# Translation as problem solving:<br>uses of comparable corpora

**Serge Sharoff**

Centre for Translation Studies
University of Leeds, Leeds, LS2 9JT, UK
s.sharoff@leeds.ac.uk

**Abstract**

The paper describes an approach that uses comparable corpora as tools for solving translation problems. First, we present several case studies for practical translation problems and their solutions using large comparable corpora for English and Russian. Then we generalise the results of these studies by outlining a practical methodology, which has been tested in the course of translation training.

## 1. The problem

It is widely accepted that translation can be viewed as problem solving: in the process of producing a translation the translator encounters problems of various sorts and uses a set of tools and resources to solve them, cf. (Levý, 1967; Reiß, 2000; Varantola, 2003). Possible problems can involve detecting properties of the source and target audiences, determining the extent of the translation brief, designing the structure of the translated document, etc.

However, problems that occur most frequently in translation of practically every sentence are those of choosing the right target word for rendering source word X in context Y. One type of word-choice problems occurs in translation of terminology: the translator may lack knowledge about the exact translation of term X in domain Z. Another type of problems concerns the choice of words from the general lexicon: the translator knows a word and the standard set of its translations, but cannot find a target word that is suitable for the current context. The obvious way to find a solution for the word-choice problem is by consulting dictionaries. However, dictionary lookup may fail in both cases: a term can be missed in available dictionaries, while translation equivalents for general words suggested in the dictionary may not be usable in the target context. In the worst possible case, a dictionary can mislead the translator by listing a term or source expression with its translation, whilst the translation is NOT used in the target language in the suggested way.

In the following sections I will present several case studies of word-choice problems of the two types and outline ways to solve them using large monolingual corpora. Parallel corpora consisting of original texts aligned with their translations offer the possibility to search for examples of translations in their context. In this respect they provide a useful supplement to decontextualised translation equivalents listed in dictionaries. However, parallel corpora are not representative: millions of pages of original texts are produced daily by millions of native speakers in major languages, while translations are produced by a small community of trained translators from a small subset of source texts. The imbalance between original texts and translations is also reflected in the size of parallel corpora, which are simply too small to account for variations in translation of moderately frequent words. For instance, *frustrate* occurs 631 times in 100 million words of the BNC, i.e. this gives on average about 6 uses in a typical parallel corpus of one million words.

The procedure is illustrated by examples of translations between English and Russian using the corpora listed in Table 1.

All corpora used in the study are quite large, i.e. their size is in the range of 100-200 million words (MW), so that they provide enough contexts for moderately frequent words such as *frustrate*. The size is especially important for the detection of collocates, as even a 10 million-word corpus with its 63 hypothetical instances of *frustrate* does not provide sufficient grounds for deciding whether a single instance of *frustrate one's efforts* represents a recurrent pattern (there are 10 instances of this expression in the BNC). However, the requirement for large corpora does not significantly limit the applicability of this study to other language pairs, as corpora of this size are increasingly available in a variety of languages. The size of about 100 million words is now the standard for so called "National Corpora", such as Czech (Kučera, 2002), Hungarian (Váradi, 2002) or Polish (Lewandowska-Tomaszczyk, 2003). The availability of huge amount of texts on the Internet in a great number of languages can produce Internet-derived corpora of practically arbitrary size, cf. (Kilgarriff and Grefenstette, 2003). What is more, an analysis of Internet corpora used in this study (they were produced by making a random snapshot of 50,000 pages indexed by Google) shows that an Internet-derived corpus is not radically different from the BNC in terms of its coverage of text types and domains. For more information about the properties of Internet-derived corpora see (Sharoff, 2006a).

Access to all corpora is available via a uniform interface (Sharoff, 2006b), which is powered internally by IMS Corpus Workbench (Christ, 1994). In comparison to other approaches using webdata as a corpus, e.g. Linguistic Search Engine (Resnik and Smith, 2003) and WebCorp (Renouf, 2003), the interface offers standard options for concordancing, queries for part-of-speech (POS) tags, detection of collocations and other statistical operations. Thus dealing with Internet corpora is not different in any respect from dealing with standard corpora, such as the BNC or British News.

