

# Teaching language variation

## The exploitation of Italian dictionaries and corpora

Isabella Chiari

Università “La Sapienza” di Roma, Dipartimento di Studi Filologici Linguistici e Letterari  
isabella.chiari@uniroma1.it

**Keywords:** Italian language, linguistic variation, electronic dictionaries, spoken and written language corpora, language teaching

### 1. Introduction

The aim of this paper is to evaluate and compare Italian spoken and written language corpora and electronic dictionaries as tools for teaching language variation to non-native learners. Language variation is traditionally one of the most crucial and complex ability that second language learners are asked to acquire: formal and informal styles (at phonological, morphological, syntactic and textual levels) and in general diaphasic variations, written vs. spoken selections, diatopic differences, diversity in linguistic style of speakers of different ages, genders, social classes or identities, terminology etc. (cfr. Biber, 1989, 1994; Lee 2001).

Is it possible to acquire this knowledge by exploiting material extracted from electronically available dictionaries and corpora, individually or in guided classroom activities? The present paper focuses on linguistic resources accessible for the Italian language: comparing and evaluating the possibilities offered by the different search tools accessible in electronic dictionaries and in reference language corpora recently designed. What kind of information about linguistic varieties is coded in dictionaries and corpora of the Italian language? Is it possible to exploit it in a learning environment?

### 2. Exploitation of Italian electronic dictionaries

Electronic dictionaries of the English language nowadays include the faculty for the individual user to access not only single annotated entries, but real examples extracted from corpora and full access to large samples of reference corpora (like the *Oxford Advanced Learner's Dictionary of Current English*, 2000, *Longman Dictionary of Contemporary English*, 2001, *Cambridge Dictionary of English*, 2001 and *Cambridge Advanced Learner's Dictionary*, 2003, *Collins Cobuild English Dictionary for Advanced Learners*, 2001). Italian electronic dictionaries do not include these feature at the moment, although in some cases examples from authentic texts is provided. Nevertheless many are the characteristics of Italian electronic dictionaries can be usefully employed in language research and teaching (cfr. Iacobini 2003; Chiari 2006), and more specifically for the development of language variation detection and usage skills in L2 learners.

A selection of the best known dictionaries of the Italian language has been analyzed in order to pick labeled characteristics that let the user extract knowledge about language variation and use from single or multiple entries, through advanced search features. Among the

dictionaries analyzed and compared are: GRADIT, *Grande Dizionario italiano dell'uso* (1999-2000), directed by Tullio De Mauro; *Dizionario della lingua italiana* (Paravia, 2000); DISC by Francesco Sabatini and Vittorio Coletti (2003); Zingarelli (2006), by Miro Dogliotti e Luigi Rosiello; Giacomo Devoto, Gian Carlo Oli, *Il Dizionario della lingua italiana 2004-2005*, edited by Luca Serianni e Maurizio Trifone.

Some of the features examined for comparing dictionaries are: capabilities to generate and export word lists, information about frequency of usage in written and spoken texts, presence/absence of searchable labels indicating formal/informal usage, spoken/written usage, regional and/or dialect variations, usage mark-up, presence of authentic examples, miscellaneous features like word games generation, and other possible teaching tools other than general retrieval software.

### **3. Exploitation of Italian available language corpora**

Nowadays linguistic reference corpora and computer learner corpora are becoming widely used for teaching purposes and a number of theoretical investigations on their potential has been pointed out (among the copious bibliography cfr. McEnery AM, Wilson A 1997; Bernardini, S. 2000; Gavioli, L. 2001). I will further identify some of the most interesting features of Italian reference corpora for the extraction of linguistic variation information as teaching aids.

Some of the major reference corpora of the Italian language have been evaluated. Written language corpora such as CORIS/CODIS (*CORpus di Riferimento dell'Italiano Scritto*, 1998), and spoken language corpora like LIP (*Corpus del Lessico di frequenza dell'italiano parlato*, De Mauro et al.1993) and C-ORAL-ROM (Cresti & Moneglia 2005) and some minor corpora have been observed. Moreover corpora have been inspected in their design and retrieval tools in order to examine their coverage of different variation attributes: characteristics of concordancing and retrieval applications, accessibility of information about phonetic, morphological, lexical and syntactic aspects of variation (in the linguistic annotation and in the retrieval process), information about frequency and usage of lexical units, access to audio samples of corpus material (for the inspection of phonetic features, ipohyperarticulation, intonation) in case of spoken language corpora, etc.

### **4. Linguistic variation accounts**

Evaluation is made according to different parameters: stage of linguistic knowledge (beginner, intermediate, advanced), level of analysis accounted for (phonetic/phonological, morphological, syntactical, lexical, textual, pragmatic), inclusion of explicit information about variation, word and other units frequency rank, exporting capabilities, etc.

Furthermore I suggest some uses of linguistic variation information in language teaching environments, such as curriculum design in the selecting and sequencing process, language materials and resources design, development and grading, in assessment procedures and in classroom work methodology.

## **References**

Bernardini, S. (2000) Systematising serendipity: Proposals for concordancing large corpora with language learners. In Burnard, L. and T. McEnery (eds). *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora*. (Lodz Studies in Language). Hamburg: Peter Lang, 183-190.

- Biber, D. (1989) A typology of English texts. *Linguistics*, 27(1), 3-43.
- Biber, D. (1994) An analytical framework for register studies. In D. Biber & E. Finegan (eds), In *Sociolinguistic perspectives on register*. New York: Oxford University Press, 31-56.
- Chiari, I. (2006) Dizionari elettronici e database online in glottodidattica: possibilità attuali e potenzialità future, *XIV Convegno GISCEL Lessico e apprendimenti*, Siena, 6-8 aprile 2006.
- Cresti E. & M. Moneglia (2005) *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*, John Benjamins, Amsterdam.
- De Mauro, T., F. Mancini, M. Vedovelli, M. Voghera (eds) (1993) *Lessico di frequenza dell'italiano parlato*, Etas libri, Milano.
- G. Aston (ed) (2001) *Learning with corpora*. Houston, TX: Athelstan.
- Godwin-Jones, B. (2001) Emerging Technologies. Tools and Trends in Corpora Use for Teaching and Learning, *Language Learning & Technology*, 5(3), 7-12.
- Iacobini C. (2003) I dizionari della lingua italiana su cd-rom come strumento per la ricerca linguistica. In F. Rainer / A. Stein (a cura di), *I nuovi media come strumento per la ricerca linguistica*, Frankfurt, Peter Lang, 61-73.
- Lee, David YW. (2001) Genres, Registers, Text Types, Domains, And Styles: Clarifying the Concepts and Navigating a Path Through the BNC Jungle, *Language Learning & Technology*, 5(3), 37-72.
- McEnery AM, Wilson A (1997) Teaching and Language Corpora (TALC). *ReCALL* 9(1), 5-14.