The following paper was presented at The 9th Workshop on Disfluency in Spontaneous Speech (DiSS 2019) held at ELTE Eötvös Loránd University in Budapest, Hungary on 12–13 September, 2019.

Title: Processing disfluencies in distinct speaking styles: Idiosyncrasies and transversality

Author(s): Helena Moniz

Abstract: This talk will tackle the idiosyncratic properties of disfluencies in distinct speaking styles, mostly university lectures (Trancoso et al., 2008) and map-task dialogues (Trancoso et al., 1998), but also featuring verbal fluency tests, and (more recently) second language learning presentations in ecological settings. It will also discuss the transversal acoustic-prosodic properties pertained across speaking styles. The main research questions are twofold: i) are there domain effects in the production of disfluencies when speakers adjust to distinct communicative contexts, as in university lectures and dialogues?; ii) if domain effects do exist, are there still acoustic-prosodic properties that can be shared across domains?

DOI: <https://doi.org/10.21862/diss-09-001-moniz>

Citation (JIPA): Moniz, Helena. 2019. Processing disfluencies in distinct speaking styles: Idiosyncrasies and transversality. In: R. L. Rose & R. Eklund (eds.), *Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech*, 12–13 September, 2019, Budapest, Hungary, 1–2.

The complete proceedings for DiSS 2019 are available as follows.

ISBN: 978-963-489-063-8

DOI: <https://doi.org/10.21862/diss-09>

DiSS 2019 was sponsored by The Faculty of Humanities, ELTE Eötvös Loránd University and the International Speech Communication Association (ISCA).

Processing disfluencies in distinct speaking styles: Idiosyncrasies and transversality

Helena Moniz

School of Arts and Humanities, University of Lisbon, Lisbon, Portugal

Spoken Language Systems Laboratory, INESC-ID, Lisbon, Portugal

This talk will tackle the idiosyncratic properties of disfluencies in distinct speaking styles, mostly university lectures (Trancoso et al., 2008) and map-task dialogues (Trancoso et al., 1998), but also featuring verbal fluency tests, and (more recently) second language learning presentations in ecological settings. It will also discuss the transversal acoustic-prosodic properties pertained across speaking styles. The main research questions are twofold: i) are there domain effects in the production of disfluencies when speakers adjust to distinct communicative contexts, as in university lectures and dialogues?; ii) if domain effects do exist, are there still acoustic-prosodic properties that can be shared across domains?

As for speaking style effects in the production of disfluencies (Moniz et al., 2014), results show that there are statistical significant differences in the acoustic-prosodic parameters when speakers adjust to distinct communicative contexts. Although there is a statistically significant cross-style strategy of prosodic contrast marking (pitch and energy increases) between the region to repair and the repair of fluency, this strategy is displayed differently depending on the specific speech task, with a stronger prosodic contrast marking of disfluency-fluency repair on university lectures. In this respect, disfluencies can also be considered as a feature of a charismatic university teacher. The speaker perceived as more fluent monitors the range of energy and f0 slopes from the *reparandum* to the repair of fluency, showing also more systematic temporal measures, mirroring the behaviour of spontaneous dialogues and further enriching the class with the dynamics of a spontaneous dialogue.

Building upon the linguistic analysis, automatic speech processing experiments will also be described aiming at shedding light on the impact of the idiosyncrasies/transversality of the disfluencies. One of those experiments discriminates between list effects, disfluencies, and other linguistic events in an animal naming task (Moniz et al., 2015a). Recordings from 42 Portuguese speakers were automatically recognized and AuToBI (Rosenberg, 2010) was applied in order to detect prosodic patterns, using European Portuguese and English models. Both models allowed to differentiate list

effects from the other events, mostly represented by the tunes: L* H/L(−%) (English models) or L*+H H/L(−%) (Portuguese models) contrasting with the *plateau* of the most frequent disfluency in the corpus, filled pauses. However, English models proved to be more suitable because they rely on substantial more training material.

