

Inter–Judge Agreement in Transcribing Dialectal Data: a Study of a Corpus of Dialectal Portuguese

U. PORTO

PEDRO TIAGO MARTINS¹, JOÃO VELOSO^{1,2}

martins@letras.up.pt, jveloso@letras.up.pt

¹Centro de Linguística da Universidade do Porto, ²Faculdade de Letras da Universidade do Porto

CLUP Centro de Linguística da Universidade do Porto

ARQUIVO Dialectal DO CLUP

ABSTRACT

In this preliminary study, we discuss how Inter–Judge Agreement—a widespread method of measuring the reliability of phonetic transcription of child and impaired speech—can also be applied to research on dialectal variation. By analyzing the results of a small scale transcribing experiment based on a corpus of dialectal samples of European Portuguese, our proposal is that inter-judge agreement enhances the accuracy and reliability of phonetic transcription of this kind of speech.

1. BACKGROUND

Some common problems with phonetic transcription (PhT) [1, 2]:

- ▶ Discrepancies in intended level of detail;
- ▶ Inaccurate productions;
- ▶ Low quality of acoustic samples;
- ▶ Lack of appropriate phonetic symbols.

Difficulties escalate when it comes to PhT of “atypical” productions [3, 4, 5, 6], e.g.:

- ▶ Child language;
- ▶ Impaired speech.

How can we deal with such difficulties? Dubious, unclear productions, possibly leading to a biased transcription (transcribers don’t transcribe what is in the acoustic signal, but what their linguistic knowledge assumes is in the signal), should be subject to either:

- ▶ **Acoustic Analysis** [7, 8]: As a way of objective detection/identification of segments and/or phonetic properties of segments;
- ▶ **Cross-checking of different PhTs**—“Inter–Judge Agreement”, IJA [4, 5]: The same phonetic sample is transcribed by different transcribers; consensus among them is then measured, as a way of improving *reliability* of problematic PhTs. Different authors propose different levels of reliability:

▶ Shriberg et al., 1999 [6]: >80% consensus → Reliable

▶ Shrout & Fleiss, 1979 [9]:

Percentage	Level of reliability
100% consensus	Full reliability
75% consensus	Excellent reliability
40%–74% consensus	Moderate reliability
<40%	Poor reliability

REFERENCES

- [1] International Phonetic Association. 1999. *Handbook of the International Phonetic Association. A Guide to the Use of the International Phonetic Alphabet*. Cambridge: Cambridge University Press.
- [2] Wells, J. C. 2006. Phonetic Transcription and Analysis. In: K. Brown (Ed.). *Encyclopedia of Language and Linguistics*. Amsterdam: Elsevier, 386–396.
- [3] Shriberg, L. D.; Kwiatkowski, J. 1982. Phonological disorders III: A procedure for assessing severity of involvement. *Journal of Speech and Hearing Research*. **47**: 256–270.
- [4] Shriberg, L. D.; Kwiatkowski, J.; Hoffmann, K. 1984. A procedure for phonetic transcription by consensus. *Journal of Speech and Hearing Research*. **27**: 456–465.
- [5] Shriberg, L. D.; Lof, G. L. 1991. Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics*. **5**(3): 225–279.
- [6] Shriberg, L. D.; Tomblin, B.; McSweeney, J. 1999. Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language and Hearing Research*. **42**: 1461–1481.
- [7] Tyler, A. A.; Edwards, M. L.; Saxman, J. H. 1990. Acoustic validation of phonological knowledge and its relationship to treatment. *Journal of Speech and Hearing Disorders*. **55**: 251–261.
- [8] Rvachew, S. 2003. Computer applications and treatment outcomes. *Perspectives on Language Learning and Education*. **10**(1): 17–20.
- [9] Shrout, P. E.; Fleiss, J. L. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*. **86**: 420–428.
- [10] Andrade, A. 1994. Estudo Acústico de Sequências de Oclusivas em Português Europeu. *Actas do IX Encontro Nacional da Associação Portuguesa de Linguística*. Lisbon: APL/Colibri, 1–15.

2. PURPOSE OF THIS STUDY

Our proposal is to apply IJA metrics to a specific set of phonetic samples consisting of a collection of oral productions from several dialects of European Portuguese in order to measure the reliability of PhTs of such corpus.

