

LOREDANA SCETTINO, SIMON BETZ, FRANCESCO CUTUGNO,
 PETRA WAGNER

Hesitations and individual variability in Italian tourist guides' speech¹

This study concerns hesitation strategies that tourist guides may use to manage their speech, with particular attention to individual variability. Previous work has pointed out that hesitation phenomena may occur as a tool to structure discourse and gain visitors' attention, and that linguistic idiosyncratic behavior may affect their production. Given these findings, the proposed investigation delves deeper into the linguistic analysis of formal, phonetic, and functional aspects of hesitations occurring in a small corpus of Italian tourist guides' speech. It aims at describing the speaker-specific and common uses of hesitation phenomena and whether different types of hesitations and their phonetic features correlate with different discourse functions. From the results, it emerges a formal differentiation between hesitations involved in speech planning for lexical coding and for the structuring of information.

Keywords: speech, disfluencies, hesitations, pauses, individual variability.

1. Introduction

Human spontaneous speech is characterized by the occurrence of phenomena connected to the online production process. Speakers can produce phonetic cues, like silent pauses or vocalizations, to suspend the speech delivery and gain extra-time for speech planning, or desert utterances due to change of projects, or alter uttered sequences that are considered somehow inaccurate or unfitting for the communicative intention. These phenomena affect to a certain extent speech fluency. For this reason, they are commonly referred to as 'disfluencies'.

The definition of what is fluent and disfluent speech is rather problematic as the issue has been tackled from different angles, different research fields, with specific perspectives, approaches, and aims: investigations on speech disorders compare pathological and non-pathological speech; investigations on foreign or second-language learning compare native and non-native speech; psycholinguistics consider disfluencies as a window into the processes involved in speech production; linguists and phoneticians are interested in describing and framing disfluencies in relation to the linguistic systems; computational linguists are engaged in the definition of disfluency models for technological applications (see Lickley, 2015, for an overview).

¹ Authors' responsibilities – Loredana Schettino: conceptualization, data curation, writing – original draft (§1 – §7); Simon Betz: theoretical and methodological advice, writing – review & editing; Francesco Cutugno: supervision, writing – review & editing; Petra Wagner: supervision, writing – review & editing.

Despite field-specific interests and differences, numerous empirical studies have shown that speech disfluencies are not just occasional and idiosyncratic production errors, but regularly occur in speech (Shriberg, 1994; Bortfeld, Leon, Bloom, Schober & Brennan, 2001) and are naturally involved in the economy of speech as a flexible and efficient tool at speakers' hand to manage the online processes of speech planning, coding, and articulation (Levelt, 1989; Allwood, Nivre & Ahlsén, 1990; Crocco, Savy, 2003; Voghera, 2017). In fact, speakers can correct already uttered sequences using *Repairs* or *Backward-Looking Disfluencies*, that is the deletion, insertion, or substitution of speech material. Furthermore, speakers can gain some extra time to organize the output message inserting pauses, fillers, lengthenings, that are commonly subsumed under the category of *Hesitations* or *Forward-Looking Disfluencies* (Ginzburg, Fernández & Schlangen, 2014).

The occurrences and realizations of disfluencies were found to vary and were associated with different functions in discourse due to contextual and individual factors. (Bortfeld et al., 2001; Betz, Lopez Gambino, 2016, a.o.). Further, it has been highly debated whether and when disfluencies may be intentionally produced to signal something or exert a certain effect. As a matter of fact, up to our knowledge, there is still no evidence of speakers' deliberate control over their production (Corley, Stewart, 2008).

This study aims at investigating hesitation strategies that tourist guides may more or less consciously use to manage their speech, with particular attention to individual variability.

The article is structured as follows: §2 presents the literature concerning hesitation phenomena and the factors affecting their productions; §3 presents the general and specific aims of the study; §4 describes the linguistic data, the annotation process, and the parameters of the linguistic analysis; in §5 and §6 the results of the analysis are reported and discussed.

2. *Related Work*

The term “hesitations” is commonly used in literature to cover a set of speech phenomena, like silent pauses, fillers, lengthenings, that realize a temporary suspension in the utterance delivery (Lickley, 2015). In the last 40 years, after the seminal work by Chafe (1980), research on hesitations has embraced a “positive” view that acknowledges the role played by these elements in speech, that is to reduce the temporal pressure due to the simultaneity of planning, production, and reception processes. On the one hand, they gain valuable time for speakers to manage the online process of speech production. On the other hand, they provide extra time for the listeners to process information (Chafe, 1980; Levelt, 1989; Corley, Hartsuiker, 2003).

Empiric investigations, including corpus-based and experimental studies, have analyzed the meaning and the factors that affect the production of hesitations (see Rochester, 1973; Crocco et al., 2003; Eklund, 2004 for a review). To this end, the

consideration of the context of occurrence is crucial given that these phenomena lack conceptual meaning² (Crible, 2018).

Generally, the occurrence of hesitation phenomena was observed to be mainly connected to moments that require a higher cognitive load in the speech planning process: a) the conceptualization and formulation of a new utterance (macro-planning, Levelt, 1989; Oviatt, 1995; Bortfeld et al., 2001); b) the search for a specific lexical item (micro-planning, Schnadt, Corley, 2006; Hartsuiker, Notebaert, 2009), for a word with low contextual probability (Beattie, Butterworth, 1979) or a word that is more difficult to retrieve for the speaker, hence low in “codability” (Chafe, 1980); c) the selection of new information in discourse (Arnold, Fagnano & Tanenhaus, 2003). In fact, it was also highlighted that hesitation phenomena play a role in structuring discourse and may indicate to the listener the arrival of important information (Chafe, 1980; Allwood et al., 1990; Kosmala, Morgenstern, 2017). Furthermore, the communicative situation, e.g., the linguistic register and the argument under discussion were found to affect the incidence of hesitations (Bortfeld et al., 2001; Moniz, Batista, Mata & Trancoso, 2014; Crible, Bunt, 2016).

Hence, if lacking in conceptual content, hesitations carry procedural meaning conveying valuable information on speech planning, structuring, and speakers’ disposition (Cataldo, Schettino, Savy, Poggi, Origlia, Ansani, Sessa & Chiera, 2019; Betz, 2020). Further studies have shown that the phonetic-prosodic features of these elements correlate with the contextual function and can discriminate between more or less “felicitous” occurrences in discourse (Swerts, 1998; Moniz, Trancoso & Mata, 2010).

