

LOREDANA SCHETTINO, IOLANDA ALFANO, VIOLETTA CATALDO,
GIOVANNI LEO

A Crosslinguistic Study on Filled Pauses and Prolongations in Italian and Spanish¹

Although disfluencies exhibit universal properties, comparative studies have demonstrated cross-linguistic differences as well. This study provides a first comparison between Italian and Spanish, investigating formal and functional features of filled pauses (FPs) and prolongations (PRLs) in dialogical speech. For both phenomena, duration was examined. As regards FPs, we looked at their segmental composition and the surrounding context; for PRLs, we considered the lexical category of the word and the position within the word in which they occur. Beyond individual variability, both Italian and Spanish speakers use more PRLs than FPs, with no interlinguistic duration differences. Furthermore, vocalic final-word PRLs are cross-linguistically preferred. However, Italian speakers present a general higher rate of disfluencies. Finally, FPs show a different segmental composition – related to the language-specific phonetic/phonological inventory –, and seem to be involved in different sub-functions.

Keywords: disfluencies, filled pauses, prolongations, Italian, Spanish.

1. *Introduction*

This study concerns two particular types of speech phenomena in Italian and Spanish task-oriented dialogues, that is Filled Pauses and Prolongations. These elements are usually considered to belong to a heterogeneous class of phenomena that characterize human spontaneous speech and either suspend or edit its production, thus apparently affecting its “fluency”; for this reason, they have been commonly referred to as “disfluencies” (see Lickley, 2015 for an overview).

In the last decades, however, numerous studies have highlighted the significance of such phenomena in the economy of speech. More specifically, human spontaneous speech results from complex online processes including planning, coding, and articulation (Levelt, 1989). Some disfluencies represent flexible and efficient tools which speakers can use to manage their own production (Allwood, Nivre & Ahlsén, 1990; Crocco, Savy, 2003; Voghera, 2017), by either editing something

¹ This article is the result of the collaboration among the authors. However, for academic purposes only, Loredana Schettino is responsible for §§ 1, 2.1, 4.2 and 5.1, Iolanda Alfano for §§ 2.2 and 5.2, Violetta Cataldo for §§ 4.3 and 5.3, Giovanni Leo for §§ 3, 4.1 and 5.4. All the authors are responsible for §§ 6 and 7.

already uttered (also known as “Repairs” or “Backward-Looking Disfluencies”) or by monitoring for something about to be uttered (also identified as “Hesitations” or “Forward-Looking Disfluencies”, Ginzburg, Fernández & Schlangen, 2014). In particular, filled pauses suspend speech with non-verbal vocalizations and/or nasalizations, i.e., *eeh*, *ehm*, *mhh*, whereas prolongations consist of marked lengthening of segmental material (Betz, 2020). Both delay the message delivery, thus reducing the temporal pressure resulting from 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 listeners to process information.

The occurrences and the surface realizations of these phenomena are found to vary due to contextual factors and the related cognitive demands (see Bortfeld, Leon, Bloom, Schober & Brennan, 2001), the contextually-determined functions in discourse (see Schettino, Betz, Cutugno & Wagner, 2021), individual factors (see Braun, Rosin, 2015), and language-specific features (see Clark, Fox Tree, 2002).

This study concerns the common and language-specific uses of filled pauses and prolongations in Italian and Spanish dialogues.

2. *Related Work*

2.1 Filled Pauses and Prolongations

Among the different phenomena that may generate suspensions in speech, “non-lexical vocalizations”, also referred to as “filled pauses” or “fillers”, represent the most salient, most recognizable, and – may be for this reason – most studied expressions of “Forward-Looking” or prospective disfluency. Since the influential study by Clark and Fox Tree (2002), arguing for the speakers’ use of different types of filled pauses to signal an upcoming short, “uh”, or long, “uhm”, delay, it has been intensively debated on the “signal” *versus* “symptom” hypotheses. That is, studies have searched for and provided evidence supporting the use of filled pauses as proper words (Clark, Fox Tree, 2002), or as a by-product of speech production processes (O’Connell, Kowal, 2005; Silber-Varod, Gósy & Lerner, 2021), which nonetheless may provide listeners with valuable information about the ongoing discourse (see Corley, Stewart, 2008; Finlayson, Corley, 2012). More recently, Tottie (2019, 2020) has proposed a different “word-class categorization” for spoken and written uses of filled pauses. According to this view, the former are ascribable to the “fuzzy category of inserts” and the latter to the category of stance adverbs (Tottie, 2019: 128). However, it has been observed that these different uses should be interpreted as preferential and not necessarily linked to the written and spoken modality (Voghera, 2017). Generally, filled pauses have been acknowledged to mark suspensions due to speech planning, corrections, turn management, discourse structure, upcoming new or prominent information, and attention recall (Kjellmer, 2003; Schegloff, 2010; Kosmala, Morgenstern, 2018; Tottie, 2020; Belz, 2021).