Brief Description of the Dialectal Corpus:

- ▶ Dialect Archive of the Center of Linguistics of the University of Porto (AD-CLUP);
- ▶ ca. 120 samples of spoken European Portuguese, encompassing several dialects;
- ▶ All samples were collected by graduate and post-graduate students of Linguistics from the University of Porto since 1994;
- ▶ Varying degrees of recording quality, methodology and duration (average is about 90s);
- ▶ All samples were subject to logistic and linguistic treatment (including exhaustive orthographic and phonetic transcriptions);
- ▶ All data (samples, transcriptions, maps, statistics, etc.) will soon be made freely available online.

Problematic Cases of Phonetic Variation in European Portuguese: In transcribing EP, the following types of variation raise a considerable amount of doubts and difficulties:

- ▶ **Schwa–realization vs. Schwa–deletion:** In EP, schwa ([i]) may be freely deleted, regardless of context; even when it is phonetically realized, acoustic analysis is generally misleading as to its occurrence, due to its “obscure” acoustic characterization (very short duration, low intensity, undefined formant patterning)[10];
- ▶ **Rhotics:** EP opposes a “simple” rhotic ([r]) to a “multiple” one, whose phonetic realization may range from [r] to [χ] and [ʁ]. This variation corresponds to an on-going change and it is not properly accounted for in the literature;
- ▶ **Voiced stops:** voiced stops /b d g/ are often realized as “assibilated” (i.e., as non-strident approximants): [β ð ɣ]. Although some contexts seem to favor this kind variation, it is largely unpredictable from context (or dialect);
- ▶ **Final unstressed syllables:** final, unstressed syllables can be produced as fully unvoiced in EP, even if phonologically voiced.

4. RESULTS/REMARKS

Sample	LJ	CP	ARV	ML	CS	VF	CA	Consensus
1	deletion	deletion	deletion	deletion	deletion	deletion	deletion	deletion**
2	deletion	deletion	deletion	deletion	deletion	deletion	deletion	deletion**
3	i	deletion	i	i	i	deletion	deletion	i*
4	deletion	deletion	deletion	i	deletion	deletion	deletion	deletion**
5	deletion	deletion	i	i	i	i	deletion	i*
6	ʁ	ʁ	r	ʁ	r	r	r	r*
7	ʁ	ʁ	r	ʁ	r	r	r	r*
8	ʁ	ʁ	r	–	r	r	r	r*
9	ʁ	ʁ	r	ʁ	r	r	r	ʁ*
10	ʁ	ʁ	r	ʁ	r	r	r	r*
11	d	d	d	d	ð	ð	ð	d (stop)*
12	b	b	b	–	β	b	b	b (stop)**
13	b	b	b	b	β	b	b	b (stop)**
14	b	b	b	b	β	b	b	b (stop)**
15	g	g	g	g	ɣ	g	g	g (stop)**
16	voiced	deletion	deletion	voiced	voiced	deletion	deletion	deletion*
17	voiced	voiced	voiced	voiced	voiced	deletion	devoiced	voiced**
18	deletion	deletion	–	voiced	voiced	deletion	voiced	–
19	voiced	voiced	voiced	voiced	voiced	voiced	voiced	voiced**
20	voiced	voiced	voiced	voiced	voiced	voiced	voiced	voiced**

Table 2: Results of the transcription task. (*)40%<IJA<75%; (**)75%≤IJA

Final Remarks:

- ▶ PhTs’ reliability is improved: initial doubts about the chosen samples were solved (see Table 2) and the final transcriptions to be included in the online material of AD-CLUP will follow these results;
- ▶ Therefore, the overall reliability of the whole linguistic study of the AD-CLUP’s material is also improved;
- ▶ IJA methodology proved efficient for the study/transcription of dialectal realizations;
- ▶ Since it has proven beneficial, this approach will be extended to the whole corpus of AD-CLUP in the future.

ACKNOWLEDGMENTS

This work was partly funded by the *Fundação para a Ciência e a Tecnologia/Centro de Linguística da Universidade do Porto* (PEst–OE/LIN/UI0022/2011). We would like to thank Luís M. T. Jesus (ESSUA & IEETA, University of Aveiro) for helping us with the experiment.