More specifically, silent pauses are, with filled pauses, one of the most studied types of hesitation phenomena. After the highly influential analysis by Sacks and colleagues (Sacks, Schegloff & Jefferson, 1978), numerous studies have acknowledged and investigated the communicative role of silence in conversation. It was found to range from emphasizing key elements (Duez, 1997; Strangert, 2003), marking discourse boundaries (Boomer, Dittman, 1962; Esposito, Stejskal, Smékal & Bourbakis, 2007), working as discourse marker for turn-taking management (Ephratt, 2008), to manifesting hesitation in speech or troubles in information processing and need for clarification. In particular, longer silences have been associated with speech processing problems (Chowdhury, Stepanov, Danieli & Riccardi, 2017; Schettino, Di Maro & Cutugno, 2020). In some studies, silences were identified as hesitant when longer than an established temporal threshold (200 ms in Beattie et al., 1979; according to the syntactic position, 200 or 500 ms in O’Shaughnessy, 1992; 60 ms in Kendall, 2009, a.o.). However, this approach is rather problematic as in some contexts silences that are shorter than these thresholds may be perceived as hesitant (Eklund, 2004; Lickley, 2015).

Numerous investigations also involved filled pauses and the correlation between their phonetic-prosodic features (segmental content, pitch values, position, length)

² As proposed in Relevance Theory (Blakemore 2002; Wilson 2011; Wilson, Sperber, 2012), different types of encoded meaning may be identified, i.e., “conceptual meaning”, which refers to entities or concepts, and “procedural meaning”, which gives the coordinates for the interpretation process.

and the associated function in context. More specifically, a longer duration is reported for filled pauses that occur at the beginning of a phrase (Swerts, 1998), that introduce a new part of the utterance (Horváth, 2010), or manifest problems in the retrieval of a specific word (Cataldo et al., 2019). The controversial claim that filled pauses may be considered as proper words intentionally used by speakers to signal upcoming delays in speech delivery was strongly debated (Clark, Fox Tree, 2002). As a matter of fact, it was highlighted that there is still no evidence of speakers' deliberate control over their production (O'Connell, Kowal, 2005; Corley et al., 2008; Finlayson, Corley, 2012). So, it was argued that filled pauses, much like intonation, may well be unconsciously used by speakers in aiming for a certain effect (Kjellmer, 2003) and "acquire their communicative function as pragmatic markers when interpreted by the listener" (Tottie, 2016: 100). Generally, speech planning can be considered the basic function of filled pauses, furthermore, their use was found to be connected to the search for a specific word, discourse structuring, turn-taking management, drawing the listener's attention on an upcoming semantically heavy element, or marking that a delicate or "dispreferred" action is about to follow (Kjellmer, 2003; Schegloff, 2010; Kosmala, Morgenstern, 2018; Tottie, 2020).

As for lexical fillers, they are commonly subsumed under the multi-functional class of discourse markers and not always acknowledged among hesitation phenomena (Schiffrin, 1987; Bazzanella, 2006; Crible, 2018). However, as in filled pauses, the primary function of discourse markers used as fillers is gaining extra time for speech planning. Then, due to their multifunctional nature, they may also provide valuable meta-discursive information about turn-taking management, discourse structure, and the speaker's disposition (Bazzanella, 2006; Schettino, Cataldo, 2019).

Finally, among the voiced hesitant pauses, the lengthenings of segmental material can be identified. Pitch values were found to distinguish hesitant prolongations from non-disfluent ones, i.e., due to accentuation. In particular, the first ones are generally realized with lower pitch range and flat contour, whereas the second ones with higher pitch range and rising contour (Eklund, 2001; Moniz, 2013; Betz, 2020). However, the study of the way lengthenings work in discourse and speech planning has only in the past about ten years gained more attention (Moniz, Mata & Céu Viana, 2007; Betz, Zarriß, Székely & Wagner, 2019).

In different languages filled pauses were found to be on average longer than lengthenings (Swedish: Eklund, 2001, 2004; German: Betz, Eklund & Wagner, 2017; Italian: Di Napoli, 2020; Cataldo et al., 2019).

Besides these general observations, studies on a range of languages and speaking styles report significant individual variability in the production of disfluencies (Shriberg, 2001; Eklund 2004; Roberts, Meltzer & Wilding, 2009; Moniz et al., 2014; Kosmala et al., 2018; Cataldo et al., 2019). Specific investigations confirm this finding by describing the emergence of speaker-dependent pausing strategies, i.e., in the choice of using silences, lengthenings, or filled pauses for the need of speech planning (Van Donzel, Koopmans-van Beinum, 1996). In particular, Betz and Lopez Gambino (2016) analyzed the individual variability in a dataset of German spontaneous speech

produced by six speakers and elicited via a Wizard of Oz experiment to control for time management. They report three main tendencies: a general avoidance of fillers and a high tolerance for silence; a balanced use of both silent and filled pauses; avoidance of both and reliance on low-content words for hesitation purposes. Furthermore, it was observed that the individual hesitation strategies characterize speakers' production also when speaking a second language (Fehring, Fry, 2007). So, based on the assumption that speech planning and production strategies may respond to individual psycho- and socio-linguistic demands, McDougall and Duckworth (2017) highlighted the speaker-discriminating role of disfluency production, which provides a further tool for forensic phoneticians.

3. *Research Aim*

This study is part of a project that aims at modeling the occurrence of hesitation phenomena in Italian semi-spontaneous discourse for technological applications. It has been developed from the CHROME project – *Cultural Heritage Resources Orienting Multimodal Experience* – generally aimed at modeling multimodal data for the design of Virtual Agents serving in museums (Origlia, Savy, Poggi, Cutugno, Alfano, D'Errico, Vincze & Cataldo, 2018).

The observations reported in the previous section §2 support the assumption that hesitation phenomena contribute to the efficacy of the interaction. So, modeling their occurrence in discourse may be useful for both improving our understanding of conversational dynamics and the implementation of dynamic and efficient human-machine interactions. More specifically, Incremental Spoken Dialogue Systems would benefit from the ability to insert hesitation phenomena as it would provide a set of tools to cover the time needed for response processing, to signal the ongoing process, and it would result in more natural productions (Adell, Escudero & Bonafonte, 2012; Skantze, Hjalmarsson, 2013; Betz, Carlmeyer, Wagner & Wrede, 2018; Origlia, Savy, Cataldo, Schettino, Ansani, Sessa, Chiera & Poggi, 2019).

Hence, the proposed study delves deeper into the linguistic analysis of structural, phonetic, and functional aspects of hesitations in a corpus of tourist guides' speech and addresses the following research questions:

1. What individual strategies can speakers use when hesitating?
2. How do different types of hesitations and their duration correlate with different discourse functions?

4. *Methodology*

4.1 Corpus and dataset

To address the research questions, we conducted a corpus-based analysis on a dataset extracted from the CHROME corpus (Origlia et al., 2018). The corpus consists of audio-visual recordings of three tourist guides (G) each leading four visits (V) at

San Martino's Charterhouse in Naples. Each visit is organized into six points of interest (P). The selected data concern the presentation of the same point of interest (P01, i.e., the *Pronao*, the entrance of the Church) in two different visits per guide, which amounts to approximately 80 minutes of semi-spontaneous speech (Table 1 reports duration, number of uttered words, and speech rate per speaker).

