Corpora for Language Learning and the Training of Translators

Federico Zanettin
University of Bologna-Forlì

0. Introduction

In this paper I discuss applications of corpus linguistics in a learning environment, focusing on the use of a specific kind or corpus, the comparable bilingual corpus (i.e. a collection of texts in two languages matched on the basis of similarity of content and function), in the training of translators. I analyse features which can be investigated using small bilingual corpora (English-Italian), showing how they can be used for translation activities aimed at developing learning skills, both as regards comprehension and production.

1. Corpora and translation

One of the major developments in linguistic research has come from the possibility of studying vast amounts of text through computer-enhanced tools¹, particularly through text retrieval and concordancing programs. Most recently corpus linguistics techniques have been put to use in the field of translation, leading to the construction of a wide array of different types of corpora for descriptive as well as for applied purposes. The criteria with which these corpora are designed reflect the purpose for which they are created.

A first type is monolingual "comparable" corpora, consisting of two sets of texts, one originally written in language A and one of similar texts translated into language A from a variety of different languages (Baker 1995, 1966, Laviosa-Braithwaite 1996). Their value is mainly theoretical: what is investigated is the nature of translated texts and translation itself. The hypothesis is that all translated texts share certain features, independently of their source. A comparison of the features which all or most translated texts share with those of comparable texts originally produced in that language may shed light on the strategies which are employed in the process of translating.

A second kind of corpus used in translation is the parallel corpus. Parallel corpora, consist of texts in language A and their translation into language B. The

¹ For an overview of a corpus-based approach to language and linguistics see for example McEnery & Wilson 1996.
relationship between texts is directional, i.e. it goes from one text (the SL text) to the other (the TL text).

Parallel corpora have been extensively used in terminology extraction (e.g. Dryberg & Tournay 1990, Laffling 1992) and machine aided translation (see Somers 1993). However, much recent research in MT aims not so much at the creation of a system able to perform automatically the job of translating a given text, but at the implementation of computerised tools to assist human translators in their tasks. Parallel corpora can also be treated as "translation memories", from which translators can retrieve chunks of translated language in order to speed up their work and ensure accurate and consistent translations. While it has been observed that translated texts are prone to being artificial and reflecting stylistic idiosyncrasies of the source language and of individual translators, the comparison between a set of texts and their acknowledged translations can show how equivalence has been established by translators under certain circumstances and provide examples of translations strategies. And if such corpora are sufficiently varied and large,

... the accumulation of motivated translation choices revealed by a concordance program allows the general patterns to be perceived. Hence generalisations emerge from the aggregation of large number of individual instances. (Barlow 1996)

The third type of corpus which may be relevant to the field of translation, and the one which I shall be dealing with here, is what I shall term bilingual comparable corpora2. These are collections of texts composed independently in the respective languages and put together on the basis of similarity of content, domain and communicative function. They are

typically unrelated except by the analyst’s recognition that the original circumstances that led to the creation of the two [sets of] texts have produced accidental similarities. (Hartmann 1980: 38)

The practice of collecting texts in different languages on the basis of similarity of content, domain and function was a common practice in translation research

---

2 According to Hartmann (1980), a third kind of bilingual texts, labelled "adaptations", is exemplified by the production of advertising copy for different countries or by authoritative multilingual versions of international law. It is characterised by simultaneous formulation with reference to a common source. While equivalence is not here based on a "textual" identity (from "original" texts to "secondary" texts) but on conceptual identity, it can be argued that all translations are adaptations to the "respective conventions of the two languages for the purpose of conveying an identical message to receivers of sometimes very different cultures" (Hartmann 1980: 38).
and training before the word corpus came to mean almost exclusively a collection of electronic texts. Snell-Hornby (1984, 1986) examines a corpus of printed English and German public notices, and Shäffner (1996) discusses how bilingual English German corpora composed of, for example, tourist brochures or instruction manuals can be used for the training of translators. They argue that the analysis of a large number of texts belonging to the same text-type can identify prototypical features, and thus provide a "profile" of the type of text which translators aim at producing in the target language.

