

Hesitations and Individual Variability in Italian Tourist Guides' Speech

Loredana Schettino¹, Simon Betz², Francesco Cutugno³, Petra Wagner²

¹University of Salerno, Italy; ²Bielefeld University, Germany; ³University of Naples "Federico II", Italy

In this study, we investigate hesitation strategies that tourist guides may use to manage speech with an emphasis on individual variability. It is by now recognized that speech disfluencies are not just occasional and idiosyncratic production errors, but naturally involved in the economy of speech and an integrative part of a language's grammatical organisation (Chafe 1980; Crocco and Savy 2003). Their use allows speakers to manage the online process of speech planning, coding, and articulation, correcting utterances in case of error and taking some extra time to organize the output message (Levelt 1989, Ginzburg et al. 2014). These "stalls" are marked by hesitations like silent pauses, lengthenings, and fillers which carry procedural communicative values, although lacking in propositional content. Their occurrence buys time for the speakers to manage their speech and for listeners to process information, conveying at the same time valuable information on speech planning, structuring, and speakers' disposition (Cataldo et al. 2019, Betz 2020). However, there is no evidence of speakers' deliberate control over their production (see Eklund 2004, Corley and Stewart 2008).

Previous investigations on Italian tourist guides' speech (Cataldo et al. 2019) have pointed out that such "disfluencies" may occur as a tool to structure discourse and gain visitors' attention, and that linguistic idiosyncratic behaviour may affect their production. The individual variability issue was also recently tackled in a study on German semi-spontaneous speech where speaker-specific hesitation strategies emerged (Betz and Lopez Gambino 2016). Moreover, McDougall and Duckworth (2017) highlighted the speaker-discriminating role of disfluency production, which provides a further tool for forensic phoneticians.

Given these findings, the proposed study delves deeper into the linguistic analysis of formal, phonetic, and functional aspects of hesitations occurring in a corpus of tourist guides' speech. We investigated whether different types of hesitations and their phonetic features correlate with different discourse functions and what individual strategies speakers may, more or less consciously, use when hesitating.

We performed a corpus-based analysis on a dataset from the C.H.R.O.M.E. corpus (Origlia et al. 2018). It consists of circa 80' semi-spontaneous speech by three female expert guides leading visits at San Martino's Charterhouse. Disfluency phenomena were annotated using a three-level annotation scheme: Disfluency Model, Disfluency Structure, Disfluency Function (see Shriberg 1994, Eklund 2004, Ginzburg et al. 2014). Phenomena falling into the set of "hesitation pauses" were associated with possible function/s according to their co-text. The types considered were: Silent Pauses (SP); Filled Pauses (FP, such as "ehm" or "eeh"); Lengthenings (LEN, as in "nell<aa> Certosa"); Lexicalized Filled Pauses (LFP, such as "let's say", "so"). Functions were classified as follows: Hesitative, referring to speech planning as hesitation basic function; Word searching, for items revealing difficulties' in retrieving target words; Structuring, when structuring discourse at syntax and information structure level; Focusing, when emphasizing upcoming semantically heavy elements (see Cataldo et al. 2019). The robustness of this categorization was tested measuring inter-rater reliability, Cohen's kappa was 0.732 ("substantial agreement", Landis and Koch 1977).

Results confirm and dig deeper into findings by Cataldo et al. (2019) reporting the emergence of idiosyncratic linguistic behaviours. The 1158 hesitation items occurring in the dataset are unevenly distributed across the three speakers (G01, G02, G03). Indeed, their G01's productions occur about twice as frequently as for the other two guides, see Table 1. Preliminary statistical computation confirms that speakers adopt different strategies in the choice of hesitation phenomena and their pragmatic functions. A Multinomial Logistic Regression Model was fitted in R, defining "type" as the dependent variable and "speaker" and "pragmatic function" as interacting independent variables. Then, a pairwise comparison among fixed levels as to the speaker effect on the hesitation choice showed the following significant results: compared to the other two speakers G01 uses more filled pauses and lengthenings, fewer lexical fillers; G02 fewer silent pauses; G03 fewer lengthenings ($p < .001$). Also, the interaction between hesitation type and function was found to be significant ($p < .001$), namely, as compared to the other guides, in G01's speech more lengthenings are used with focusing function and more filled pauses with structuring function, in G02's speech more lexical fillers are used to convey hesitative and word searching function, in G03's speech more silent pauses are used for hesitative function whereas fillers and lengthenings for word searching function.