Given the setting and the communicative goal of tourist guides, the elicited speech presents specific characteristics: a high degree of discourse planning, as the production is based on a partially pre-structured descriptive text; a low degree of interaction and participation in the discourse construction, because of the asymmetrical relationship between the guide and the audience. Hence, this kind of speech can be defined as semi-spontaneous and semi-monological (Voghera, 2017; Cataldo et al., 2019).

Table 1 - *Selected Dataset*

Speaker	File - ID	Duration (min)	Words	words/minute
G01	V01P01	11:33	1491	129.0
G01	V02P01	13:22	1705	129.2
G02	V01P01	16:00	2788	174.3
G02	V02P01	13:32	2349	173.6
G03	V01P01	14:29	2138	147.6
G03	V03P01	11:48	1706	145.1
<i>Total</i>		80:34	12183	151.1

4.2 Linguistic Analysis

The scheme adopted for the annotation of disfluency phenomena is a revised version of the one tested in a previous pilot study (Cataldo et al., 2019). The system integrates the identification of disfluencies' formal structures and their functions in the context of occurrence (Schettino, Cataldo, Alfano & Leo, forthcoming) and includes the following levels of annotation:

1. *Disfluency Model*, on the first level, the macro-structure of disfluencies is labeled. Namely, the region to be repaired, the repaired one, and the one in which the delay occurs (see Shriberg, 1994).
2. *Disfluency Structure*, the second level serves for labeling the micro-structure embodying the disfluency. Here, disfluent items are categorized as *Insertion*, *Deletion*, *Substitution*, *Repetition*, *Silent Pause*, *Lengthening*, *Filled Pause*, *Lexicalized Filled Pause* (Eklund, 2004).
3. *Disfluency Function*, on the third level, each item is assigned its macro-function, *Backward-Looking* or *Forward-Looking* (Ginzburg et al., 2014).
4. *Hesitation Function*, on a fourth level, Forward-Looking items – hesitations marking a temporary delay in speech – are associated with more specific functions regarding their co-text.

Unlike the categories of the first three levels, on the fourth functional level, given the fact that hesitation phenomena may carry out more than one function, categories are not mutually exclusive.

The procedure was conducted using the work tool ELAN (2020, Sloetjes, Wittenburg, 2008), a software for multilevel linguistic annotations, which allows to assign labels from pre-specified vocabularies and define the relationship between categories per each tier.

In this study the following hesitation types are considered:

- *Silent Pause* (SP) identifies marked silences perceived as a hesitant pause in the context of occurrence (breath pauses are not included);
- *Filled Pause* (FP) identifies non-verbal filler, vocalizations and/or nasalizations, i.e., *eeh*, *ehm*, *mbh*;
- *Lexicalized Filled Pause* (LFP) identifies verbal fillers, meaning lexical items that are semantically strongly bleached in their context of occurrence, i.e., *well*, *let's say*, *so*, a.o.;
- *Lengthening* (LEN) identifies marked prolongation of segmental material (Betz, 2020).

The identification of these phenomena does not depend on absolute measures but is related to the context of occurrence, which entails subjective perceptual judgment. Hence, its reliability was tested measuring Cohen's K for the ratings by two expert annotators (K = 0.91, high agreement; Landis, Koch, 1977).

Each type of hesitation is then associated with specific functions according to their context of occurrence. Functions were classified as follows:

- *Word Searching* (WS) is the label assigned to items involved in the lexical retrieval and selection of a target word (Tottie, 2020), for example to the filled pause, silent pause, lengthening, and repetitions preceding the word “beauty” in the following utterance:

(a) *potete intuire <ehm> <sp> la<aa> la bellezza*
 “you can grasp <uhm> <sp> the<ee> the beauty”

- *Structuring* (STR) is the label assigned to items that play a structuring role in discourse occurring on syntax (clause) and information structure (topic-comment) boundaries. The following utterance provides an example of a silent pause and a filled pause co-occurring on the boundary between two clauses, in both cases a lengthening and a filled pause occur between the topic and the relative comment.

(b) *“la<aa> Certosa di San Martino qui a Napoli<ii> <ehm> ha almeno due anime <sp> <eeh> una<aa> <ehm> racconta la storia...”*
 “the<ee> Charterhouse here in Naples<vv> <ehm> has two souls <sp> <eeh> one<ee> <ehm> tells the story...”

- *Focusing* (FOC) is the label assigned to items preceding keywords or key concepts, i.e., semantically heavy and often emphasized elements in the context (Kjellmer,

2003). The following utterance provides an example of a silent pause preceding the main information in the utterance: the origin of the Charterhouse.

- (c) *“quindi la<aa> Certosa ha un’origine <sp> trecentesca”*
 “so the<ee> Charterhouse has a <sp> 14th century origin”

– *Hesitative* (HES) is the category standing for hesitations’ basic function of speech planning. It is assigned to items for which no other function can be identified.

- (d) *“non possiamo vedere<ee> molto bene”*
 “we can’t see<ee> it properly”

Also in this case the robustness of the categorization was tested measuring Cohen’s K for the inter-annotator agreement ($K = 0.73$, substantial agreement; Landis et al., 1977).

Furthermore, the analysis concerns the duration variation of SP, FP, and LEN as a function of hesitations type and function.

4.3 Statistical Analysis

The statistical analysis was performed using the R software (R Core Team, 2020). To define speakers’ systematic preferences and strategies, the occurrence of the selected hesitation types was modeled as a function of the associated function and the speaker. As reported in the previous paragraph, §4.2, four types of hesitation phenomena are considered, hence a Multinomial Logistic Regression Model was built using the ‘multinom’ function from the ‘nnet’ package (Venables, Ripley, 2002). The model included Hesitation Type as multinomial dependent variable, whereas Speaker and Function as interacting independent variables. This approach is useful to consider the speaker’s effect on the occurrence of hesitations and to observe their interaction with the effect of function, but to inspect the way hesitations are commonly used, a method that allows controlling for individual variability is needed.

Therefore, to analyze the role of functions on the occurrence of the four different types of phenomena independently from speaker-specific strategies, Mixed Models were implemented selecting Speaker and Item as random effects (‘lme4’ package, Bates, Maechler, Bolker & Walker, 2015). More specifically, to determine whether and how the use of hesitation types depends on the functions, Generalized Linear Mixed Models were fitted. In this case, each type of hesitation was processed as binomial dependent variable and Function was selected as the independent variable. Furthermore, Linear Mixed Models were fitted to inspect the systematic variation of hesitations duration, introduced as the dependent variable, as a function of Type and Function, selected as interacting independent variables.

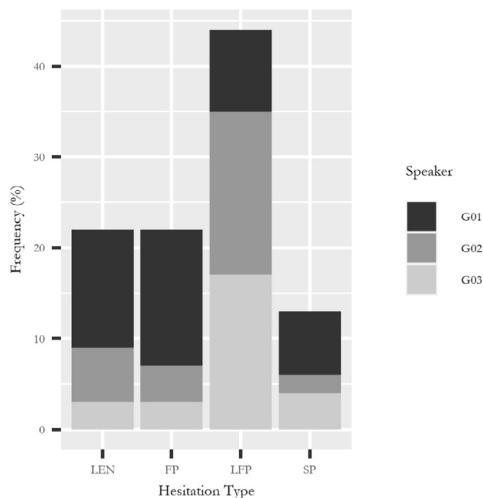
Post-hoc analyses were conducted to inspect the levels within the main effects and the interactions using pairwise comparisons (‘emmeans’ package, Length, 2020). P-values were calculated using Tukey’s HSD adjustment.

