

C-ORAL-ROM – Integrated reference corpora for spoken romance languages

Umfangreiches Korpus zur gesprochenen Sprache in Italien, Frankreich, Spanien und Portugal, entstanden in Zusammenarbeit von Dozenten der Universitäten Florenz, Aix-Marseille, Madrid und Lissabon zum Ziele des gesamtromanischen Sprachvergleichs.

Die Sprachaufnahmen sind sowohl als Audiofiles als auch als transkribierte Textdateien (ca. 300.000 Wörter pro Sprache) verfügbar. Die Transkriptionen sind lemmatisiert und PoS-Annotiert. Zudem enthält die DVD die Metadaten der Sprecher, eine Übersicht über die diversen Kommunikationssituationen, in denen die Aufnahmen entstanden sind, Frequenzlisten sowie Software zur Analyse der Korpora.

Sprache	Italienisch, Französisch, Spanisch, Portugiesisch
Sprachliche Realisierung	mündlich
Umfang	ca. 1,2 Mio. Wörter (300.000 pro Sprache)
Medium	Sprachaufnahmen in diversen Kommunikationssituationen (privat und öffentlich, Nähe und Distanz, Monolog und Dialog, formell und informell)
Geographischer Ursprung	Italien, Frankreich, Spanien, Portugal
Zeitliche Einordnung	ab 1970
Form der Daten	Audiofiles und transkribierte Textdateien; die DVD enthält Software zur Bearbeitung und Analyse der Daten
Format	WAV, CHAT, TXT, IMDI, XML
Annotation	prosodische Annotation ("terminal" und "non terminal breaks"), lemmatisiert, part-of-speech-annotiert
Mögliche Suchabfragen	Suche nach <i>key words in context</i> , Wortfolgen und Wortarten sowie Erstellung von Frequenzlisten möglich
Buch + DVD	Cresti, Emanuela/Moneglia, Massimo (eds.) (2005): <i>C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages</i>. Amsterdam / Philadelphia: Benjamins.
Infos online	C-ORAL-ROM Vollversion
Zugang im FU-Netz	Datenbanken der FU und Philologische Bibliothek
Zum Zitieren:	Cresti, Emanuela & Massimo Moneglia (eds.). 2005. <i>C-Oral-Rom: Integrated Reference Corpora for Spoken Romance Languages</i> . C-ORAL-ROM. Amsterdam: Benjamins.