| **Corpus** | **Size** | **Time frame** |
|---|---|---|
| The British National Corpus | 100 MW | 1970-1992 |
| A corpus of major British newspapers | 200 MW | 2004 |
| The English Internet Corpus | 130 MW | 2005 |
| The Russian National Corpus, a representative Russian corpus comparable to the BNC in its design(Sharoff, 2005) | 100 MW | 1970-2004 |
| A corpus of major Russian newspapers | 70 MW | 2002-2004 |
| The Russian Internet corpus, | 130 MW | 2005 |

Table 1: English and Russian corpora used in the study

## 2. Case studies

The general principle followed in the case studies below assumes gathering a set of expressions in the source language (most typically collocates of the source word or expression), making hypotheses about their translations and testing the hypotheses in the context of target language expressions. All original examples are taken from one of the corpora used (mostly from the newspaper corpus), while translations are provided by the author.

### 2.1. Terminology detection

Rapid development of a field of scientific research or political process produces a host of new concepts which are somehow rendered in both the source and target languages, but are not reflected in dictionaries. However, if they can be found in corpora, there is a possibility of finding a link between them.

For instace, recent political changes in Russia produced a new expression представитель президента (‘representative of president’), which is as yet too novel to be listed in dictionaries or glossaries. At the same time we can use news corpora to identify the people that perform this duty: Драчевский, Латышев, Полтавченко, Черкесов. This can be done by building the list of collocates for the original expression (представитель президента) or by simply browsing through concordance lines. The hypothesis for translation is straightforward: we can search for the English transcription of their names, because they offer more or lesss stable translations. However, even in this simple case there is some variation in the way Cyrillic characters are rendered in English, e.g. letters like -ы- or endings like -ский, which can be rendered as *-sky, skiy, ski* or *-skij*. So it is safer to make a query:

[lemma=$'Drachevsk.*|Lat.shev|Poltavchenko|Kirienko'$][1]

Note that it is unwise to include the first name of the person in question, even if it is frequently supplied in the original Russian text, because it can be omited in English or again transliterated in a less-standard way. The target names in British newspapers are accompanied with the following expressions *Putin’s personal envoy* (twice) and *Putin’s regional representative* (once). From this we can assume that no specific term has been established for this purpose in the British media, but either translation should be acceptable.

A similar technique can be used for the detection of possible translations of a technical term *environmental enforcement*, which is not listed in major English-Russian dictionaries. The most frequent collocates of this expression (counted for the span of 3 words) are *agency, authorities, government, office*. Given that the standard translation of *environment* is окружающая среда, we can make a query in which this term combines with a variety of expressions for government offices, agencies, etc. The frequency of окружающая среда in the three Russian corpora is about 3600, which gives sufficient evidence for detecting its collocates. The range of expressions to be found in this way includes departments and agencies for: охрана (‘protection’, 724 instances), защита (‘guarding’, 234), гигиена (‘hygiene’, 26) of the environment, as well offices for природопользования (‘nature use monitoring’, 82). Again this suggests the lack of a single translation equivalent, but corpora can guide translators about the range of expressions possible for naming environmental enforcement agencies in Russian.

### 2.2. Translating words from the general lexicon

Terminology in any established domain should be stable and allow one-to-one correspondence between the source and target languages. However, as we noticed in the examples above, there is some variation in the use of newly coined terms in domains of rapid development. Anyway we can assume that terminology in such domains will eventually settle down, be recorded in dictionaries and translated consistently. On the other hand, translations of words from the general lexicon depend on the context of their use, so that a dictionary can never give a complete record for all possible translations.

For instance, the Oxford Russian Dictionary lists three Russian translations for *frustrate*: разочаровывать, расстраивать, обескуражен. Yet in the majority of cases the most natural translation into Russian uses a word that does not belong to this set, e.g.

(1) En: *Saddam’s ambition ... is frustrated by the presence of UN inspectors.*
Ru: Стремлению Саддама ... мешает пребывание инспекторов ООН.
Gloss: ‘Saddam’s ambition ... is hampered by the presence of UN inspectors.’

(2) En: *The share offer opens the possibility for thousands of frustrated commuters to air their grievances*

[1] The dot character in regular expressions refers to an arbitrary character, the asterisc to a sequence of such characters, the pipe character (|) to the disjunction operator.