5. Results

5.1 Idiosyncratic Hesitation Strategies

In the selected dataset 1158 hesitation items occur. Among these, as illustrated in Figure 1, lexical fillers are by far the most frequent (45%), filled pauses and lengthenings are equally frequent (21%), whereas silent pauses are less frequent (13%).

Figure 1 - Frequency of hesitation types per speaker



Results confirm and dig deeper into some of the findings by Cataldo et al. (2019) reporting the emergence of idiosyncratic linguistic behaviors. As shown in Table 2, the occurrences are unevenly distributed across the three speakers (G01, G02, G03). Indeed, G01 produces approximately 20 hesitations per minute and 16 per 100 words, which is about twice as much as the productions by the other two guides. The statistical computation yielded significant results as for the speaker effect and the interaction between speaker and function, which confirms that speakers systematically adopt different strategies in the choice of hesitation phenomena and associated pragmatic functions (see Fig. 1 and Fig. 2).

Table 2 - Per minute, per word hesitation rate (column 2 and 3) and occurrences of hesitations by Speaker

Speaker	n° hes /minute	n° hes /100 word	LEN	FP	LFP	SP
G01	20.12	16	144 (30%)	166 (34%)	98 (20%)	78 (16%)
G02	11.55	7	64 (19%)	42 (13%)	210 (63%)	19 (6%)
G03	12.10	8	32 (10%)	32 (10%)	198 (63%)	53 (17%)

Table 3 - Significant results yielded by the post-hoc analysis for the multinomial logistic models

HesType	Function	Speakers contrast	Estimate	SE	df	t.ratio	p.value
LEN	/	G01 – G03	0.113	0.035	36	3.255	0.007
FP	/	G01 – G02	0.172	0.028	36	6.145	<.0001
FP	/	G01 – G03	0.103	0.039	36	2.641	<.0001
LFP	/	G01 – G02	-0.330	0.029	36	-11.328	<.0001
LFP	/	G01 – G03	-0.129	0.030	36	-4.302	<.0001
SP	/	G01 – G02	0.113	0.018	36	6.156	<.0001
SP	/	G02 – G03	-0.199	0.036	36	-5.416	<.0001
LEN	FOC	G01 – G02	0.100	0.034	36	2.909	0.0166
LEN	FOC	G01 – G03	0.088	0.033	36	2.632	0.0325
FP	STR	G01 – G02	0.371	0.053	36	6.956	<.0001
FP	STR	G01 – G03	0.318	0.058	36	4.649	0.0001
LFP	HES	G01 – G02	-0.334	0.066	36	-5.020	<.0001
LFP	HES	G02 – G03	0.350	0.080	36	4.347	0.0003
LFP	WS	G01 – G02	-0.120	0.053	36	-2.283	0.0711
LFP	WS	G02 – G03	0.180	0.053	36	3.376	0.0049
SP	HES	G01 – G03	-0.413	0.116	36	-3.543	0.0031
SP	HES	G02 – G03	-0.457	0.115	36	-3.956	0.0010

The *post-hoc* analysis reveals that as compared to the other two speakers: G01 uses significantly more filled pauses (34%) and lengthenings (30%), fewer lexical fillers (20%); G02 uses fewer silent pauses (6%). Also, concerning the interaction between hesitation type and function, as compared to the other guides' speech, in G01's speech, more lengthenings are used with focusing function (FOC), and more filled pauses with structuring function (STR); in G02's speech, more lexical fillers are used to convey hesitative (HES) and word searching (WS) function; in G03's speech, more silent pauses are used for the hesitative function (significant results are reported in Table 3).

5.2 Hesitation Type and Functions

Besides the observed inter-speakers variability, the mixed models used to investigate the general systematic relationship between hesitation types and the function associated in context yielded significant results (Fig. 2).

Figure 2 - Frequency of hesitation functions per speaker grouped by hesitation type

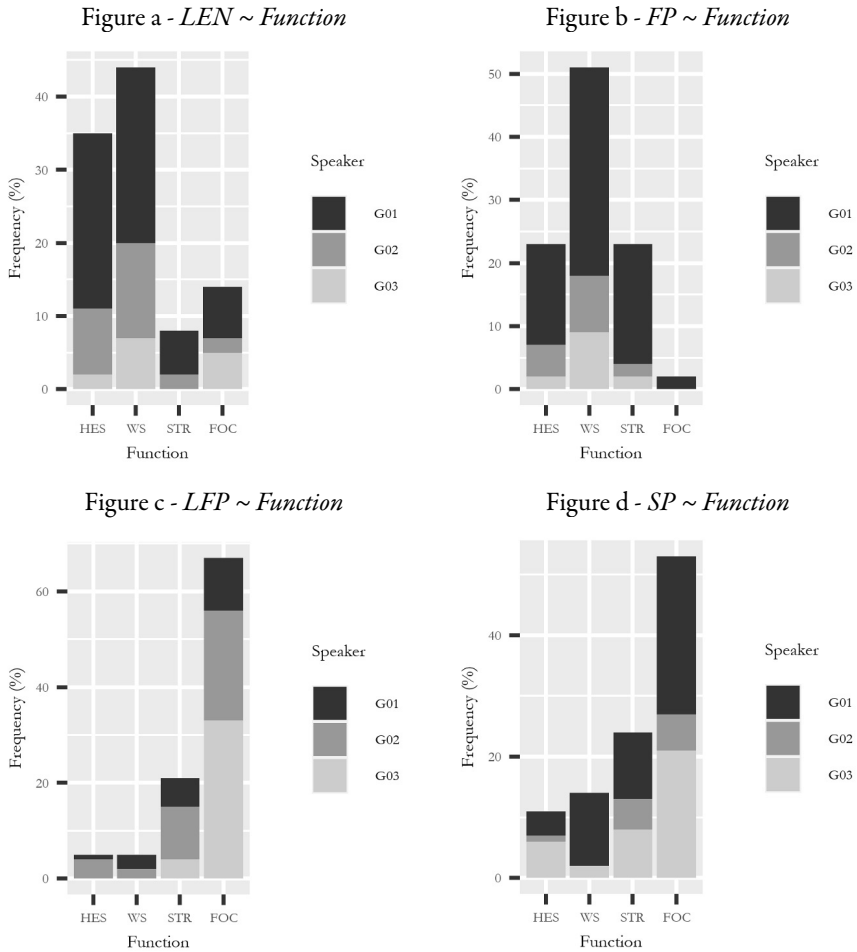


Figure 2a shows that lengthenings are more frequently associated to hesitant and word searching functions and significantly less to structuring (Est: -2.17; SE: 0.28; z: -7.75; $p < .001$) and to focusing (Est: -2.33; SE: 0.40; z: -5.80; $p < .001$). Similarly, Figure 2b shows that fillers generally carry out word searching functions (Est: 0.62; SE: 0.19; z: 3.18; $p < .001$) and significantly less often focusing functions (Est: -4.10; SE: 0.95; z: -4.30; $p < .001$).

