

SABRINA AULITTO*

Corpus multilingue: strumento di osservazione e strategie d'uso

SOMMARIO: 1. Introduzione. – 2. Definire un corpus. – 3. Il contenuto del corpus. – 4. Corpus-Didattica: un binomio possibile? – 5. Conclusione.

Un corpus n'est pas un objet simple, mais le fruit des choix des linguistes qui le rassemblent. L'objet corpus est censé représenter quelque chose, et ne doit pas être une simple collection de phrases ou un sac de mots¹. L'ensemble qui forme un corpus est un assemblage de textes, il ne faut pas que la forêt cache l'arbre textuel².

1. Introduzione

«Una delle discipline che affonda le sue radici nelle origini della linguistica computazionale stessa è la linguistica dei corpora, la quale si occupa del trattamento del linguaggio naturale ai fini della produzione di lessici di testi specifici, di concordanze e di studi statistici»³.

Grandi corpora di testi sono da tempo strumenti d'uso quotidiano nelle mani dei linguisti per ricerca pura o applicata alla redazioni di dizionari, parallelamente, in un contesto di studi universitari queste risorse sono impiegate ai fini di un'esplorazione del linguaggio nelle sue molteplici sfaccettature.

Le applicazioni ideate e prodotte dai Centri di ricerca per eccellenza italiani e stranieri, sono state multiformi e nelle varie lingue euro-americane, interessandosi all'elaborazione di banche dati relative a plurimi set-

* Università Cattolica di Milano.

¹ F. RASTIER, *Enjeux épistémologiques de la linguistique de corpus*, in «TEXTO!» [En ligne], URL: <http://www.revue-texto.net/index.php?id=543>.

² G. WILLIAMS, *Introduction*, «TEXTE ET CORPUS», 3, août, 2008, Actes des Journées de la linguistique de Corpus 2007, p. 7.

³ Cfr. *ivi*, p. 23.