University of León
E-24071 León
Tf. +34 987 291078
Fax. +34 987 291099
E-mail: dfmnrg@unileon.es

BUILDING AN ENGLISH-SPANISH PARALLEL CORPUS FOR TEACHING AND RESEARCH: THE *ACTRES* PROJECT

Noelia Ramón García
University of León (Spain)

INTRODUCTION

ACTRES Project: English-Spanish Corpus-based CA
⇩
Applications in Translation

Step 1: Using semantic criteria ⇨ comparable corpora: Cobuild & CREA (Corpus de Referencia del Español Actual).

Step 2: Using formal criteria ⇨ ready-made parallel corpus.

METHOD

Compilation, selection, & scanning, if necessary, of English originals and their Spanish translations.
Corpus description:
➤ Current size: 447,449 words
➤ Intended full size: 2 million words
➤ Date of texts: from 2000 on.
➤ Length of texts: from ±100 to ± 10,000.
➤ Text types: books, press & miscellanea, intended to be used as a reference corpus.
➤ Language varieties: British English & European Spanish.

	ENGLISH	SPANISH	TOTAL
Books - fiction	84,412	87,195	171,607
Books- non-fiction	35,271	37,806	73,077
Magazines	41,165	46,113	87,278
Newspapers	18,529	20,830	39,359
Miscellanea	34,282	41,846	76,128
TOTAL	213,659	233,790	447,449

Table 1. Internal structure and number of words currently included in the English-Spanish parallel corpus.
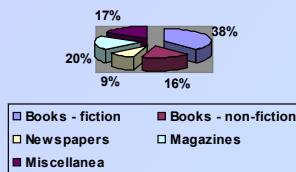

Figure 1. Corpus distribution across text types.

Once in electronic format, parallel texts can subsequently be aligned by means of different programs. One user-friendly option is MULTICONC. A minimum of mark-up will be necessary, basically sentence and paragraph marks.

APPLICATIONS

Intended applications:
➤ TEACHING: Advanced language teaching, English-Spanish Contrastive Analysis & Translation Practice.
➤ RESEARCH: Cross-linguistic studies, descriptive translation studies, translation quality assessment, etc.

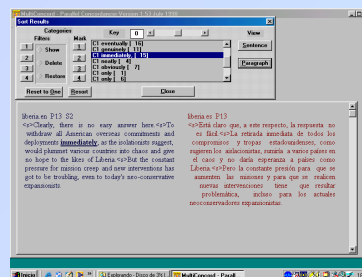Sample alignments with MULTICONC for current research topics:

Figure 2. -ly adverbs and their translations into Spanish. Single sample.
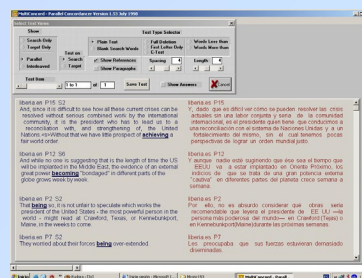

Figure 3. –ing forms and their translations into Spanish. Multiple sample.

CONCLUSIONS

Exploiting our English-Spanish parallel corpus will mean a considerable improvement in teaching procedures involving Advanced language learning or language contrast, as well as a wider range of possible research topics in cross-linguistic studies at all linguistic levels.

REFERENCES

Botley, S.P., A.M. McEnery and A. Wilson (eds.) 2000: *Multilingual Corpora in Teaching and Research*. Amsterdam: Rodopi.
Granger, S., J. Lerot and S. Petch-Tyson (eds.) 2003: *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*. Amsterdam: Rodopi.
Véronis, J. (ed.) 2000: *Parallel Text Processing. Alignment and Use of Translation Corpora*. Dordrecht: Kluwer.