

Português Falado – Documentos Autênticos: gravações áudio com transcrição alinhada

The four published CD-ROMs include a spoken Portuguese corpus collected among sociolinguistically diverse speakers having Portuguese as mother tongue or as second language. This corpus consists of informal conversations between acquaintances, friends or relatives as well as formal acts as, for instance, radio programs or conferences. In a total of 86 recordings, the texts exemplify the Portuguese spoken in Portugal (30), in Brazil (20), in the African countries with Portuguese as its official language – Angola, Cape Verde, Guinea-Bissau, Mozambique and São Tomé and Príncipe (5 each) –, in Macao (5), in Goa (3) and in East-Timor (3), corresponding to 8h44m of recording and to 91.966 tokens. The recordings cover a period that goes from 1970 to 2001, falling upon about 70% on the last decade.

These samples of Portuguese varieties are distributed in the four CD-ROMs in the following way:

- 1 – Portugal (recordings from the nineties);
- 2 – Portugal (recordings from the seventies and the eighties), Macao, São Tomé and Príncipe and East-Timor;
- 3 – Angola, Cape Verde, Guinea-Bissau and Mozambique;
- 4 – Brazil and Goa.

Finally, 94 speakers appear in the recordings; their characterizations (origin, sex, age, professional status, level of education) are visible on the header of each transcription, in which is also given information about the place, date and situation in which the recording was made, as well as other relevant types of information.

For each recording the user will find text and sound files, which can be handled on their own (when opened with normal text or sound editors), and a software tool that, together with the application files included in the CD-ROMs, allow text to sound alignment. This application starts automatically when the CD-ROM is inserted in the computer. The user can, then, open the file he wants to work with; in the toolbar he has buttons similar to those in any record player that allow him to handle easily with the document. After pushing the play button, a colored light runs over the transcription of the sequence that is being listened to. The user can control what he is listening to, repeating sequences or jumping parts of the text, either using the buttons on the toolbar or clicking with the mouse on the preferred parts of the text.

LINK