From a formal point of view, filled pauses have been described in the literature as non-lexical elements that are realized by a close to a mid-central vocalic and/or nasal phone (see Lickley, 2015; Belz, 2021). However, these vocalizations have been observed to vary across different languages and varieties, as speakers tend to use articulatory models that are strictly linked to their native phonological inventory (Clark, Fox Tree, 2002; Giannini, 2003a; Lickley, 2015). Furthermore, the phonetic-prosodic features of these phenomena have been observed to correlate with functional differences. In particular, a longer duration characterizes occurrences involved in cognitively demanding production processes, i.e., when they occur at the beginning of a phrase (in Dutch monologues, Swerts, 1998), when they introduce new information (in Hungarian monologues, Horváth, 2010), or when they signal problems in the retrieval of a specific word (in Italian monologues, Cataldo, Schettino, Savy, Poggi, Origlia, Ansani, Sessa & Chiera, 2019).

Speakers can temporarily suspend their speech also by prolonging segments beyond their normal duration. Such a lengthening may serve different functions, e.g., marking prosodic phrase boundaries (Albano Leoni, Maturi, 2002), cueing prominence (Bishop, Kuo & Kim, 2020), or signalling hesitation (Eklund, 2004). “Disfluent” lengthening has been described as “a marked prolongation of one or more phones, resulting in above-average syllable and word duration [...] This coincides with a local reduction in speech rate that is not expected by the listener, causing an impression of disfluency and hesitation” (Betz, 2020: 14). Disfluent and non-disfluent lengthening has been observed to be characterized by specific pitch features: prolongations are generally realized with lower pitch range and/or a slowly falling contour, whereas phrase-final lengthenings are usually associated with higher pitch range and boundary (e.g., rising/falling) pitch contour (Savino, Refice, 2000; Shriberg, 2001; Giannini, 2003b; Moniz, 2013; Betz, Eklund & Wagner, 2017). However, given the difficulties and relevance of context in discriminating prolongations, using a “pragmatic” approach based on perceptive criteria has been suggested to provide a “safer” and more suitable solution than relying on temporal thresholds or pitch features (Lickley, 2015).

In different languages, filled pauses have been found to have longer average duration than lengthenings (Swedish, Tok Pisin: Eklund, 2001, 2004; German: Betz et al., 2017; European Portuguese: Moniz, Mata & Viana, 2007; Italian: Giannini, 2003a, 2003b; Cataldo et al., 2019; Di Napoli, 2020). This suggests that these two types of disfluency phenomena may be involved in different ways in the online speech planning. Moreover, in his thesis, Betz (2020) argues that lengthenings represent less salient disfluency elements than silences or filled pauses “which are islands in the speech signal, whereas lengthening stretches the message by ongoing phonation [...] [and is] the softest measure a speaker can apply to solve problems in speech planning” (Betz, 2020: 14).

Besides discourse contextual factors, even individual psycho- and socio-linguistic demands can shape speech planning and production strategies (McDougall, Duckworth, 2017), resulting in speaker-specific uses of disfluency phenomena (see Van Donzel,

Koopmans-van Beinum, 1996; Betz, Lopez Gambino, 2016; Llisterri, Machuca & Ríos, 2019a, b). Hence, various studies suggest that disfluencies may represent a further tool forensic phoneticians may utilize for the identification of speakers (Ishihara, Kinoshita, 2010; Braun, Rosin, 2015; McDougall, Duckworth, 2017).

Moreover, it has been observed that disfluency phenomena may be subjected to the structural and usage constraints imposed by different linguistic systems, as described in the next paragraph.

2.2 Language-specific Features

As noted by McDougall and Duckworth (2017), comparisons across studies are difficult because of differences in the taxonomies used, in the speech style, and in the ways each study compares the occurrence of disfluencies against the whole speech sample (e.g., per 100 syllables, per 100 words, or per minute of speech). Furthermore, some authors rely on perceptive criteria, while others use acoustic duration thresholds.

Even if disfluencies seem to exhibit universal properties, comparative studies have demonstrated cross-linguistic differences, thus giving support to Clark and Fox Tree's (2002) argument that fillers are language-specific. Besides, it has been shown that the use of different hesitation markers and their interplay may respond not only to phonological, syntactic, and semantic constraints, but also to pragmatic culture-specific dynamics (Betz et al., 2021).

McDougall and Duckworth (2017) for British English and Llisterri, Machuca and Ríos (2022) for Spanish analyze disfluency phenomena with the aim of testing the extent to which they are employable for forensic discrimination across speakers. The results of these studies indicate that each speaker presents specific disfluency traits, although not all of these contribute equally to speaker identification. Cross-linguistically, the most frequent disfluencies in Spanish are vocalic prolongations and silent pauses (Llisterri et al., 2022), while the results on British English reveal the presence of silent and filled pauses, but a lower occurrence of segmental prolongations compared to Spanish (McDougall, Duckworth, 2017).