Technological advances have now made it possible to construct easily small bilingual comparable corpora. Teachers, students and researchers as well as professional translators can make use of corpora and text analysis software as a complementary resource to printed books and other materials. Comparable corpora can be created from a variety of sources: collections of texts distributed in electronic format (e.g. newspaper archives on CD-ROM, the Internet, etc.), or even from scanned or typewritten material. Criteria for creating comparable corpora depend on the homogeneity of texts, both within and across languages, in terms of features such as subject domain, author-reader relationship, text origin and constitution (i.e. "single" or "joint"/"composite" texts), factuality, technicality and intended outcome (i.e. communicative function) (see Pearson 1996).

The size of comparable corpora can of course vary, depending on how far they meet these criteria. In the case of texts with a high degree of technicality, written by subject specialists for their peers, it is highly likely that relevant information can be derived even from a small number of texts.

In the following section I will describe analyses carried in translation courses at the Scuola Superiore di Lingue Moderne per Interpreti e Traduttori of the University of Bologna (at Forlì) using collections of newspaper articles downloaded from CD-ROMs in Italian and in English. The software used to analyse these bilingual corpora was WordSmith Tools (Scott 1996).

### 2. Corpora in the language classroom

In the classroom, comparable corpora can be used to confirm translation hypotheses and to suggest possible solutions to actual translation problems related to a specific text. This process is text-based: the corpus can offer

---

3 See Zanettin 1994 for details on the design of such corpora.
4 Other text analysis and concordancing tools include freely available programs such as Tact and Corpus Wizard, and commercially available software such as MicroConcord, Longman MiniConcordancer, DBT, etc. For further information see Corpus Linguistics Page at http://www.sslmit.unibo.it/zanettin/cl.tm
information about terminology and concepts, and about the attestedness of expressions within a certain context. Comparable corpora can also provide a means of investigating similar genres across languages.

The analysis of texts produced in comparable communicative situations can help learners investigate the respective expectations, experience and knowledge of the linguistic communities involved. In the process of establishing equivalence between comparable sets of texts, learners acquire information about the way in which discourse is laid down in the two languages. They can use the attested evidence which corpora provide and create new texts which are partly made of citations from the target language adapted to the new occasion. By looking for coincidence across language and manipulating units which approximate the concepts and functions they want to convey, they engage in a meaning creation activity and develop procedural skills.

The basic investigation procedure for querying text corpora consists in producing multiple concordance lines, a so-called Key Word In Context – or KWIC – concordance, for a specified string of characters – a word, a lemma or a phrase. The citations thus obtained can be sorted to reveal recurring clusters of words. Analysis of these recurring patterns highlights the behaviour of actual language in context, and complements and sometimes challenges the information from standard reference tools such as dictionaries and grammars.

For example, sports articles abound in figurative language. It may be difficult to establish whether an expression such as "salire il gradino più alto del podio" (to climb onto the highest step of the podium") would sound "native-like" if translated literally (figure 2):

Figure 2

L'ammiraglio Straulino, con a prua Niccolò Rode, negli anni 50 vince tre campionati mondiali, dieci europei, sale il gradino più alto del podio alle Olimpiadi finlandesi del 1952 e quattro anni dopo conquista la medaglia d'argento ai giochi olimpici australiani.


Looking at a 65,000 word corpus of articles on the 1992 Olympic games taken from Il Sole 24 Ore, we find that this expression is recurrently used in Italian to mean "win the gold medal": a search for the word podio within a context of 5 words of gradino produced the following concordance lines:
Corpora for Language Learning and the Training of Translators

Figure 3
Search word: PODIO + GRADINO (5 words to left or right)
1. dieci europei, sale il gradino più alto del podio alle Olimpiadi finlandesi del
2. appena conquistato il gradino più alto del podio nella 4x100 (Ashford, Jones,
3. Oggi favoriti per il gradino più alto del podio Scarpa e Josefa Idem; outside
4. la gara era mista) sul gradino più alto del podio, la cinesina Zhang Shang. La
5. costretto a cedere il gradino più alto del podio alla Germania. Appassionante i
6. ecutiva ai Giochi, sul gradino più alto del podio ben cinque volte nell'edizione

(corpus: 65,000 words from Il Sole 24 Ore of 1992 about the Barcelona Olympic Games)

Looking at a 250,000 word corpus of English articles dealing with the same topic, a search for the words "podium*" produced 22 citations (figure 4).