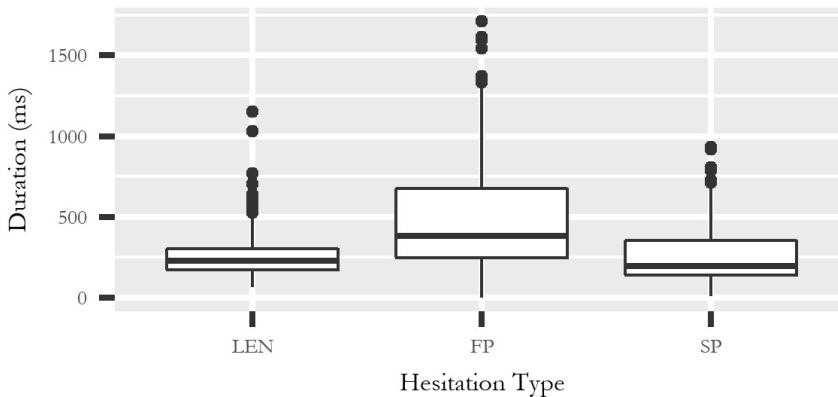
Conversely, as illustrated in Figure 2c, lexical fillers are more frequently associated to structuring (Est: 1.76; SE: 0.24; z: 7.10; $p < .001$) and focusing functions (Est: 2.78; SE: 0.27; z: 9.97; $p < .001$).

Lastly, a similar trend is reported for silent pauses in Figure 2d. In this case, silent pauses are systematically less associated to the hesitant function (Est: -2.34; SE: 0.49; z: -4.76; $p < .001$).

5.3 Hesitations, Duration and Functions

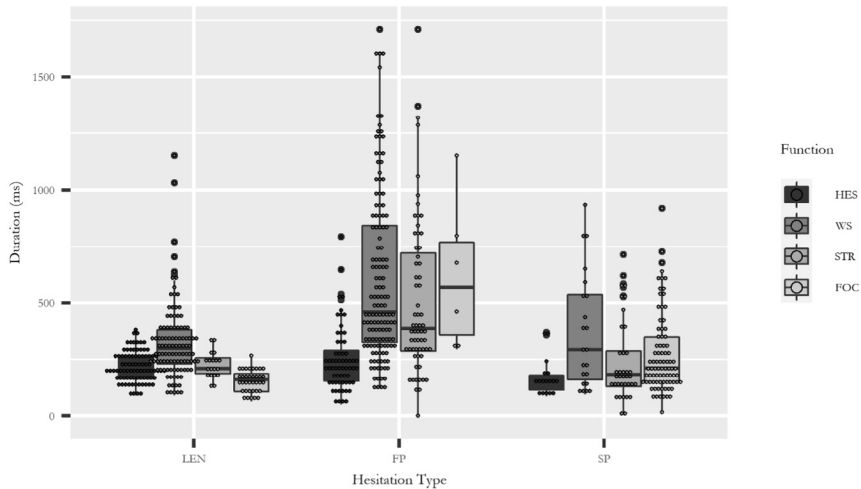
Next, we investigated the variation of the duration of hesitations as a function of their type and associated function. On average, filled pauses (mean: 409 ms) are significantly longer than silent pauses (mean: 243 ms; Est: 165.7; SE: 26.6; $t: 6.22$; $p < .001$) and lengthenings (mean: 191 ms; Est: 217.8; SE: 24.6; $t: 8.84$; $p < .001$).

Figure 3 - *Duration of hesitations per type*



As for function, all hesitation types were found to be significantly longer when associated to the word searching function. More specifically, WS lengthenings show a mean length of 280 ms vs. 184 ms (Est: 96.6; SE: 28.6; $t: 3.38$; $p = 0.03$), WS filled pauses are on average 513 ms vs. 187 ms (Est: 326; SE: 31.3; $t: 10.42$; $p < .00001$), and WS silent pauses are on average 351 ms long vs. 256 ms (Est: 94.8; SE: 43.1; $t: 2.201$; $p = 0.03$).

Interestingly enough, lengthenings were found to be significantly shorter when co-occurring with focusing function (148 ms vs. 246 ms; Est: -98.3; SE: 16.1; $t: -6.102$; $p < .0001$).

Figure 4 - *Duration of hesitations per type and function*

6. Discussion

The interpretation of the presented results must consider the specific features of the analyzed speech. As previously mentioned, tourist guides' speech is mostly characterized by a high degree of speech planning and a descriptive-informative style that results from the type of asymmetrical relationship between the interlocutors: the guide and the tourists. In this setting, the speaker aims to engage the listeners and provide the required information clearly and efficiently. So, the consistent occurrence of hesitation phenomena in these pre-scripted, repeated productions corroborates the assumption of hesitations' supportive role in achieving the communicative purpose in the desired manner rather than obstructing it. For this reason, studying hesitation phenomena in this speech style, i.e., the informative style, proves to be particularly helpful to get insight of how they work as linguistic resources.

The analysis confirms that speakers choose different specific hesitation strategies (Cataldo et al., 2019), which, as observed by McDougall and Duckworth (2017), may reflect individual responses to psycho- and socio-linguistic demands. Indeed, the first guide manifests a higher tolerance for hesitations producing twice as many as the other two guides, particularly, several filled pauses and lengthenings. In her production, more than in the others', filled pauses are also used to mark sentence and topic boundaries and lengthenings to mark semantically heavy keywords, besides buying time for the online speech processing, in what could be defined as a more spontaneous 'on the fly' production. The second guide produces fewer hesitations, more specifically, she avoids hesitant silent pauses altogether and tends to avoid lengthenings and filled pauses relying on lexicalized filled pauses for word searching and general hesitation functions as well as for structuring and focusing. Then, the third guide adopts a more controlled, 'rhetorical' style, using mainly lexicalized

filled pauses and silent pauses to structure her discourse and to mark important information, besides general hesitation. The very few lengthenings and filled pauses she produces are associated with casual word retrieval difficulties.

Overall, in the analyzed speech, hesitation is mostly marked by lexicalized filled pauses, which means that the three tourist guides prefer to cover the time needed for speech planning producing fillers that consist of a proper lexical form (though semantically bleached and not adding anything to the propositional content of the utterance; Bazzanella, 2006; Crible, 2018) rather than other hesitation pauses such as vocalizations and silences that may be perceived as being more salient and disruptive (Betz, Bryhadyr, Kosmala & Schettino, 2021).

Despite interspeaker variability, a structural and phonetic differentiation emerges between hesitations associated with different functions.