Language-specific patterns may concern not only a different distribution, but also a specific phonetic realization. Candea, Vasilescu and Adda-Decker (2005) study the vocalic peculiarities of "autonomous fillers" (our filled pauses) in several languages, i.e., the realization of a central or non-central timbre of their vocalic support, nasalizations or diphthongized segments. They examine the following eight languages: standard Arabic, Mandarin Chinese, French, German, Italian, European Portuguese, American English, and Latin American Spanish; they base fillers extraction on a minimum duration threshold of 200 ms. They consider three parameters: the duration, the F1/F2 values of the fillers' vocalic segments, and the fundamental frequency (f_0). Whereas f_0 and duration do not show significant differences among the eight languages, the acoustic analysis of F1/F2 reveals language-dependent characteristics. The number of occurrences of disfluencies per language is not considerable for Italian and Spanish corpora (57 and 93 occurrences,

respectively). However, their results strengthen the hypothesis of cross-linguistic timbre differences of the vocalic support of the autonomous fillers, showing that different languages admit various vocalic realizations. Spanish employs the closed-mid vowel [e], while Italian makes use of both central and non-central vocalic supports, i.e., the front open-mid vowel [ɛ].

De Leeuw (2007) reports significant differences among English, German, and Dutch hesitation markers in the proportion of vocalic, vocalic-nasal, and nasal markers, as well as in their positioning.

In the last decades, Eklund (2001) and various colleagues conducted a number of investigations on the characteristics and uses of prolongations in different languages. It has been observed that all types of segments could be subjected to lengthening, but vowels and sonorants are usually more prone to it within the language-specific phonological constraints. For instance, since vowel length is a distinctive feature in German, prolongation tends to be avoided on short vowels (Betz et al., 2017). Lengthened segments generally occur in the word-final syllable, although the distribution within word-initial/medial/final segments seems to be highly dependent on language-specific syllable structure and phonotactic rules. Prolongations tend to occur on functional rather than content words in English (O'Shaughnessy, 1995), Swedish, Tok Pisin (Eklund, 2001), German (Betz et al., 2017), Hungarian (Gósy, Eklund, 2017), whereas no clear-cut distinction was found in Mandarin (Lee, He, Huang, Tseng & Eklund, 2004) and Japanese (Den, 2003).

As far as Italian and Spanish are concerned, the relevant literature on fillers indicates that the most frequent vocalization is a vocalic element representable in Spanish as [e:] (Rebollo, 1997; Machuca, Llisterri & Ríos, 2015) and in Italian as [ə:] or [əm:] (Cataldo et al., 2019). Regarding the lexical category of the word affected by lengthening, most cases in Spanish were found to correspond to functional words, in particular, prepositions represent the highest percentage of cases (Machuca et al., 2015); in Italian, instead, a rather balanced distribution among open class (most frequently verbs, nouns, and adverbs) and closed class (mostly prepositions, conjunctions, and determiners) words was found (Di Napoli, 2020). As far as we know, there are no studies comparing Italian and Spanish by considering the same speech style and adopting the same annotation scheme.

3. *Research Aim*

This study sets out to investigate formal and functional features of filled pauses (FPs) and prolongations (PRLs) in dialogical speech of Italian and Spanish from a crosslinguistic perspective. As shown in §1, FPs and PRLs, like other disfluency phenomena, are present in every spoken language, but at the same time possess language-specific traits, which are strongly related to the linguistic (e.g., phoneme inventory, syntactic structures) and extralinguistic constraints (e.g., communicative context) of the language(s) in question. Hence, the investigation of FPs and PRLs in two genealogically related languages, Italian and Spanish, will add to the pre-existing

knowledge on the phenomena by highlighting commonalities and specificities in formal and functional terms by taking count of several factors, among which distributional and acoustic features (§4.3). To summarize, this work will attempt to provide an answer to the following questions:

- What are the common traits exhibited by FPs and PRLs in Italian and Spanish? What are, on the other hand, the language-specific ones?
- What level(s) (e.g., formal or functional) do the aforementioned patterns refer to?

4. Method

4.1 Corpus

The corpus for the analysis consists of 4 task-oriented dialogues (2 in Italian and 2 in Spanish), elicited through the “spot the difference” technique (Péan, Williams & Eskénazi, 1993). Table 1 summarizes the dataset by showing the duration of each dialogue per language. The Italian dialogues belong to the Neapolitan variety, whereas the Spanish ones belong to the variety of Spanish spoken in Barcelona. In all, 50 minutes of speech have been examined, about 25 minutes in Italian and about 26 in Spanish.