Figure 4
Search word: PODIUM*
Sort: 1st word to left then 2nd word to left
1. ay Michael Carruth was standing on a podium in the boxing arena here, listening t
2. nywhere better than an Olympic Games podium' By COLIN GIBSON KEVIN YOUNG last
3. ever stepped on to the Olympic medal podium he would have been identified as one
4. sure he said farewell from the medal podium. What the former Army sergeant could
5. ionships I have stood on three medal podiums. Next year I will be running around
6. I am going to do when I get on that podium,' he says engagingly, 'and I have nev
7. tation the three men embraced on the podium for the last time as team-mates befo
8. good. I'm proud to have been on the podium, though, and it's a great way to fini
9. expected to take their place on the podium at Barcelona. Edgington conceded that
10. e acclaim; Skah shuffled over on the podium to take, rather than shake, his hand
11. me away from a definite place on the podium after crushing Australia 98-65 in the
12. in the world. Only as he stood on the podium, the first Briton to earn an Olympic
13. in. But I was proud to stand on the podium after a race like that. It's a great
14. at Britain team stood proudly on the podium in Seoul, with the Union Flag flutter
15. t finish anywhere better than on the podium of an Olympic Games.' Ever since his
16. s carefully mapped-out route to the podium. Childerley Page -2- Printed Tue Jul
17. dal and was twice called back to the podium by the enthusiastic crowd. The overall
18. has even opened up the route to the podium in demanding prestige events such as
19. .60m. Despite his climbing on to the podium along with Zelezny and Raty, there wa
20. old real chances of making it to the podium ... Paul Brotherton, 25, and Andy
21. as Skah stepped jauntily out to the podium matched that which accompanied his
22. two Americans stood on the winner's podium to salute the anthem. True that duo,

(corpus: 250,000 words from The Independent and The Daily Telegraph of 1992 about the Barcelona Olympic Games)
They confirm that indeed podium refers to the same concept as podio and is used figuratively, but that it does not occur in conjunction with "the highest step" to mean "win the gold medal". In this case the evidence provided by the corpus may help avoid a literal translation which, even if possible in principle, may have the effect of sounding awkward and non-native like.

Some recurring collocational patterns, however, can be seen in the English citations. The phrase "stand on the podium" appears in 6 lines out of 22. In lines 3, 4, and 5 we find the phrase "medal podium" and in line 22 "two Americans stood on the winner's podium". A translator may thus decide to use this information and write, for example, that "Admiral Straulino ... stood on the winner's podium..." or adhere more closely to the source text and simply write that he won the gold medal, depending on the target communicative context or on target textual constraints.

Corpora made up of specialised texts, such as the "hepatitis corpus" described by Laura Gavioli in this volume, can be a useful source of terminology and content information. One of the main feature of medical texts is the high percentage of specialised terminology, often of Latin or Greek origin, resulting in many "look-alike" terms in English and Italian. Terms in the target language can thus be easily identified and their collocates compared to those of the corresponding terms in the source language to reveal diverging patterns. In the following sentence:

Due mesi dopo il trapianto fu eseguita una biopsia epatica che evidenziò alterazioni ascrivibili a rigetto.

We find the term biopsia epatica which seems apparently unproblematic, since biopsy and hepatic are both English words belonging to medical language, as any dictionary can reveal.

