	5	4	25.00
t81	20	19	5.26	5	13	61.54

Appendix 2

Activity 1. Concordance analysis

Read the Italian translated and interpreted versions of the following excerpts from EU Parliamentary speeches delivered in English, and comment on the choices that were made. Are there any choices that you consider inappropriate or infelicitous? How would you explain to the translator/interpreter why you disagree with their choices? Can you provide alternative renderings that you believe would improve the Italian texts?

a)

ST: I call on the Commission to ensure that there is a proper and exhaustive investigation into this problem.

Translated TT: Mi appello alla Commissione affin ch  assicurati lo svolgimento di indagini adeguate ed esaurienti sulla questione.

Interpreted TT: Mi rivolgo alla Commissione per assicurare che effettivamente ci sia un'investigazione molto approfondita in merito.

b)

ST: It would be most unfortunate if the Pakistani authorities were to believe that they can resolve the Tasman Spirit case by detaining the crew.

Translated TT: Sarebbe deplorabile se le autorit  pakistane credessero di poter risolvere il caso della Tasman Spirit tenendo in carcere l'equipaggio.

Interpreted TT: Sarebbe stato un vero peccato se le autorit  pakistane credessero veramente di poter risolvere questo caso della petroliera Tasman Spirit detenendo semplicemente l'equipaggio.

c)

ST: President Mesic, it is my great pleasure to welcome you here today, though this pleasure is clearly tinged with such great shared sadness at the news of the death of President Boris Trajkovski.

Translated TT: Presidente Mesic,   con grande piacere che le porgo il benvenuto qui oggi, sebbene questo piacere sia chiaramente venato di profondo rammarico per la notizia della morte del Presidente Boris Trajkovski.

Interpreted TT: Signor Presidente   un grande piacere per me darle il benvenuto qui quest'oggi bench  chiaramente siamo colpiti da una grave tristezza data la notizia della morte del presidente Trajkovski.

Appendix 3

Activity 2. Translation of short passages.

Read the following excerpts from speeches delivered in English at the EU Parliament. First sight-translate them, then produce a written, revised version, using general language corpora of English and Italian (e.g.ukWaC and itWaC) for reference.

1)

I urge Russia to withdraw its troops from Transnistria, allow a peaceful reunification of Moldova and consign the Tiraspol gangster regime to the dustbins of history.

2)

The Council passed a grudging directive on long-term immigrants and one on family reunification.