Ru: Благодаря этой доле тысячи недовольных пассажиров получают возможность выразить свои жалобы
Gloss: ‘Thanks to this offer thousands of angry passengers get the opportunity to express their complaints’

There are natural limits on the number of translation equivalents to be listed in a bilingual dictionary, imposed by its size and usability. A printed dictionary cannot afford to give separate translations for derived forms or list dozens of translation equivalents for a relatively unambiguous word, such as *frustrate* (for instance, English monolingual dictionaries list no more than two or three senses for it). As for usability, it is impossible to use a (printed or electronic) dictionary in which the relevant translation is buried in the long list of potential translation equivalents: a translator or a student will not find a translation they want. Entries for polysemous words have already too many suggested translations. For example, the entry for *strong* in the Oxford Russian Dictionary has 57 subentries and yet it fails to mention many word combinations frequent in the BNC, such as *strong feeling, field, opposition, sense, voice*.

The obvious strategy for finding translation equivalents for such examples is to check collocates of target words that are more straightforward for translation. For instance, *voice* in the context of *Her voice was surprisingly strong and powerful* can be reliably translated as голос, so we can produce a list of adjectives collocating with it. The resulting list is long (over 100 adjectives), varied and similar to the collocates the English word *voice* has, including женский (female), громкий (loud), глухой (husky), слабый (feeble), ровный (level), etc. The last adjective is particularly interesting, as the Oxford dictionary gives no suggestion on translating ровный голос, the expression *level voice* is possible in English, but it is nowhere as frequent as the corresponding Russian expression (11 vs. 327 instances in BNC-sized corpora). However, ровный голос fits perfectly into the context for the source example giving a smooth translation

(3) Она сказала это на удивление ровным и властным голосом
‘She said this in a surprisingly level and powerful voice’

What is more this expression ровный голос can be used in the majority of contexts in which *strong voice* occurs in the BNC (unless *strong voice* implies ‘loud voice’), so it can be treated as a reliable translation equivalent worth including in dictionaries.

In the next case study we will encounter a shift in the link between the two languages. If we want to find a translation equivalent for *strong feeling* as in

(4) *In Eastern Europe, meanwhile, . . . nationalist feeling is exceptionally strong*

neither of the two words (*feeling* and *strong*) provides a bridge between the source and target languages. However, *nationalistic* is translated in a restricted number of ways, which helps in building this bridge in two steps. First, we can find nouns correlating with националистский and националистический as two possible translations of *nationalistic*. Nouns that can be relevant in the current context include проявления (manifestations), риторика (rhetoric), убеждения (beliefs), настроения (attitudes), страсть (passion), etc. A separate study of concordance lines discovers that intensifiers for words from the list combined with *nationalist* do not typically come in the form of adjectives (like *strong* in English); they are either nouns or verbs: разгул (raging), разжигать (to fuel), усиление (strengthening). The latter expression can be further intensified by резкий (sharp), if this is what the translator wants to emphasise:

(5) В Восточной Европе тем временем произошло резкое усиление националистических настроений
‘In Eastern Europe, meanwhile, sharp strengthening of nationalistic attitudes has happened’

In the last case study, the context of a problematic expression does not provide any reliable clues for its translation. The translation of *daunting experience* in the following examples:

(6) *Hospital admission can prove a particularly daunting experience.*

(7) *Even though you knew that what you said didn’t matter, it was a daunting experience.*

does not depend on hospital admission or cross-examination, while neither *daunting* nor *experience* can be reliably translated using dictionary equivalents. One way to generalise the context in this case is by using “similarity classes", i.e. groups of words with lexically similar behaviour, cf. Chapter 8.5 in (Manning and Schütze, 1999). The similarity class of a word defines the paradigmatic relationship between it and other words that can appear in similar contexts. This is analogous to the definition of the relationship of synonymy in a thesaurus, but there is a difference, in that the notion of similarity classes is based on the affinity between the contexts in which the words occur. For instance, *strong* has the following similarity class: *powerful, weak, strength, potent, heavy, good, overwhelming, intense, robust, tough, weaken, compelling, fierce*.[2]. It is not the case that *strong* is synonymous with *good, heavy* or *weak*, but this is the case that they all occur in similar contexts. The notion of similarity classes provides an automatic procedure for generalising the contexts of a word in question.

If we compute similarity classes for *daunting* and *experience*,[3] we will get:

(8) daunting ∼ insurmountable (0.347), apprehensive (0.338), alarming (0.328), onerous (0.317), unfamiliar (0.314),

[2] There is no requirement that words in the similarity class have the same POS, even though it happens quite frequently that their POS is also the same because of the similarity of contexts.