A search for biopsy, however, shows that the phrase hepatic biopsy does not occur, while liver biopsy would appear to be a more satisfactory translation equivalent with its 39 occurrences. A comparison between the words liver and fegato, and between hepatic and epatico shows that in English the adjective hepatic collocates only with generic words such as disease, lesion or failure. In Italian, on the other hand, the adjective epatico is preferred to compounds with the word fegato, the only exception being trapianto di fegato.

The second approach to comparable corpora involves using them to investigate a particular genre and/or topic area, as in pre- or post-translation activities. Comparable corpora can be the source of a potentially endless serendipity process, as one word or phrase leads to another, depending on learners' intuition and on their individual proficiency, interests or needs. Comparable corpora provide learners with a means of testing the relationship between items of language which they perceive as holding some kind of
similarity or equivalence. Learners can be instructed to look for similarities between languages and to compare words and phrases by identifying categories which have a strong formal resemblance, such as proper names and cognates (see Partington 1995), or which are proposed as translation equivalents in dictionaries.

For example, a search for the proper name *Mitterrand* in a comparable corpus of foreign news articles about France (from *The Independent*, *The Daily Telegraph*, and *Il Sole 24 Ore* of 1992, of about 152,000 in English and 102,000 words in Italian) shows some differences in the way Mitterrand is talked about in the two languages. The search produced 152 and 188 concordance lines in Italian and in English respectively (0.13% in both corpus components). Figure 5 lists all two-words clusters occurring in these concordance lines with a frequency higher than ten.

Figure 5
2-word clusters from MITERRAND

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Italian</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 ... francos mitterrand ...</td>
<td>41</td>
<td>1 ... mr mitterrand ...</td>
</tr>
<tr>
<td>2 ... di mitterrand ...</td>
<td>21</td>
<td>2 ... francos mitterrand ...</td>
</tr>
<tr>
<td>3 ... il presidente ...</td>
<td>21</td>
<td>3 ... president mitterrand ...</td>
</tr>
<tr>
<td>4 ... mitterrand e ...</td>
<td>16</td>
<td>4 ... president francos ...</td>
</tr>
<tr>
<td>5 ... presidente francos ...</td>
<td>14</td>
<td>5 ... m mitterrand ...</td>
</tr>
<tr>
<td>6 ... a mitterrand ...</td>
<td>11</td>
<td>6 ... mitterrand has ...</td>
</tr>
<tr>
<td>7 ... mitterrand che ...</td>
<td>11</td>
<td>7 ... mitterrand said ...</td>
</tr>
<tr>
<td>8 ... mitterrand ha ...</td>
<td>11</td>
<td>8 ... mr mitterrand's ...</td>
</tr>
</tbody>
</table>

A first observation is that the Italian texts seem to be much more at ease in calling politicians by their full or last name, e.g. *Francois Mitterrand* or simply *Mitterrand*, whereas the English texts preferably call him *President Mitterrand* or *President Francois Mitterrand* and *Mr* or *M Mitterrand*. This may seem somehow counter-intuitive, as it is known that Italians are generous in using titles, we all are *dottori, professori* or *geometri*.

This observation can be followed up by a search for all contexts where the words *presidente* and *francese* and *French* and *President* respectively co-occur.

Figure 6
Search word: PRESIDENTE + FRANCESE (5 words to left or right)
Sort: 1st word to right then 2nd word to right

1. ezzata anche Francois Pèrigot, presidente della Cnpf, la Confindustria francese
2. francese su tre, Jacques Delors, presidente della Commissione di Bruxelles ha "ri
to _ afferma Micha Spierenburg, presidente della filiale francese della banca d'
3. esentato al popolo francese dal presidente della Repubblica autorizzando la rati
4. rancese, Pierre Bèrègovoy. Il presidente della Repubblica, Francois Mitterrand
6. a Parigi per esaminare con il presidente francese Francois Mitterrand i prossi
7. PARIGI_ La mina vagante che il presidente francese Francois Mitterrand ha lanci
8. CIDENTALE Didascalia FOTO-01 Il presidente francese Francois Mitterrand arriva a
9. monopolio dell'iniziativa, e il presidente francese ha detto che altri Paesi, It
10. leanza. All'inizio del 1991 il presidente francese Mitterrand e il cancelliere t
11. litica nella zona del Golfo. Il presidente francese Mitterrand, secondo due lead
12. contro Francois Mitterrand. Al presidente francese, che la sera prima aveva cri
13. edì' a Parigi per incontrare il presidente francese, Francois Mitterrand. I coll
14. ormai consumato. Ieri l'anziano presidente, in un'intervista al quotidiano franc