More specifically, lengthenings and filled pauses are mostly involved in word retrieval problems and general planning problems, rather than structuring discourse. Conversely, lexicalized filled pauses and silent pauses are mostly associated with the structuring and focusing functions.

Looking at the results concerning duration values, across hesitation types, phenomena associated with the word searching function suspend speech delivery for more time, which leads to acknowledging the search for and the selection of a specific lexical item as more time-consuming processes.

Furthermore, the opposition between the word searching and focusing functions was found to be encoded in lengthenings' duration, respectively longer and shorter.

Given these results, the higher duration registered for filled pauses as compared to silent pauses and lengthenings (also in Eklund, 2001; 2004; Betz et al., 2017; Di Napoli, 2020; Cataldo et al., 2019) may be explained by the fact that filled pauses are the type of hesitation more frequently associated to the word searching function.

The observed structural and phonetic differentiation of hesitation phenomena could be interpreted as a reflection of different processes involved in speech planning. Two main processes could be identified, namely, on the one end the conceptualization and translation of thought into linguistic objects (lexical-semantic level), on the other the structurization of information (pragmatic level). Indeed, as already observed by Levelt (1983, 1989) and Chafe (1980) speakers may need time to check for what to say, how to code their thought into language, and make sure that their communicative intention is conveyed appropriately and efficiently.

The described analysis has contributed to shedding light on the way speakers may use hesitation phenomena as resources to efficiently manage their speech while aiming at reaching their communicative goal, in this case, providing the required information clearly and efficiently. Although to date it is not clear the degree of consciousness involved in this process, it emerged that speakers may enact different hesitation strategies by employing these phenomena in the way that they deem as most suitable and appropriate to the task. This choice may further be subjected to intra-speaker variation: synchronically – due to the specific communicative situation, the roles of the interlocutors, the speakers' degree of self-consciousness

as they speak – and diachronically – due to the over time ongoing building of the individual cognitive representations that underlie language use (respectively referred to as “intra-speaker” and “intra-individual” variation in Bülow, Pfenninger, 2021).

7. Conclusion

To conclude, this study has contributed to describing the use of hesitations as speech management phenomena in Italian tourist guides’ speech considering the relevance of inter-speaker variability.

The identification and modeling of this kind of linguistic behavior provide influential findings for a range of technological speech applications such as interactive speech synthesis. In particular, the performances, i.e., the naturalness and communicative efficacy, of a system supporting a Virtual Agent serving in museums may result improved by a theoretically motivated insertion of hesitations. Furthermore, the observation of speaker-specific strategies leads us to be cautious towards developing synthesis systems based on averaged behaviors. Follow-up studies are meant to experimentally evaluate listeners’ perception to test relevant corpus-based observations.

Finally, this investigation has concerned a specific type of speech and the production of a limited number of speakers, which works for the intended specific application. However, future analyses should consider further speaking styles gathered from a larger number of speakers to gain a better understanding of hesitation pauses’ communicative functions and their contribution to discourse and communication.

Acknowledgements

Work funded by the Italian National Project PRIN Cultural Heritage Resources Orienting Multimodal Experiences (CHROME)(B52F15000450001).

Bibliography

- ADELL, J., ESCUDERO, D. & BONAFONTE, A. (2012). Production of filled pauses in concatenative speech synthesis based on the underlying fluent sentence. In *Speech Communication*, 54 (3), 459-476. <https://doi.org/10.1016/j.specom.2011.10.010>
- ALLWOOD, J., NIVRE, J. & AHLSEN, E. (1990). Speech management – on the non-written life of speech. In *Nordic Journal of Linguistics*, 13, 3-48. <https://doi.org/10.1017/S0332586500002092>
- ARNOLD, J.E., FAGNANO, M. & TANENHAUS, M.K. (2003). Disfluencies signal thee, um, new information. In *Journal of psycholinguistic research*, 32 (1), 25-36. <https://doi.org/10.1023/A:1021980931292>

- BATES, D., MAECHLER, M., BOLKER, B. & WALKER, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. In *Journal of Statistical Software*, 67 (1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- BAZZANELLA, C. (2006). Discourse markers in Italian: towards a 'compositional' meaning. In *Approaches to discourse particles*, 1, 504-524.
- BEATTIE, G.W., BUTTERWORTH, B.L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. In *Language and speech*, 22 (3), 201-211. <https://doi.org/10.1177/002383097902200301>
- BETZ, S. (2020). Hesitations in Spoken Dialogue Systems. PhD Dissertation, Universität Bielefeld.
- BETZ, S., LOPEZ GAMBINO, M.S. (2016). Are we all disfluent in our own special way and should dialogue systems also be?. In *Proceedings of Elektronische Sprachsignalverarbeitung (ESSV)*, Leipzig, Germany, 2-4 March 2016, 81.
- BETZ, S., EKLUND, R. & WAGNER, P. (2017). Prolongation in German. In *Proceedings of DiSS 2017 The 8th Workshop on Disfluency in Spontaneous Speech*, Stockholm, Sweden, 18-19 August 2017, 13-16.
- BETZ, S., CARLMEYER, B., WAGNER, P. & WREDE, B. (2018). Interactive hesitation synthesis: modelling and evaluation. In *Multimodal Technologies and Interaction*, 2 (1), 9. <https://doi.org/10.3390/mti2010009>
- BETZ, S., ZARRIESS, S., SZÉKELY, É. & WAGNER, P. (2019). The green tree-lengthening position influences uncertainty perception. In *20th Annual Conference of the International Speech Communication Association: Crossroads of Speech and Language, INTERSPEECH 2019*. Graz, Austria, 15-19 September, 3990-3994. <https://doi.org/10.21437/Interspeech.2019-2572>
- BETZ, S., BRYHADYR, N., KOSMALA, L. & SCHETTINO, L. (2021). A Crosslinguistic Study on the Interplay of Fillers and Silences. In *Proceedings of DiSS 2021, The 10th Workshop on Disfluency in Spontaneous Speech*, Paris, France, 25-27 August 2021, 47-52.
- BLAKEMORE, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers* (Vol. 99). Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486456>
- BOOMER, D.S., DITTMANN, A.T. (1962). Hesitation pauses and juncture pauses in speech. In *Language and speech*, 5 (4), 215-220. <https://doi.org/10.1177/002383096200500404>
- BORTFELD, H., LEON, S.D., BLOOM, J.E., SCHOBER, M.F. & BRENNAN, S.E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. In *Language and speech*, 44 (2), 123-147. <https://doi.org/10.1177/00238309010440020101>
- BÜLOW, L., PFENNINGER, S.E. (2021). Introduction: Reconciling approaches to intra-individual variation in psycholinguistics and variationist sociolinguistics. In *Linguistics Vanguard*, 7 (s2), 1-10. <https://doi.org/10.1515/lingvan-2020-0027>
- 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. In *XV Convegno Nazionale AISV. Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale*. Studi AISV 6, 205-231.