[3] We use similarity classes computed using Singular Value Decomposition, as implemented by (Rapp, 2004). Figures in brackets show the relative similarity to the source word (*daunting*) according to the SVD measure.

forgivable (0.306), disconcerted (0.303), trepidation (0.300), incongruous (0.290), complicated (0.289), bleak (0.279), convincing (0.272),

(9) experience ∼ knowledge (0.357), opportunity (0.343), life (0.330), encounter (0.317), skill (0.317), feeling (0.316), reality (0.310), sensation (0.307), dream (0.296), vision (0.279), learning (0.277), perception (0.265), learn (0.263), training (0.263)

In the next step we produce an equivalence class, consisting of translations of words in the similarity class. As the list is large, it is easier to do so using an electronic bilingual dictionary (Oxford Russian Dictionary, in our case). For instance, the equivalence class of the Russian word опыт (experience) includes:

(10) ability, acquire, aptitude, capability, capacity, competence, courage, evidence, experience, experiment, expertise, feasibility, hypothesis, ingenuity, intelligence, knowledge, laboratory, learning, method, opportunity, perception, qualification, rat, research, skill, stamina, statistical, strength, study, talent, technique, test, training, vision.

The result reflects the ambiguity of опыт, which can mean ‘experience’, as well ‘experiment’ (hence the presence of *hypothesis, laboratory* and *rat* in the equivalence class), however it does preserve the semantic core of опыт, which is about skills and abilities.

In the final step we check target language corpora for uses of collocations consisting of members of the two equivalence classes. Even if an equivalence class contains some words that are not relevant to the source example, e.g. *hypothesis* or *rat*, those words create little noise, as they rarely collocate with words in the second equivalence class, e.g. *insurmountable* or *onerous*. Usually, this step brings 30-50 collocates whose relevance to the source language examples can be easily assessed, e.g. it should be obvious for the student that expressions like эффект устрашения (‘deterrent effect’) have nothing to do with the original query *daunting experience*. Then, the contexts of the remaining 5-7 relevant examples can be explored manually. For instance, *daunting experience* brings the following relevant collocates: безрадостный ситуация (dismal situation), волнующая возможность (worrying possibility), мрачный впечатление (gloomy impression), тягостное чувство (onerous feeling), устрашающее впечатление (intimidating impression).

Similarly, for *frustrated commuter/passenger* the procedure brings the following set of potential equivalents: пострадавший пассажир (suffered passenger), неудачный посадка (unfortunate boarding), недовольный пассажир (angry passenger), with the latter being the closest to *frustrated commuters* from the original example (2).

## 3. Considerations for the general methodology

This set of case studies can help in drawing generalisations about the use of corpora for problem solving. Baiscally this involves searching for ‘islands’ of stability in translation, around which we explore and compare contexts in the source and target languages.

In the first step we analyse the context of an expression in question (*environmental enforcement, strong voice, strong feeling*) in order to identify the functions performed by this expression in the source example and possibly in other similar contexts. The second step is to generalise the context of the original example by defining words indicative of the situation in question and extending the list with other words that can perform the same function. If contexts defy a reasonable generalisation, it is possible to use similarity classes, which statistically accumulate contexts most specific for the source expression. The third step is to build a bridge between monolingual corpora in the two languages by translating words with more obvious translation equivalents, such as names, *voice* or *nationalistic*. This step can be facilitated by the availability of a large-scale bilingual dictionary in machine-readable form, in order to produce equivalence classes without human intervention. The case studies presented above used the Russian Oxford Dictionary, some other studies conducted with my students used German and Spanish bilingual dictionaries, also provided by the Oxford University Press. However, it is possible to rely on one’s intuition or to use traditional dictionaries, as it was the case with examples of *personal envoy* or *strong voice*. The final step in the methodology is to study the results of a number of queries in the target language that consist of words in the equivalence class in order to find lines which suggest suitable translation equivalents. If the number of occurrences of equivalent words is not large, as it was the case with the names of relatively obscure Russian political figures, it is possible to start with the study of concordance lines. If the number of concordance lines is too large to allow its direct exploration, as it was the case with *nationalistic* or *voice*, it is easier to study the most significant collocations for words in the equivalence class and then to study patterns consisting of these words with their collocations. Finally, if we use two very large equivalence classes, as it was the case with *daunting experience*, it is reasonable to intersect them in order to find expressions that regularly occur in the target language.