Figure 7
Search word: FRENCH + PRESIDENT (5 words to left or right)
Sort: 1st word to right then none
1. unacceptable consequences for French agriculture, he said. President Mitterra
2. ortend a convulsion within the French body politic. President Francois Mitterta
3. sit the Louvre, centrepiece of French culture and also of President Mitterrand'
4. e same,' Bernard Lapasset, the French federation president, said. Of those queu
5. lation,' said Simone Veil, a French former president of the European parliame
6. . It will be the full fig. The French, from the President downwards, have obvio
7. rove the Bill submitted to the French people by the President of the Republic a
8. ld radically redraw the map of French politics. The Socialist President's swift
9. only a brief meeting with the French President and was given finger-wagging le
10. ows to defend sterling in ERM, French President champions European union: Andre
11. the German Chancellor and the French President issued no formal statement afte
12. r Gorbachev wisely ignored the French president's pleas. François failure to ma
13. espondent THE president of the French Red Cross and three government advisers r
14. ard Lapasset, president of the French Rugby Federation, said yesterday he hoped
15. nophobia. The president of the French rugby federation last week pointed to a p
16. was elected president of the French Senate yesterday after a battle between o
17. ois Theodore, president of the French stock market ruling body, the Societe des
18. i de Benoist, president of the French wheat-producers' association, told Agence

Corpus: foreign news articles about France (from The Independent, The Daily Telegraph, and Il Sole 24 Ore of 1992, about 152,000 in English and 102,000 words in Italian)

In the Italian corpus we find that the phrases Il presidente francese Mitterrand or Il presidente francese Francois Mitterrand are quite common, while the English texts rarely use the adjective French and prefer to avoid long phrases such as the French President Francois Mitterrand or the French President Mitterrand. A further activity may be to enlarge the context of the citations and investigate the structural position of these references. What emerges is that English articles use the name Mitterrand without a title only in headlines. The body of the article
usually introduces him as *President Mitterrand* and later on refers to him as *Mr Mitterrand* or *the President*. In the Italian articles he is as often first mentioned as *Francois Mitterrand* as *il Presidente Mitterrand*, and subsequent references are to *il presidente / il presidente francese*, or simply to *Mitterrand*. There are of course exceptions, but general patterns of reference can be observed and compared. These can then be used as stylistic models to relate to when writing in the foreign language.

This corpus can also be used to highlight differences in the lexis used to introduce direct speech. Most of the time President Mitterrand is quoted as saying something. In English, quotations seem to be mostly introduced by the verb "say". A concordance of all occurrences of pronouns, *Mr, Mrs, Ms, and Pr* followed or preceded by a quotation mark (within a span of 7 words on each side) allows the learner to compare how direct speech is reported in the Italian and the British articles.

A search for *ha* in the vicinity of a quotation mark in the Italian texts identifies a list of functionally equivalent verbs to "dire" (in the third person of passato prossimo). By scanning the many concordance lines thus generated (a few hundreds) and grouping them through different sortings, recurring lexical choices can be isolated and further investigated. In the English texts, beside *say* (which accounts for more than one third of the occurrences), and *add* (about a fourth), the following verbs are recurrently used in conjunction with reported speech: *admit, announce, ask, insist, declare, and complain*. In Italian the most frequent are: *dire, dichiarare, aggiungere, affermare, chiamare, concludere, insistere, ribadire*. *Dire* and *aggiungere* are the most common, but they account for less than half of the total of the verbs used to introduce direct speech unlike their English dictionary equivalents *say* and *add* which account for more than 60%. Some of the verbs recurrently used in Italian and in English have cognates or other accepted dictionary equivalents which do not always seem to have the same distribution or frequency. For example, dictionaries give as equivalents for *affermare* a cognate (*affirm*), together with *assert* and *state*. However, while a search for *affermare* produces a considerable list of occurrences, a search for forms of the verbs *assert, affirm* and *state* does not produce a single citation.