Table 1 - *Dataset for the analysis*

<i>Italian Dialogue</i>	<i>Duration (min)</i>	<i>Spanish Dialogue</i>	<i>Duration (min)</i>
<i>TDA01N</i>	<i>14.18</i>	<i>TDA01BCN</i>	<i>12.14</i>
<i>TDA02N</i>	<i>10.16</i>	<i>TDA02BCN</i>	<i>14.03</i>
<i>total</i>	<i>24.5</i>	<i>total</i>	<i>26.3</i>

The elicitation and the transcription of each dialogue has been conducted within the framework of the CLIPS project (Savy, Cutugno, 2009). The “spot the difference” elicitation method consists in a game, during which the communicating dyad has to identify the differences on two apparently identical pictures by relying solely on the verbal channel, without seeing each other. The peculiarity of this method is that both participants are placed on a par, which results in a highly interactive task. Although this elicitation technique has limits, it provides a certain naturalness as for the phonetic-prosodic traits, thus resulting in semi-spontaneous speech, with a low degree of discourse planning.

4.2 Disfluency Annotation

Given the cross-linguistic nature of the present study, we identified formal, functional, and structural parameters for filled pauses (FPs) and prolongations (PRLs) that can exhibit common or language-specific uses in Italian and Spanish dialogues.

We carried out the analysis by relying on an annotation scheme designed in previous studies (Cataldo et al., 2019; Schettino et al., 2021; Schettino, 2022). The scheme allows for a multi-layered annotation of disfluency phenomena, based on both

their formal structure and their contextual functions. Accordingly, the ELAN software was used (Sloetjes, Wittenburg, 2008), which permits multilevel linguistic annotation.

For the present study, we have focused on FPs and PRLs (Eklund, 2004; Betz, 2020). Each occurrence has been identified and labelled as follows: FP for non-verbal fillers realized as vocalization and/or nasalization; PRL for marked segmental lengthening. Since the annotation was previously designed for and applied to a different type of speech, namely almost monologic speech of Italian tourist guides (Cataldo et al., 2019; Schettino et al., 2021), we tested the inter-annotator agreement for the identification of the types of disfluency phenomena in dialogic speech conducted. The measured Cohen's κ reached 0.92, which stands for "high agreement" according to Landis, Koch (1977).

The two disfluency types serve a broad Forward-Looking function, namely they mark a temporary speech delay (Ginzburg et al., 2014). In the annotation system, such a function is further specified on the basis of the context of occurrence (for more details, see Schettino et al., 2021; Schettino, 2022).

- a. Word Searching (WS), when disfluencies are involved in lexical retrieval or lexical selection purposes (Tottie, 2020).
- b. Structuring (STR), for disfluencies occurring at the boundaries of syntactic or information structure, e.g., clauses and topic-comment, respectively.
- c. Focusing (FOC), associated with disfluencies marking upcoming "semantically heavy concepts or words" (Kjellmer, 2003)².
- d. Hesitative (HES), for disfluencies not fitting into any of the preceding sub-functions but triggered only by broad speech planning.

The inter-rater's agreement on the assignment of sub-functions to disfluency items reached Cohen's κ of 0.78 (substantial agreement, Landis, Koch, 1977).

4.3 Analysis Parameters

FPs and PRLs were then analyzed in Praat (Boersma, Weenink, 2018). Due to the structural differences of the disfluency types under investigation, we selected both general and disfluency-specific parameters. Accordingly, for both FPs and PRLs, we considered:

- Duration (ms).

As regards FP, we looked at:

- Segmental composition, namely the phonetic realization of vocalization and nasalization; specifically, we labelled the sequence of phones constituting the FP, e.g., [em], [ə];
- The surrounding context, i.e., preceding and following the FP, for each occurrence we annotated whether the FP was preceded or followed by speech or silence.

As for PRL, the analysis concerned:

² The label FOC was not assigned to phenomena signalling properly focalized elements, but rather to items involved in the planning and production of key information, e.g., new or emphasized elements, independently from syntactic structures.

- The lexical category of the word affected by lengthening, in order to test whether this disfluency type is more likely to affect functional (e.g., articles, conjunctions) or content words (e.g., nouns, verbs).
 - The position of PRL within the affected word, namely initial, medial or final position.
- The statistical analysis was performed in *R* (R Core Team, 2020), by means of Generalized Linear and Linear Mixed Models in order to control for individual variability ('lme4' package, Bates, Maechler, Bolker & Walker, 2015). Disfluency Type and Duration were set as the dependent variable, Function and Language as interacting independent variables, while Speaker was set as a random effect.

5. Results

A total of 406 disfluency items were analyzed, namely 142 FPs and 264 PRLs. First, the overall results on the frequency of occurrence and on the functions of the disfluencies are presented. Then, results per parameter are reported.

The Italian data present a higher incidence of disfluency rate (10,8 per minute) than Spanish data (5,4 per minute). Figure 1 shows the frequency of occurrence of the two types of phenomena in Italian and Spanish dialogues. Generally, PRLs are more frequent than FPs in both languages, although this difference is significantly larger in Spanish than in Italian (estimate: -2.34; SE: 0.58; z-value: -4.03; $p < 0.001$).

Actually, Italian speakers make greater use of both types of disfluencies (FPs = 110, PRLs = 154) with respect to Spanish (FPs = 31, PRLs = 110). The higher incidence per speaker in Italian compared to Spanish dialogues is reported in Table 2.

