

Representing and analysing bilingual conversations

María del Carmen Parafita-Couto
Bangor University
maricarme@gmail.com

Kevin Donnelly
Bangor University
kevin@dotmon.com

Abstract

We present methods of representing and analysing bilingual conversations using software available under the GPL or other open-source licenses. The software has been used to record and analyse 3 corpora of almost a million words in total, in three languages (Welsh, Spanish and English). We discuss some of the practical lessons learned, and review the scope for computer-aided linguistic analysis of the corpora.

1 Introduction

Este é o estilo usado pelos artigos da revista *Linguamática*. O estilo pode ser usado em L^AT_EX (Mittelbach et al., 2004), Microsoft Word ou OpenOffice.org.

Many of the conversations going on in the world right now are taking place between people who speak more than one language. Not widely studied until recently. Trad ling concentrates on one language, mod ling concentrates on micro-aspects of a single language. Some major researchers, and refs.

The ESRC Centre - history and work. The corpora.

1.1 Siarad

1.2 Patagonia

1.3 Miami

2 *Representing bilingual conversations*

2.1 Fieldwork

Location, informants, equipment, procedure.

2.2 CLAN

History, usage, Talkbank. Other software.

2.3 Transcription

Conventions, proofreading, checking. Version control.

2.4 Glossing

Tagging - MOR, not available for Welsh. Autoglossing, accuracy, additional benefits. Aperium.

3 *Analysing bilingual conversations*

3.1 Presentation

Type-setting.

3.2 Dependency analysis

Whenever! Currently using very unsophisticated clause-splitter.

3.3 Conversation depiction

Profile images. Word-cloud?

3.4 Variation analysis

Goldvarb, Rbrul.

3.5 Codeswitched collocations

MC stuff.

3.6 Triggered codeswitching

Diana stuff.

4 *Conclusions*

Major high-quality resources for further study. Type of speech that is not commonly recorded.

Time-consuming to prepare to a high standard.

4.1 FLOSS

FLOSS crucial in preparing the material - open science, reproducible science.

4.2 Computer-assisted

Allows new aspects to be looked at that would not be practicable otherwise.

×	1	2	3	4	5
1	1	2	3	4	5
2	2	4	6	8	10
3	3	6	9	12	15
4	4	8	12	14	20
5	5	10	15	20	25

Table 1: Tabela multiplicativa.

5 *Fórmulas Matemáticas*

Sugere-se que as fórmulas matemáticas sejam devidamente numeradas para que se possam referir posteriormente.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

E agora é possível comentar devidamente a fórmula 1.

6 *Tabelas e Figuras*

As tabelas e figuras devem ser devidamente numeradas, e devem ter uma legenda. Caso a tabela ou figura não caiba numa única coluna é possível obrigar o L^AT_EX a ocupar a largura da página com o ambiente `table*` e `figure*` respectivamente. You can refer to Table 1.

Acknowledgements

The support of the Arts and Humanities Research Council (AHRC), the Economic and Social Research Council (ESRC), the Higher Education Funding Council for Wales and the Welsh Government is gratefully acknowledged. The work presented in this paper was part of the programme of the ESRC Centre for Research on Bilingualism in Theory and Practice at Bangor University.

References

Mittelbach, Frank, Michel Goossens, Johannes Braams, David Carlisle, and Chris Rowley. 2004. *The L^AT_EX Companion*. Addison-Wesley, 2 edition, April. ISBN-10: 0201362996 ISBN-13: 978-0201362992.