Translators can draw the conclusion that *affermare* could be used to translate verbs such as *say* and *declare*, while to translate *affermare* into English with *affirm* or *assert* are not the safest bets in this type of context. Of course, no corpus can tell us if something is not possible (i.e. cannot be said), but it can provide some evidence as to how likely a word or expression is to occur, i.e. how routine it is.

Comparison between related words can also highlight aspects of syntactic preferences. In many cases, learners may not be aware of these differences: they may assume that their knowledge of the language is adequate to provide an
appropriate translation. For instance, words like *prices* and *prezzi* may not seem to present any problems as translation equivalents. A comparison of two concordances, however, reveals some interesting features. Figures 8 and 9 show 1 out of 5 randomly selected occurrences in a corpus of about 500,000 words for each language from Italian and English quality newspapers.

Figure 8
Search word: **PREZZI**. Sort order: 2nd word to left, then 1st word to left

1. recente, sensibile abbassamento dei *prezzi* sui mercati tedeschi ha reso meno
2. tto Saddam", che si è abbatuto sui *prezzi* petroliferi dall'inizio di
3. e di flessione si è avuto anche sui *prezzi* del provolone, mentre il burro ha
4. riforme. Gli stock aumentano, alcuni *prezzi* scendono e inevitabilmente
5. di attenderesi ulteriori aumenti dei *prezzi*". Secondo la fonte, l'incertezza
6. lettriche. Concedendo l'aumento dei *prezzi* chiesto dall'Enel, il Governo
7. erso il terzo round e un aumento dei *prezzi* petroliferi in una condizione
8. ella degli interessi; l'aumento dei *prezzi* si scaricherà su BoT e CcT e
9. usa prima del recupero autunnale dei *prezzi*. Il movimento rialzista non era
10. l mercato comunitario, la caduta dei *prezzi* e la necessità per la
11. in malinconici ammassi; il calo dei *prezzi* internazionali del frumento fa
12. a scorta di manovra ogni volta che i *prezzi* scendevano sotto un livello
13. na allo Stne il tasso di crescita dei *prezzi* in Gran Bretagna dovrà scendere
14. a del mercato del Golfo e crollo dei *prezzi*, anche sull'onda della caduta
15. nsie, hanno generato un crollo dei *prezzi* e reso necessarie delle misure d'
16. he da carne). Domanda molto debole e *prezzi* nettamente cedenti, per contro,
17. Medio Oriente ha mantenuto elevati i *prezzi* delle scadenze lontane e non ha
18. l'aumento del livello generale dei *prezzi* si trasformi in una rincorsa tra
19. ILANO_ Nuova violenta impennata dei *prezzi* sui mercati petroliferi, che
20. ale francese. Anche ieri, intanto, i *prezzi* del petrolio sono aumentati,
21. s produttivi aveva fatto lievitare i *prezzi* del latte. Alla fine _ denunciano
22. 91. Sono infine ancora in recupero i *prezzi* dei cruscam, a seguito della
23. azione che accompagnano i rialzi dei *prezzi* petroliferi". In sostanza, l'
24. ato importanti effetti al rialzo sui *prezzi* mondiali, e al ribasso sull'
25. o l'Europa trascinando al ribasso i *prezzi* della carne. Si è deciso infatti
26. aggior risparmio dalla riduzione dei *prezzi* che sarà provocata dall'
27. è del 3,5%, a causa del rincaro dei *prezzi* dei derivati petroliferi, della
28. sile pubblicato ieri. Il rincaro dei *prezzi* petroliferi innescati dalla crisi
29. etterà probabilmente di capire se i *prezzi* petroliferi sono destinati a
Corpora for Language Learning and the Training of Translators

