

Multimedia Corpora for Applied Linguistic Contexts

Ylva Berglund Prytz,¹ Sabine Braun,²
Kurt Kohn³ and Pascual Pérez-Paredes⁴

Abstract

Applied linguistic uses of corpora are still restricted by a number of factors. Broadly speaking, traditional corpora usually follow the requirements of descriptive linguistics, whereas applied linguistics has specific needs in terms of corpus content and size, annotation and query (cf. Braun 2005). The need for pedagogically motivated corpora is well reflected in the many initiatives to create ‘home-made’, ‘ad-hoc’ or ‘DIY’ corpora for language learning, teacher education or translator/interpreter training (cf. e.g. Amador Moreno, Chambers and O’Riordan 2006; Aston 2002; Ghadessy, Henry and Roseberry 2001; Tribble 1997; Varantola 2003; Zanettin 2001) and in Seidlhofer’s (2002) call for ‘local learner corpora’.

These initiatives have focussed on written corpora. Spoken discourse is becoming increasingly important in international communication and hence in applied linguistics. The usefulness of spoken language corpora in the pedagogical context has been emphasized by many researchers (e.g. Mauranen 2004; McCarthy 1998). Spoken corpora in multimedia format (sometimes called multimodal corpora) will greatly facilitate corpus use in applied linguistic contexts, as they foster the contextualisation and open up entirely new ways of exploitation (e.g. listening comprehension activities). On the other hand, the creation of multimedia corpora is a labour-intensive enterprise, which would seem to invalidate any approach that focuses too narrowly on one specific purpose.

In this paper we will discuss the challenges involved in the creation of spoken multimedia corpora for applied linguistic purposes. The points for discussion will include corpus contents and transcription issues, relevant query options and their implications for corpus annotation and media alignment. The paper will be based on two current projects in this area: the EU-funded project SACODEYL, which aims to create teenage language corpora in seven languages, and the ELISA corpus, which includes interviews with native speakers of different varieties of English giving accounts of their professional lives.

References

- Amador Moreno, C., A. Chambers and S. O’Riordan (2006) Integrating a corpus of classroom discourse in language teacher education: the case of discourse markers *ReCALL* Vol 18 (1): 83–104.

¹ Research Technologies Service, University of Oxford

² Centre for Translation Studies, University of Surrey
e-mail: S.Braun@surrey.ac.uk

³ Applied English Linguistics, University of Tuebingen

⁴ Departamento de Filología Inglesa, Campus de la Merced, Universidad de Murcia

- Aston, G. (2002) The learner as corpus designer. In B. Kettemann and G. Marko (eds) *Teaching and learning by doing corpus analysis*. Amsterdam: Rodopi, pp. 9–25.
- Braun, S. (2005) From pedagogically relevant corpora to authentic language learning contents, *ReCALL* Vol 17 (1): 47–64.
- Ghadessy, M., A. Henry and R. Roseberry (eds) (2001) *Small corpus studies and ELT Theory and practice*. Amsterdam: Benjamins.
- Mauranen, A. (2004) Spoken corpus for an ordinary learner. In J. Sinclair (ed) *How to Use Corpora in Language Teaching*. Amsterdam: Benjamins, pp. 89–104.
- McCarthy, M. (1998) *Spoken language and Applied Linguistics*. Cambridge: CUP.
- Seidlhofer, B. (2002) Pedagogy and local learner corpora: working with learning-driven data. In S. Granger, J. Hung and S. Petch-Tyson (eds) *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: Benjamins, pp. 213–34.
- Tribble, C. (1997) Improvising corpora for ELT quick-and-dirty ways of developing corpora for language teaching. In B. Lewandowska-Tomaszczyk and P. Melia (eds) *PALC'97 practical applications in language corpora*. Lodz: Lodz University Press, pp. 106–17.
- Varantola, K. (2003) Translators and disposable corpora. In F. Zanettin, S. Bernardini and D. Stewart (eds) *Corpora in Translator Education*. Manchester: St Jerome, pp. 55–70.
- Zanettin, F. (2001) DIY corpora the WWW and the translator. In B. Maia, H. Haller and M. Urlrych. *Training the language services provider for the new millennium*. Porto Faculdade de Letras: Universidade do Porto, pp. 239–48.