The possibility of applying this methodology is based on several assumptions. First, translators need to have skills in making queries to corpora and analysing lists of collocations and concordance lines. The latter involves skills in vertical reading of concordance lines, as the methodology crucially depends on the ability to notice and describe lexical patterns in raw data. Skills for vertical reading of concordance lines sorted around a keyword are different from those required for horizontal reading of a continuous text. Even if modern-day translators typically cannot do this type of research, a growing number of students in translation studies receive training in corpus linguistics and acquire skills for reading of concordance lines and detecting collocations.

The methodology also assumes the existence of sufficiently large source and target language corpora, such as the BNC as a general-purpose English corpus or the British news corpus for journalistic texts. As noted above, such corpora are increasingly available for a large number of languages. On the other hand, terminology in specific problem domains and register-specific word uses can be studied on the basis of much smaller specialised corpora cf. related work (Bennison and Bowker, 2000; Zanettin, 2002b). For such tasks small disposable corpora can be even more useful, since they include more instances of terms and register-specific constructions to make generalisation specific to this domain. For instance, in a 5 MW corpus of software annotations collected from the Internet using BootCat (Baroni and Bernardini, 2004), there are 35 instances of the expressions *written in Java* and the majority of instances of *written in* are followed by the name of a prorgamming language. In contrast in the 200 MW corpus of British News there are only two instances of *written in Java*, while *written in* is typically followed by dates, locations and names of human languages.

## 4. Conclusions

When large corpora of the type of the BNC are used by translators, they typically provide a confirmation service: they are used to check whether a hypothetical translation equivalent is attested in authentic texts and, if yes, whether it is used in the same function as expected by the corpus user (Varantola, 2003; Zanettin, 2002a). Also students in translation classes can take part in lexicographic excercises which compare the contexts and functions of potential translation equivalents, for instance, *absolutely* and *assolutamente* (Partington, 1998).

In this study we went one step further and proposed a methodology that helps in solving the problem of choosing the right word for an expression. Even if the case studies discussed above solve problems of translation between English and Russian, we tried several exercises of this for various languages, such as Chinese, French, German and Spanish (the other language was English).

The methodology is especially useful for trainee translators. Professional translators have vast experience in finding lexical items that fit well into the context of translation. Some maintain “non-systematic” dictionaries (Palazhchenko, 2002), which highlight words that can cause troubles in translation and interpreting and explain contexts for their translations. Trainee translators on the other hand trust dictionaries, tend to use translations offered in dictionaries and feel frustrated when dictionaries do not provide them with solutions of their problems. Some of the case studies discussed above are not suitable for the practice of professional translators, either because the solution is immediately obvious for them or because finding a solution in this way takes too much of their time. However, the results are rewarding for trainees, because the final description covers not only the translation of a specific word in the context of a single example, but a wider range of contexts in which such words as *voice* and голос are used, as well as conditions for possible translations. This naturally fits into the education plan of trainee translators, which involves equipping them with a range of resources for finding contextually appropriate translations that go beyond what is offered in dictionaries.

The same methodology can be also of help for professional translators, if it is accompanied with automated means for generalising contexts and building bridges between the source and target languages. This link is explored in the ongoing ASSIST project (Sharoff et al., 2006), using semantic tags that are designed as uniform for the two languages, and USAS-EST, a software system for automatic semantic analysis of text that was designed at Lancaster University (Rayson et al., 2004). The semantic tagset used by USAS was originally loosely based on Tom McArthur’s Longman Lexicon of Contemporary English (McArthur, 1981). It has a multi-tier structure with 21 major discourse fields, subdivided into 232 sub-categories.[4] In the ASSIST project, we have been working on a tool that should assign syntactic and semantic tags to texts in comparable corpora and present source and target language examples that are similar in their semantic and syntactic contextual features. We expect that the use of similarity between contexts should reduce the number of irrelevant collocates and present only examples that can be potentially useful in the context of the current problem.

## Acknowledgments

Research presented in this paper has been partly supported by EPSRC grant EP/C005902. I’m grateful to Bogdan Babych, Tony Hartley and Paul Rayson for useful comments on the earlier drafts of the paper.

## 5. References

Marco Baroni and Silvia Bernardini. 2004. Bootcat: Bootstrapping corpora and terms from the web. In *Proc. of the Fourth Language Resources and Evaluation Conference, LREC2004*, Lisbon.