- 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. <https://doi.org/10.1515/9783110816570.169>
- CHOWDHURY, S.A., STEPANOV, E., DANIELI, M. & RICCARDI, G. (2017). Functions of silences towards information flow in spoken conversation. In *Proceedings of the Workshop on Speech-Centric Natural Language Processing*, Copenhagen, Denmark, September, 1-9. <https://doi.org/10.18653/v1/W17-4601>
- CLARK, H.H., FOX TREE, J.E. (2002). Using uh and um in spontaneous speaking. In *Cognition*, 84 (1), 73-111. [https://doi.org/10.1016/S0010-0277\(02\)00017-3](https://doi.org/10.1016/S0010-0277(02)00017-3)
- CORLEY, M., HARTSUIKER, R.J. (2003). Hesitation in speech can... um... help a listener understand. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Boston, Massachusetts, USA, 31st July-2nd August, 25.
- CORLEY, M., STEWART, O.W. (2008). Hesitation disfluencies in spontaneous speech: The meaning of um. In *Language and Linguistics Compass*, 2 (4), 589-602. <https://doi.org/10.1111/j.1749-818X.2008.00068.x>
- CRIBLE, L., BUNT, H. (2016). Discourse markers and disfluencies: Integrating functional and formal annotations. In *Proceedings of the LREC 2016 Workshop ISA-12*, Portorož, Slovenia, May 23-28, 38-45.
- CRIBLE, L. (2018). *Discourse Markers and (Dis)fluency: Forms and functions across languages and registers*, (Vol. 286). Amsterdam/Philadelphia: John Benjamins Publishing Company. <https://doi.org/10.1075/pbns.286>
- CROCCO, C., SAVY, R. (2003). Fenomeni di esitazione e dintorni: una rassegna bibliografica. In CROCCO, C., SAVY, R. & CUTUGNO, F. (a cura di), *API. Archivio di Parlato Italiano*, DVD.
- DI NAPOLI, J. (2020). Filled pauses and prolongations in Roman Italian task-oriented dialogue. In *Laughter and Other Non-Verbal Vocalisations Workshop: Proceedings*. Bielefeld, 5 October 2020, 24-27.
- DUEZ, D. (1997). Acoustic markers of political power. In *Journal of Psycholinguistic Research*, 26 (6), 641-654. <https://doi.org/10.1023/A:1025008124631>
- EKLUND, R. (2001). Prolongations: A dark horse in the disfluency stable. In *Proceedings of DiSS 2001 Disfluency in Spontaneous Speech*, Edinburgh, Scotland, UK, 29-31 August 2001, 5-8.
- EKLUND, R. (2004). Disfluency in Swedish human-human and human-machine travel booking dialogues. PhD Dissertation, Linköping University Electronic Press.
- EPHRATT, M. (2008). The functions of silence. In *Journal of pragmatics*, 40 (11), 1909-1938. <https://doi.org/10.1016/j.pragma.2008.03.009>
- ESPOSITO, A., STEJSKAL, V., SMÉKAL, Z. & BOURBAKIS, N. (2007). The significance of empty speech pauses: Cognitive and algorithmic issues. In *International Symposium on Brain, Vision, and Artificial Intelligence*. Naples, Italy, 10-12 October, 542-554. https://doi.org/10.1007/978-3-540-75555-5_52
- FEHRINGER, C., FRY, C. (2007). Hesitation phenomena in the language production of bilingual speakers: The role of working memory. In *Folia Linguistica*, 41 (1-2), 37-72. <https://doi.org/10.1515/flin.41.1-2.37>

- FINLAYSON, I.R., CORLEY, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? In *Psychonomic bulletin & review*, 19 (5), 921-928. <https://doi.org/10.3758/s13423-012-0279-x>
- GINZBURG, J., FERNÁNDEZ, R., SCHLANGEN, D. (2014). Disfluencies as intra-utterance dialogue moves. In *Semantics and Pragmatics*, 7 (9), 1-64. <https://doi.org/10.3765/sp.7.9>
- HARTSUIKER, R.J., NOTEBAERT, L. (2009). Lexical access problems lead to disfluencies in speech. In *Experimental psychology*, 57, 169-177. <https://doi.org/10.1027/1618-3169/a000021>
- HORVÁTH, V. (2010). Filled pauses in Hungarian: Their phonetic form and function. In *Acta Linguistica Hungarica* (Since 2017 *Acta Linguistica Academica*), 57 (2-3), 288-306. <https://doi.org/10.1556/ALing.57.2010.2-3.6>
- KENDALL, T.S. (2009). Speech rate, pause, and linguistic variation: An examination through the sociolinguistic archive and analysis project. PhD Dissertation, Duke University.
- KJELLMER, G. (2003). Hesitation. in defence of er and erm. In *English Studies*, 84 (2), 170-198. <https://doi.org/10.1076/enst.84.2.170.14903>
- KOSMALA, L., MORGENSTERN, A. (2017). A preliminary study of hesitation phenomena in l1 and l2 productions: a multimodal approach. In EKLUND, R., ROSE, R. (Eds.), *Proceedings of DiSS 2017 The 8th Workshop on Disfluency in Spontaneous Speech*, KTH, Royal Institute of Technology, Stockholm, Sweden, 18-19 August 2017, 37-40.
- KOSMALA, L., MORGENSTERN, A. (2018). Should 'uh' and 'um' be categorized as markers of disfluency? the use of fillers in a challenging conversational context. In *Fluency and disfluency across languages and language varieties*, 4, 67.
- LANDIS, J., KOCH, G. (1977). The measurement of observer agreement for categorical data. In *Biometrics*, 33 (1), 159-74. <https://doi.org/10.2307/2529310>
- LENGTH, R. (2020). emmeans: Estimated Marginal Means, aka Least-Squares Means. [R package] Version 1.4.6.
- LEVELT, W.J. (1983). Monitoring and self-repair in speech. In *Cognition*, 14 (1), 41-104.
- LEVELT, W.J. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press. [https://doi.org/10.1016/0010-0277\(83\)90026-4](https://doi.org/10.1016/0010-0277(83)90026-4)
- LICKLEY, R.J. (2015). Fluency and Disfluency. In REDFORD, M.A. (Ed.), *The handbook of speech production*. John Wiley & Sons, 445-474. <https://doi.org/10.1002/9781118584156.ch20>
- MCDUGALL, K., DUCKWORTH, M. (2017). Profiling fluency: an analysis of individual variation in disfluencies in adult males. In *Speech Communication*, 95, 16-27. <https://doi.org/10.1016/j.specom.2017.10.001>
- MONIZ, H. (2013). Processing disfluencies in European Portuguese. PhD Dissertation, University of Lisbon.
- MONIZ, H., MATA, A.I. & CÉU VIANA, M.C. (2007). On filled-pauses and prolongations in European Portuguese. In *8th Annual Conference of the International Speech Communication Association*. Antwerp, Belgium, 27-31 August, 2820-2824 <https://doi.org/10.21437/Interspeech.2007-695>
- MONIZ, H., TRANCOSO, I. & MATA, A.I. (2010). Disfluencies and the perspective of prosodic fluency. In ESPOSITO, A., CAMPBELL, N., VOGEL, C., HUSSAIN, A. & NIJHOLT,

A. (Eds.), *Development of multimodal interfaces: active listening and synchrony*. Springer, 382-396. https://doi.org/10.1007/978-3-642-12397-9_33

MONIZ, H., BATISTA, F., MATA, A.I. & TRANCOSO, I. (2014). Speaking style effects in the production of disfluencies. In *Speech communication*, 65, 20-35. <https://doi.org/10.1016/j.specom.2014.05.004>

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.

