

Bilingual reference corpora for translators and translation studies

Carol Peters	Eugenio Picchi
Istituto di Elaborazione della	Istituto di Linguistica
Informazione	Computazionale
Consiglio Nazionale delle Ricerche	Consiglio Nazionale delle Ricerche
Via S. Maria, 46, 56126 Pisa	Via della Faggiola 32, 56126 Pisa
Italy	Italy
Tel: +39 50 593429	Tel: +39 50 560481
Fax: +39 50 554342	Fax: +39 50 589055
E-mail: carol@iei.pi.cnr.it	E-mail: picchi@ilc.pi.cnr.it

A recent trend in corpus linguistics is the construction of bilingual (or multilingual) reference corpora. So far most studies on bilingual text archives have been on parallel corpora, i.e. sets of translationally equivalent texts. Such corpora are important sources of data when studying the translation process. However, it has been suggested that a more reliable resource for other types of contrastive linguistic studies and applications is a comparable corpus, i.e. sets of texts from pairs or multiples of languages that concern a given domain and can be contrasted because of their common features. In the first case, translation equivalents can be found; in the second case, it should be possible to identify natural language lexical equivalents. We will present procedures that have been developed at the Istituto di Linguistica Computazionale, Pisa, to construct and query both parallel and comparable bilingual text corpora and to extract significant data for translation purposes and contrastive textual studies, giving examples of the different types of results that can be obtained. The procedures form part of an integrated mono- and bilingual text and lexical database system, implemented in a prototypal version at the Institute.