Results on cross-domain experiments and the robustness of acoustic-prosodic features will be presented (Moniz et al., 2015b). The main trend found is that models can be quite robust across corpora for this task, despite their distinct nature. The model trained on dialogues proved to be the more robust one, possibly due to the fact that dialogues contain more contrastive tempo characteristics, while sharing with university lectures most of the pitch and energy patterns on disfluent sequences. Therefore, a model created with such data generalizes better.

Recently, cross-domain analysis of disfluencies has also been tackled in a holistic view, i.e. discussing how distinct/similar disfluencies are in the discourse markers ecosystem (Cabarrão et al., 2018). The study shows that turn-initial discourse markers are usually easier to classify than disfluencies. Our in-domain experiments achieved an accuracy of about 87% in university lectures and 84% in dialogues. The results for cross-domain are about 11%–12% lower, but still the data from one domain can be used to classify the same events in the other. Ultimately, using exclusively acoustic-prosodic cues, discourse markers can be fairly discriminated from disfluencies and sentence-like units.

In order to better understand the contribution of each feature, we have also reported the impact of the features in both the dialogues and the university lectures. Pitch features are the most relevant ones for the distinction between discourse markers and disfluencies, namely pitch slopes. Furthermore, although they have idiosyncratic properties, disfluencies, particularly filled pauses, may share with discourse markers the same prosodic properties, as the plateau contours contrasting with rises in the following prosodic constituents.

Future challenges will encompass human-computer interactions both with virtual and embodied agents aiming at simulating both the

idiosyncratic traits of the domains and the shared acoustic-prosodic features across such domains.

Acknowledgements

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, and under Ph.D. grant SFRH/BD/96492/2013, and Post-doc grant SFRH/PBD/95849/2013. The author is very grateful to work with her co-authors and friends, Vera Cabarrão, Fernando Batista, Rubén Solera Ureña, Jaime Ferreira, Ana Isabel Mata, Isabel Trancoso. Furthermore, the author is also very grateful for discussions and feedback provided by several researchers, with a special thank you for Julia Hirschberg, Ralph Rose, and Vered Silber-Varod.

References

- Cabarrão, V., H. Moniz, F. Batista, J. Ferreira, I. Trancoso & A. I. Mata. 2018. Cross-domain Analysis of Discourse Markers in European Portuguese. *Dialogue & Discourse* 9(1): 79–106.
- Moniz, H., F. Batista, A. I. Mata & I. Trancoso. 2014. Speaking Style Effects in the Production of Disfluencies. *Speech Communication* 65: 20–35. <https://doi.org/10.1016/j.specom.2014.05.004>
- Moniz, H., A. Pompili, F. Batista, I. Trancoso, A. Abad & C. Amorim. 2015a. Automatic Recognition of Prosodic Patterns in Semantic Verbal Fluency Tests – an Animal Naming Task for Edutainment Applications. In: The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of The 18th International Congress of Phonetic Sciences* (ICPhS 2015), 10–14 August 2015, Glasgow, Scotland, paper number 0997.1-5.
- Moniz, H., J. Ferreira, F. Batista & I. Trancoso. 2015b. Disfluency Detection across Domains. In: *DiSS 2015, Proceedings of the 7th Workshop on Disfluency in Spontaneous Speech*, 8-9 August 2015, University of Edinburgh, Scotland, UK.
- Rosenberg, A. 2010. AuToBI – A Tool for Automatic ToBI Annotation. In: T. Kobayashi, K. Hirose & S. Nakamura (eds.), *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (Interspeech 2010), 26–30 September 2010, Makuhari, Chiba, Japan, 146–149.
- Trancoso, I., M. C. Viana, I. Duarte & G. Matos. 1998. Corpus de Diálogo CORAL. In: *Proceedings of Proceesing of the Portuguese Language* (PROPOR 1998), 3–4 November 1998, Portalegre, Brazil.
- Trancoso, I., R. Martins, H. Moniz, A. I. Mata & M. C. Viana. 2008. The LECTRA Corpus – Classroom Lecture Tanscriptions in European Portuguese. In: Nicoletta Calzolari (Conference Chair), K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis & D. Tapias (eds.), *Proceedings of the Sixth International Conference on Language Resources and Evaluation* (LREC 2008), 28–30 May 2008, Marrakesh, Morocco, 1416–1420.