ORIGLIA, A., SAVY, R., CATALDO, V., SCETTINO, L., ANSANI, A., SESSA, I., CHERA, A. & POGGI, I. (2019). Human, all too human. Towards a disfluent Virtual Tourist Guide. In *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*, Larnaca, Cyprus, 9-12 June 2019, 393-399. <https://doi.org/10.1145/3314183.3323866>

O'CONNELL, D.C., KOWAL, S. (2005). Uh and um revisited: Are they interjections forsignaling delay? In *Journal of Psycholinguistic Research*, 34 (6), 555-576. <https://doi.org/10.1007/s10936-005-9164-3>

O'SHAUGHNESSY, D. (1992). Recognition of hesitations in spontaneous speech. In *Proceedings: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92)*, San Francisco, USA, 23-26 March 1992, 521-524. <https://doi.org/10.1109/ICASSP.1992.225857>

OVIATT, S. (1995). Predicting spoken disfluencies during human-computer interaction. In *Computer Speech and Language*, 9 (1), 19-36. <https://doi.org/10.1006/csla.1995.0002>

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

ROBERTS, P.M., MELTZER, A., & WILDING, J. (2009). Disfluencies in non-stuttering adults across sample lengths and topics. In *Journal of communication disorders*, 42 (6), 414-427. <https://doi.org/10.1016/j.jcomdis.2009.06.001>

ROCHESTER, S.R. (1973). The significance of pauses in spontaneous speech. In *Journal of Psycholinguistic Research*, 2 (1), 51-81. <https://doi.org/10.1007/BF01067111>

SACKS, H., SCHEGLOFF, E.A. & JEFFERSON, G. (1978). A simplest systematics for the organization of turn taking for conversation. In SCHENKEIN, J. (Ed). *Studies in the organization of conversational interaction*. Elsevier, 7-55. <https://doi.org/10.1016/B978-0-12-623550-0.50008-2>

SCHEGLOFF, E.A. (2010). Some other "uh (m)"s. In *Discourse Processes*, 47 (2), 130-174. <https://doi.org/10.1080/01638530903223380>

SCETTINO, L., CATALDO, V. (2019). Lexicalized pauses in Italian. In *Proceedings of the 10th International Conference of Experimental Linguistics*, Lisbon, Portugal, 25-27 September, 189-192. <https://doi.org/10.36505/ExLing-2019/10/0047/000409>

SCETTINO, L., DI MARO, M. & CUTUGNO, F. (2020). Silent pauses as clarification trigger. In *Laughter and Other Non-Verbal Vocalisations Workshop: Proceedings*, Bielefeld, 5 October 2020, 51-54.

SCETTINO, L., CATALDO, V., ALFANO, I. & LEO, G. (forthcoming). A Protocol for Formal and Functional Annotation of Disfluency.

- SCHIFFRIN, D. (1987). *Discourse markers*. Number 5. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511611841>
- SCHNADT, M.J., CORLEY, M. (2006). The influence of lexical, conceptual and planning based factors on disfluency production. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vancouver, Canada, 26-29 July 2006, 28 (28), 750-755.
- SHRIBERG, E. (1994). Preliminaries to a theory of speech disfluencies. PhD Dissertation, University of California, Berkeley.
- SHRIBERG, E. (2001). To 'errrr' is human: ecology and acoustics of speech disfluencies. In *Journal of the International Phonetic Association*, 31, 153-169. <https://doi.org/10.1017/S0025100301001128>
- SKANTZE, G., HJALMARSSON, A. (2013). Towards incremental speech generation in conversational systems. In *Computer Speech & Language*, 27 (1), 243-262. <https://doi.org/10.1016/j.csl.2012.05.004>
- SLOETJES, H., WITTENBURG, P. (2008). Annotation by category-ELAN and ISO DCR. In *Proceedings of the 6th international Conference on Language Resources and Evaluation (LREC 2008)*, Marrakesh, Morocco, 28-30 May 2008, 816-820.
- STRANGERT, E. (2003). Emphasis by pausing. In *Proceedings of the 15th international congress of phonetic sciences*, Barcelona, Spain, 3-9 August 2003, 2477-2480.
- SWERTS, M. (1998). Filled pauses as markers of discourse structure. In *Journal of pragmatics*, 30 (4), 485-496. [https://doi.org/10.1016/S0378-2166\(98\)00014-9](https://doi.org/10.1016/S0378-2166(98)00014-9)
- TOTTIE, G. (2016). Planning what to say: Uh and um among the pragmatic markers. In KALTENBÖCK, G., KEIZER, E. & LOHMANN, A. (Eds.). *Outside the clause: Form and function of extra-clausal constituents*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 178, 97-122. <https://doi.org/10.1075/slcs.178.04tot>
- TOTTIE, G. (2020). Word-Search As Word-Formation?: The Case Of “Uh” And “Um”. In NÚÑEZ-PERTEJO, P., LÓPEZ-COUSO, M.J., MÉNDEZ-NAYA, B. & PÉREZ-GUERRA, J. (Eds.), *Crossing linguistic boundaries: systemic, synchronic and diachronic variation in english*. London: Bloomsbury Academic, 29-42. <https://doi.org/10.5040/9781350053885.ch-002>
- VAN DONZEL, M.E., KOOPMANS-VAN BEINUM, F.J. (1996). Pausing strategies in discourse in Dutch. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96. 2*. Philadelphia, PA, USA, 3-6 October, 1996, 1029-1032.
- VENABLES, W.N., RIPLEY, B.D. (2002). *Modern Applied Statistics with S*. Fourth Edition. New York: Springer. <https://doi.org/10.1007/978-0-387-21706-2>
- VOGHERA, M. (2017). *Dal parlato alla grammatica*. Roma: Carocci.
- WILSON, D. (2011). The conceptual-procedural distinction: Past, present and future. In *Procedural meaning: Problems and perspectives*, 25, 3-31. [https://doi.org/10.1108/S1472-7870\(2011\)0000025005](https://doi.org/10.1108/S1472-7870(2011)0000025005)
- WILSON, D., SPERBER, D. (2012). Linguistic form and relevance. In WILSON, D., SPERBER, D. (Eds.), *Meaning and Relevance*. Cambridge: Cambridge University Press, 149-168. <https://doi.org/10.1017/CBO9781139028370.010>