To get general correlations between hesitations' formal and functional features besides inter-speakers' variability, mixed models considering "speaker" and "item" as random intercepts were employed – statements about cause-effect are avoided considering the number of variables at play in semi-spontaneous speech. Generalized Linear Mixed Models with hesitations' type as dependent variable and function as independent

variable showed that fillers are generally used with word searching function (SE: 0.19, $p < .001$) and not with focusing function (SE: 0.95, $p < .001$); lengthenings are also less used with focusing (SE: 0.38, $p < .001$), and structuring function (SE: 0.28, $p < .001$); conversely, lexical fillers are generally used to convey structuring (SE: 0.25, $p < .001$), and focusing function (SE: 0.28, $p < .001$). The correlation between hesitations' duration and the type was tested fitting a Linear Mixed Model, with duration as the dependent variable and type as the independent variable. It showed fillers (mean: 0.46 s) to be significantly longer than silent pauses (mean: 0.27 s; SE: 0.03, $p < .001$) and lengthenings (mean: 0.26 s; SE: 0.02, $p < .001$) To test the correlation between hesitations' duration and function, a Linear Mixed Model was built for each type defining the function as the independent variable. All hesitation types were found to be significantly longer when carrying out word searching function (**LEN** means: 0.33 s vs. 0.20 s; SE: 0.02, $p < .0001$. **FP** means: 0.60 s vs. 0.30 s; SE: 0.04 $p < .0001$. **SP** means: 0.37 s vs. 0.27 s; SE: 0.04, $p = .03$). Interestingly enough, lengthenings were found to be significantly shorter when conveying focusing function (means: 0.16 s vs. 0.28 s; SE: 0.02, $p < .0001$).

To conclude, these results show how speakers use hesitation phenomena to manage their discourse and may choose different "hesitation strategies". Indeed, G01 prefers an 'on the fly' production, employing several filled pauses and lengthenings, also with, respectively, structuring and focusing function; G02 avoids silent pauses all together and tends to avoid lengthenings and filled pauses preferring lexicalized filled pauses for word searching and general hesitation functions as well as for structuring and focusing; whereas G03 adopts a more controlled, 'rhetorical' style, using mainly lexicalized filled pauses and silent pauses for emphasizing, structuring information, and for general hesitations, then, very few lengthenings and filled pauses occur when searching for a certain word. Overall, lengthenings and filled pauses were found to correlate with general hesitations and word retrieval problems, rather than for structuring and focusing functions, the latter being generally conveyed by lexicalized filled pauses instead. Noteworthy, the opposition between the word searching and focusing functions was found to be encoded in lengthenings' duration, respectively longer and shorter. Finally, although focusing on a restricted number of speakers, this study further uncovers the role of interspeaker variability when describing hesitations and the way they come into play in speech production and perception also in Italian, which may be a highly influential finding for a range of speech technological applications such as interactive speech synthesis,... Furthermore, it provides a better understanding of hesitation pauses' communicative functions and their contribution to discourse and communication.

Table 1 – Per minute, per word hesitation rate (first two columns) and instances of hesitations by Speaker

Speaker	n° hes/ minute	n° hes/ word	SP	FP	LEN	LFP
G01	20,12	0,16	78 (16%)	166 (34%)	144 (30%)	98 (20%)
G02	11,55	0,07	19 (6%)	42 (13%)	64 (19%)	210 (63%)
G03	12,10	0,08	53 (17%)	32 (10%)	32 (10%)	198 (63%)

References

- Betz, S., Lopez Gambino, M. S. (2016). Are we all disfluent in our own special way and should dialogue systems also be?. *Elektronische Sprachsignalverarbeitung (ESSV)* 2016, 81.
- Betz, S. (2020). Hesitations in Spoken Dialogue Systems. Ph.D. dissertation, Universität Bielefeld, Bielefeld.
- Cataldo, V., Schettino, L., Savy, R., Poggi, I., Origlia, A., Ansani, A., Sessa, I., Chiera, A. (2019). Phonetic and functional features of pauses, and concurrent gestures, in tourist guides' speech. *Studi AISV*.
- Chafe, W. (1980). Some reasons for hesitating. In Dechert, H.W., Raupach, M. (Eds), *Temporal variables in speech: Studies in Honour of Frieda Goldman-Eisler*. The Hague: Mouton, 169-180.
- Corley, M., Stewart, O. W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. *Language and Linguistics Compass*, 2(4), 589-602.
- Crocco, C., Savy, R. (2003). Fenomeni di esitazione e dintorni: una rassegna bibliografica. In Crocco, C., Savy, R. & Cutugno, F. (Eds.), *API. Archivio di Parlato Italiano*, DVD.
- Eklund, R. (2004). Disfluency in Swedish human-human and human-machine travel booking dialogues. Ph.D. dissertation, Linköping University Electronic Press.
- Ginzburg, J., Fernández, R. & Schlangen, D. (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, 7(9), 1-64.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33 1, 159-74.
- Levelt, W.J. (1989). *Speaking: From intention to articulation* (Vol. 1). Cambridge, MA: MIT Press.
- Lickley, R. J. (2015). Fluency and Disfluency. Redford, M.A. (ed.). *The handbook of speech production*, 445-474. John Wiley & Sons.
- McDougall, K. and Duckworth, M. (2017) Profiling fluency: an analysis of individual variation in disfluencies in adult males. *Speech Communication* 95: 16–27.
- Origlia, A., Savy, R., Poggi, I., Cutugno, F., Alfano, I., D'Errico, F., Vincze, L. & Cataldo, V. (2018). An Audiovisual Corpus of Guided Tours in Cultural Sites: Data Collection Protocols in the CHROME Project. In *Proceedings of the 2018 AVI-CH Workshop on Advanced Visual Interfaces for Cultural Heritage*, Castiglione della Pescaia, Italy, 29 May 2018, 1-4.
- Shriberg, E. (1994). Preliminaries to a theory of speech disfluencies. Ph.D. dissertation, University of California, Berkeley.