The observed cross-linguistic difference is significant beyond individual variability. Indeed, all speakers confirm this trend, especially for Spanish speakers presenting a higher incidence of PRLs compared to FPs.

Figure 1 - *Frequency of occurrence of FP and PRL in Italian (left) and Spanish (right) dialogues*

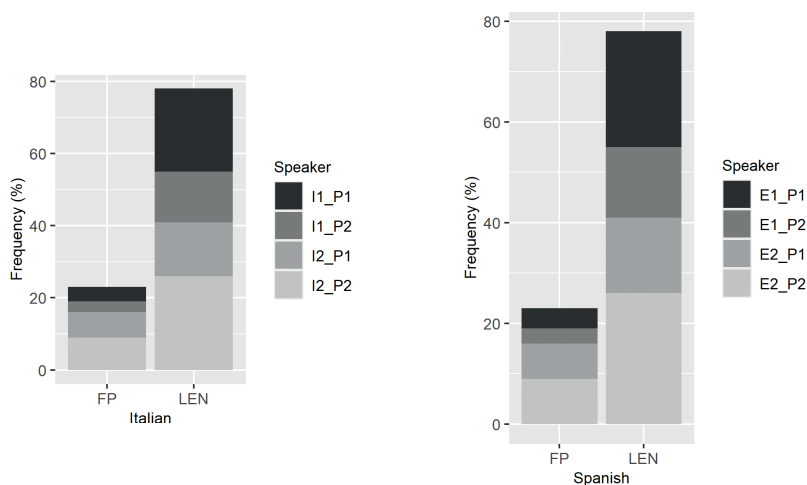
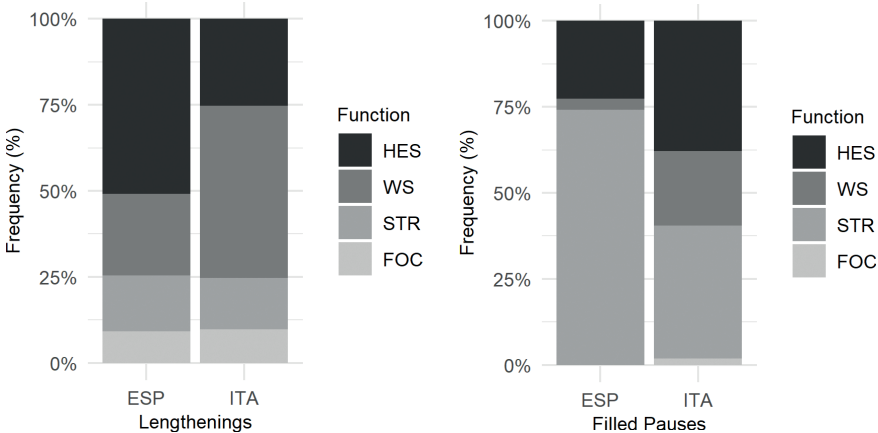


Table 2 - Incidence (%) of FP and PRL across speakers of Italian (left) and Spanish (right)

	<i>I1_P1</i>	<i>I1_P2</i>	<i>I2_P1</i>	<i>I2_P2</i>	<i>E1_P1</i>	<i>E1_P2</i>	<i>E2_P1</i>	<i>E2_P2</i>
<i>FP (%)</i>	31	29	10	41	5	4	10	12
<i>PRL (%)</i>	32	57	28	37	32	20	21	37

As for sub-functions (Fig. 2), in Spanish, FPs mostly serve a structuring function, whereas in Italian FPs exhibit a more even distribution. Note that FPs are rarely associated with the focusing function, which is, in any case, an infrequent sub-function in this corpus. On the other hand, PRLs are in fact significantly more used for hesitant, generic planning, and word searching, and far less for structuring. Indeed, comparing the use of the two selected phenomena, FPs are associated with the structuring function significantly more than PRLs in both Italian and Spanish (estimate: -2.54; SE: 0.56; z-value: -4.58; $p < 0.001$).

Figure 2 - Sub-functions of FP (left) and PRL (right) in Italian and Spanish dialogues



5.1 Duration

Duration was found to vary significantly between disfluency types but, crucially, not between the two languages. Indeed, in both Italian and Spanish, FPs are significantly longer than PRLs (estimate: -255.16; SE: 41.43; t value: -6.159), as shown in Table 3 and Figure 3.

Table 3 - Duration per disfluency type and language

<i>Language</i>	<i>Hes type</i>	<i>Dur (ms)</i>	<i>St. dev.</i>	<i>SE</i>
<i>ESP</i>	<i>FP</i>	530.11	336.65	60.46
<i>ESP</i>	<i>PRL</i>	272.07	111.09	10.59
<i>ITA</i>	<i>FP</i>	495.86	294.09	27.91
<i>ITA</i>	<i>PRL</i>	280.31	137.66	11.09

We also looked at the correlation between the duration of disfluencies and the corresponding sub-functions (Fig. 4). FPs are longer than PRLs regardless of their sub-function, whereas PRLs involved in word searching are on average longer than any other PRLs. This difference is significant for Italian PRLs (estimate: 199.36; SE: 66.05; t value: 3.018), while in Spanish it is a tendency that does not reach statistical significance.