Figure 9
Search word: PRICES. Sort order: 1st right, 2nd right

1. ...ations against the steep rise in food prices. A court in Marrakesh sentenced
2. ... tonnes might be delivered, squeezing prices acutely. In that event, smaller
3. ... day, but said no relief from soaring prices and food queues could be expected
4. ... will step in themselves to maintain prices, as they did in the wake of the
5. ... eir nadir because of the hike in oil prices. At the same time inflation and
6. ... ven in the 1973/74 collapse, nominal prices did not fall. The "wealth effect"
7. ... s. Properties will stick rather than prices fall." Barclays Bank said that t
8. ... p 3.9 points at 1,775.6. Bombay: Prices fell sharply for the second day
9. ... are traded publicly. Last month junk prices fell sharply on news that Campeau
10. ... market drifted in sympathy. Bombay: Prices fell on heavy speculative selling
11. ... poll discloses a decline in selling prices for the second successive quarter
12. ... weekend sales in New York and Paris prices for post-war French artists soared
13. ... ominated a moderate session in which prices generally fell back from early ad
14. ... delayed to the end of 1991." House prices have already begun to fall across
15. ... ial said. Jaguar, which dropped its prices in North America last year, now
16. ... yield the best opportunities, as the prices of companies fall. Jon Moulton,
17. ... st's cocoa crop indirectly pushed up prices on the London Futures and. Option
18. ... ors. Page 22 World Markets New York: Prices plunged and by the close the Dow
19. ... bank lending. US September consumer prices. Business and City Page 22
20. ... wo years, it doubled again. Property prices soared. But some people did worry
21. ... be reintroduced into the agreement. Prices steadied and ended the day largel
22. ... three years. Many in the trade expect prices to go up further today, and poss
23. ... tly most analysts had expected house prices to begin to recover in early 1991
24. ... per cent higher at 1,169. Brussels: Prices tumbled on the last day of the fo
25. ... pressed concern that the increase in prices was becoming more generalised. Ex
26. ... is evidence to support the case that prices will remain high for a while. Des
27. ... edge that during the contract period prices would remain within the general

What is noted here will depend on the learners’ interests, abilities or needs. For instance, some may note that while prezzi is usually preceded by the definite article i, prices is not. Others may notice instead that nominal groups such as l’aumento dei prezzi or la discesa dei prezzi are more common than verbal groups in Italian, while verbal groups such as prices soared or prices fell are more common in English. Others may enlarge their vocabulary by observing the range of verbs which collocate with prices: prices go up, shoot up, increase, soar, rise and rocket or go down, abate, fall, plunge, tumble, and plummet.
Contrastive analysis of comparable corpora can reveal how similar ideas and concepts are realised linguistically in similar texts in different languages. The analysis may regard stylistic preferences, related conventions of rhetorical and propositional structure, figurative language, and lexico-grammatical features. Comparable corpora provide information on the way discourse is structured as text in different languages, understood as an indication of recurrent patterning rather than as a normative statement.

Corpora reveal regularities, not rules, and the evidence emerging from data is subject to interpretation and always needs to be verified against larger corpora. Learners should also be aware that the evidence provided by, for example, a corpus of quality press articles may differ from those of popular press. The extent to which the features of a specific corpus can be generalised can certainly be the subject of further investigation. While all language learners can benefit from activities based on comparable corpora, trainee translators in particular can be highly motivated in using them, as they not only need to enhance their linguistic competence but also to acquire specific skills related to translation, e.g. cross-linguistic mediation, and knowledge of text type conventions as represented in different languages.

Comparable bilingual corpora are a resource that learners can use to enhance both their analytical skills and their ability to produce accurate target texts.

References

Gavioli L., 1998, "Corpora and the concordancer in learning ESP. And experiment in a course for interpreters and translators", in this volume.