Peter Bennison and Lynne Bowker. 2000. Designing a tool for exploiting bilingual comparable corpora. In *Proceedings of LREC 2000*, Athens.

Oliver Christ. 1994. A modular and flexible architecture for an integrated corpus query system. In *COMPLEX’94*, Budapest.

Adam Kilgarriff and Gregory Grefenstette. 2003. Introduction to the special issue of the web as corpus. *Computational Linguistics*, 29(2):333–347.

K. Kučera. 2002. The Czech National Corpus: Principles, design and results. *Literary and Linguistic Computing*, 17:245–257.

Jiří Levý. 1967. Translation as a decision process. In *To Honor Roman Jakobson: essays on the occasion of his seventieth birthday*, volume II, pages 1170–1182. Mouton, The Hague.

Barbara Lewandowska-Tomaszczyk. 2003. The PELCRA project — state of art. In B. Lewandowska-Tomaszczyk, editor, *Practical Applications in Language and Computers*, pages 105–121. Peter Lang, Frankfurt.

Christopher Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA.

Tom McArthur. 1981. *Longman Lexicon of Contemporary English*. Longman.

Pavel Palazhchenko. 2002. *Moj nesistematicheskij slovar*. Valent, Moscow. (My non-systematic dictionary, in Russian).

[4] For the full tagset, see `http://www.comp.lancs.ac.uk/ucrel/usas/`

Alan Partington. 1998. *Patterns and meanings: using corpora for English language research and teaching*. John Benjamins, Amsterdam.

Reinhard Rapp. 2004. A freely available automatically generated thesaurus of related words. In *Proceedings of the Forth Language Resources and Evaluation Conference, LREC 2004*, pages 395–398, Lisbon.

Paul Rayson, Dawn Archer, Scott Piao, and Tony McEnery. 2004. The UCREL semantic analysis system. In *Proceedings of the workshop on Beyond Named Entity Recognition Semantic labelling for NLP tasks in association with LREC 2004*, pages 7–12, Lisbon.

Katharina Reiß. 2000. Type, kind and individuality of text: decision making in translation. In L. Venuti, editor, *The translation studies reader*, pages 160–171. Routledge, London. Reprinted from 1981.

Antoinette Renouf. 2003. Webcorp: providing a renewable data source for corpus linguists. *Language and Computers*, 48(1):39–58.

Philip Resnik and Noah Smith. 2003. The web as a parallel corpus. *Computational Linguistics*, 29(3):349–380.

Serge Sharoff, Bogdan Babych, Tony Hartley, Paul Rayson, Olga Mudraya, and Scott Piao. 2006. ASSIST: Automated semantic assistance for translators. In *Proc. of the European Association of Computational Linguistics, EACL 2006*, Trento.

Serge Sharoff. 2005. Methods and tools for development of the Russian Reference Corpus. In D. Archer, A. Wilson, and P. Rayson, editors, *Corpus Linguistics Around the World*, pages 167–180. Rodopi, Amsterdam.

Serge Sharoff. 2006a. Creating general-purpose corpora using automated search engine queries. In Marco Baroni and Silvia Bernardini, editors, *WaCky! Working papers on the Web as Corpus*. Gedit, Bologna.

Serge Sharoff. 2006b. A uniform interface to large-scale linguistic resources. In *Proceedings of the Fifth Language Resources and Evaluation Conference, LREC 2006*, Genoa.

Tamás Váradi. 2002. The Hungarian National Corpus. In *Proceedings of the Third Language Resources and Evaluation Conference, LREC 2002*, pages 385–389, Las Palmas de Gran Canaria.

Krista Varantola. 2003. Translators and disposable corpora. In Federico Zanettin, Silvia Bernardini, and Dominic Stewart, editors, *Corpora in Translator Education*, pages 55–70. St Jerome, Manchester.

Federico Zanettin. 2002a. Corpora in translation practice. In Elia Yuste-Rodrigo, editor, *Language Resources for Translation Work and Research, LREC 2002 Workshop Proceedings*, pages 10–14, Las Palmas de Gran Canaria.

Federico Zanettin. 2002b. DIY corpora: the WWW and the translator. In B. Maia, J Haller, and M Ulrych, editors, *Training the Language Services Provider for the New Millennium*, pages 239–248. Porto.