Figure 3 - Duration (ms) of FP and PRL in Spanish and Italian dialogues

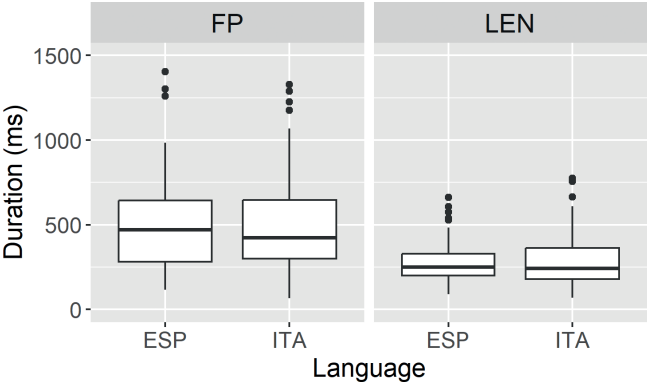
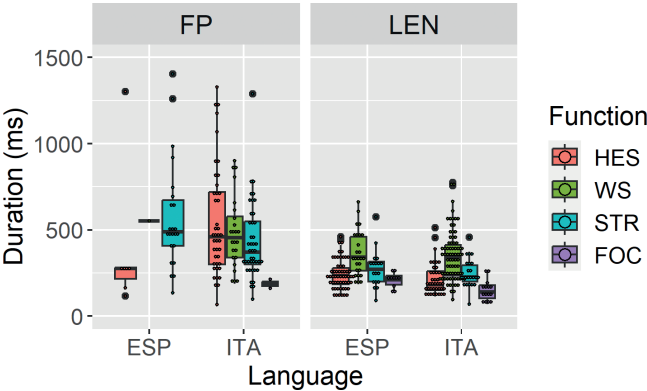


Figure 4 - Duration (ms) of FP and PRL per function in Spanish and Italian dialogues



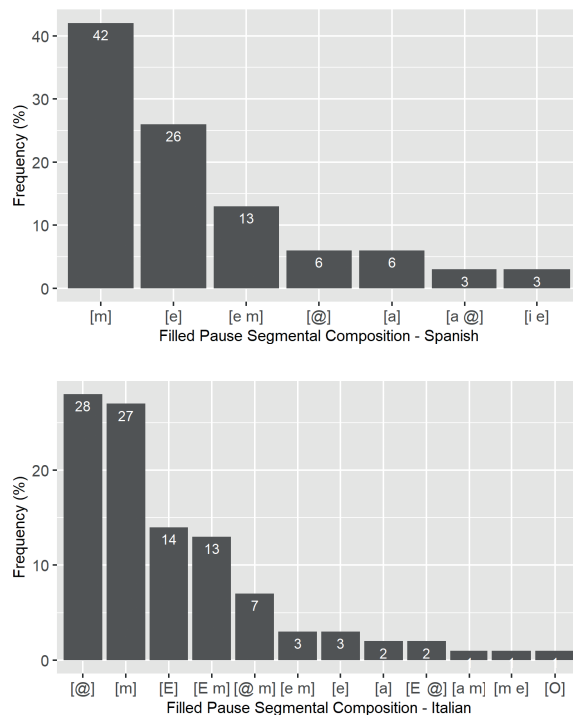
5.2 Segmental Composition

In Spanish, non-verbal vocalizations are slightly more frequently realized by nasalizations, namely [m] (55% of cases), followed by [e] and by their sequence, [em], and less frequently by [ə] (6% of cases).

On the other hand, in Italian, mid central vowel, schwa, turns to be the most frequent realization for FPs (28% of cases), followed by [m], [ɛ], and [ɛm] sequences. Generally, in the distinction between nasalized and non-nasalized items, FPs are evenly realized by nasalizations and vocalic sounds (51% and 49% of cases, respectively). Results on FP segmental composition are provided in Figure 5.

In these datasets, no differences were found in the correlation between specific phonetic realizations and sub-functions.

Figure 5 - *FP segmental composition in Spanish (top panel) and Italian (bottom panel) dialogues*

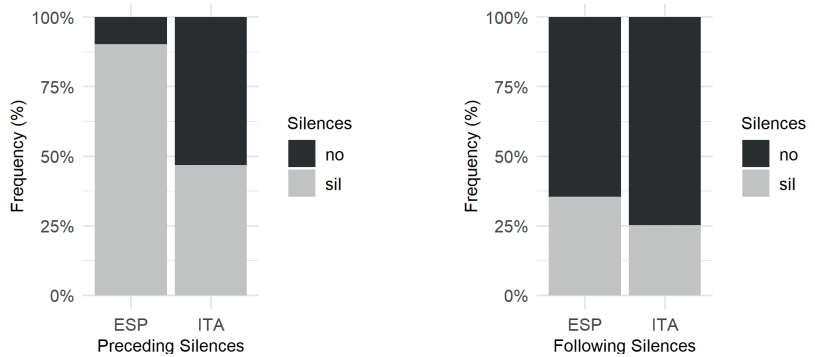


5.3 Surrounding Context

Considering the contextual elements (Fig. 6), Spanish FPs are almost exclusively preceded by a silence, whereas in Italian they are preceded by silences as often as speech. This cross-linguistic difference is statistically significant (estimate: -2.08; SE: 0.73; z value: -2.84; $p < 0.005$). Moreover, regardless of the language, the presence of a silence before a FP is significantly related to the structuring function (estimate: 0.81; SE: 0.43; z value: 2.11; $p < 0.035$).

As far as the following context is concerned, in both languages FPs are mostly followed by speech and not by silences.

Figure 6 - *Preceding (left) and following (right) FP in Spanish and Italian dialogues*

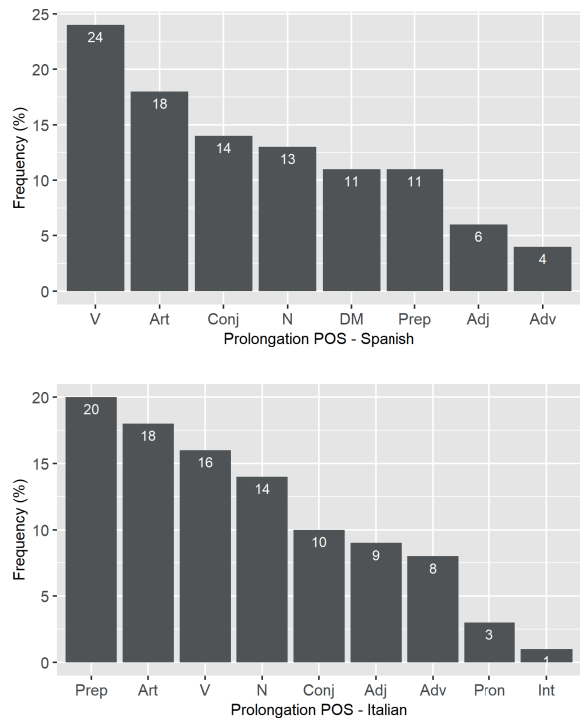


5.4 Prolongation Word Category and Position

As for the lexical categories affected by prolongations (Fig. 7), in Italian and Spanish, PRLs equally occur both on functional words (53% and 54%, respectively) and on content words (47% and 46%, respectively). The Italian distribution confirms previous studies (Di Napoli, 2020), while the distribution in Spanish does not show a marked preference for functional words and prepositions (contra Machuca et al., 2015).

Specifically, in the two languages, PRLs more commonly affect verbs, articles, and nouns.

Figure 7 - *PRL lexical category in Spanish (top panel) and Italian (bottom panel) dialogues*



As for the within-word position of PRLs, similar tendencies were observed in the two languages. Few cases of prolongations in initial (4% in Spanish and 3% in Italian) and medial (7% in Spanish and 1% in Italian) position. The vast majority of PRLs occur in word-final position, namely in 89% of cases in Spanish and 96% of cases in Italian.

6. *Discussion*

In this study, we explored speech management strategies adopted by speakers of different languages in the same communicative situation, that is in task-oriented dialogues in Italian and Spanish. More specifically, we focused on forward-looking disfluencies that generate a suspension in speech due to planning and production demands by means of non-lexical vocalizations (Filled Pauses) and marked lengthenings of word segments (Prolongations); we then analysed the way these phenomena are used by Italian and Spanish speakers based on their distribution, their contextually determined functions, their phonetic aspects, and possible correlations between these features.

From the results emerges, as expected, some individual variability in the use of both Filled pauses and Prolongations (see Betz, Lopez Gambino, 2016; McDougall, Duckworth, 2017). Beyond this source of variability, the analysis highlights, on the one hand, uses of these speech management phenomena that are common to Italian and Spanish speakers, and on the other hand, language-specific structures and dynamics.

Both Italian and Spanish speakers consistently use disfluency phenomena. In particular, segmental prolongation emerges as a more frequent hesitation strategy compared to filled pauses (as also found by Eklund, 2004). Speakers mostly rely on final-word prolongations for delaying speech due to generic planning and word searching, whereas filled pauses are more frequently used to take more time for planning processes involving the introduction of a new proposition.

Furthermore, in both languages, duration values characterize the two disfluency types differently: filled pauses are on average significantly longer than prolongations (as also found in previous literature involving different languages and styles, see §2.1).

These findings support the interpretation of prolongations as a more convenient and subtle means of taking time as opposed to filled pauses that represent, instead, more “salient” phenomena used to provide extra time when the planning and construction processes require it.

Then, prolongations occur quite evenly on content words (mostly verbs and nouns) and functional words (mostly articles, conjunctions, and prepositions). These findings are similar to those on Italian dialogues by Di Napoli (2020), but quite different from previous observations on Spanish by Machuca and colleagues (2015). However, this result may be due to the specific speech style of the analysed data and the resulting different frequency of lexical categories. Indeed, the observed task-oriented dialogues, aimed at the identification of differing details in similar

pictures, may be characterized by a relatively higher rate of content words, e.g., nouns and verbs.

Lastly, final-word position for segmental prolongation is cross-linguistically preferred, which may be due to the fact that Italian and Spanish display a similar distribution of syllabic structures, i.e., prevalently CV structures. However, the CVC type is slightly more frequent in Spanish, which reaches around 20%, than in Italian. The difference increases when considering the final-word position in stressed syllables, a position for which Spanish presents a clear majority of the CVC type (Alfano, 2008). Given the preference to lengthen vowel segments, one would have expected a different finding in this respect. Our data indicate rather that the preference for prolongation in word-final position seems to be not directly dependent on syllabic composition.

Besides these common patterns, a number of language-dependent uses and features have emerged from the analysis.

Firstly, compared to the Spanish dialogues, the Italian ones show a higher rate of disfluencies. Some differences also concern the distribution of the sub-functions, especially for filled pauses. While in Italian this type of forward-looking disfluency may be almost evenly involved in all the considered functions, in Spanish, filled pauses seem to correlate more frequently with the structuring function. Moreover, in the Spanish dialogues, filled pauses are almost exclusively preceded by a silence, whereas they seem to be equally preceded by speech or a silence in Italian. These observations may suggest a more controlled use of disfluency phenomena by Spanish speakers and, conversely, a higher tolerance for these elements by Italian speakers.

A feature that appears to be peculiar to Italian concerns the correlation between the duration of prolongations and specific sub-functions. As already observed in Cataldo et al. (2019), longer prolongations are more likely to be involved in lexical retrieval processes. Accordingly, this emerges as a robust feature in Italian monologic as well as dialogic speech.

Then, the feature that most of all reveals language-related specificities is the segmental composition of filled pauses. Indeed, it corroborates the assumption that speakers of different languages tend to generate realizations that are strictly linked to their phonological inventory even for the production of non-lexical vocalizations (Clark, Fox Tree 2002; Giannini, 2003a; Giannini 2003b; Ginzburg et al., 2014). In line with the literature, Spanish mostly employs [e:] and [m:] (Rebollo, 1997; Machuca et al., 2015), while in Italian the most frequent sounds are [ə:], [ɛ:], followed by nasals, e.g., [ə:m], [ɛ:m] (Giannini, 2003; Cataldo et al., 2019). In fact, the mid-central [ə:] phone is not acknowledged as an Italian phoneme. Therefore, in their contrastive corpus study, Candea et al. (2005) claim that Italian is the only language with a vocalic support which is not part of the vocalic system. However, this finding should be interpreted in the light of the regional variety examined in their work, which cannot be deduced in the paper. Indeed, it has been attested that the mid-central vocalic variant characterizes the dialectal substrate of the Neapolitan variety of Italian (Pellegrini, 1977; Loporcaro, 2009, 2016; Ledgeway, 2016). So, as

already observed by Giannini (2003), this realization appears to be connected to underlying dialectal sounds percolating into the local variety of Italian, rather than to other processes, such as the speaker's articulatory economy.

7. Conclusions

To conclude, this study suggests cross-linguistically shared uses of filled pauses and prolongations, in terms of phonetic "salience", and common characteristics resulting from similarities between the Italian and Spanish linguistic structures or the specific communicative context and goals. Language-specific uses concern, instead, the relative tolerance for the observed phenomena and phonological differences between Italian and Spanish.

This study has benefitted from the consideration of Italian and Spanish datasets collected by using the same elicitation technique and, therefore, constituting comparable speech data. Given the difficulties in comparing cross-study findings on disfluencies' uses and features, since they rely on different speech types and adopt different approaches, focused cross-linguistic investigations seem to be particularly relevant to the literature and provide more systematic and feasible comparisons in order to shed light on common and divergent disfluency features across languages. Nonetheless, caution is required in the interpretation of the described findings considering the rather small size of the datasets involved. Also, this investigation concerned two types of forward-looking disfluencies, but speakers may use other phenomena to suspend their speech for planning, such as silences or lexical fillers. Hence, the results emerged are only an indication of how and when the phenomena in question may be used in these languages. So, studying the different hesitation tools at speakers' disposal, and their interplay, may contribute to shedding light on more cross-linguistic different uses. For example, Spanish speakers may compensate for the scarce amount of filled pauses by relying on "lexical fillers", such as discourse markers that serve a planning function. Future investigations may follow this lead by employing larger datasets and including other disfluency types in order to deepen our understanding and gain a clearer picture of the strategies speakers may enact to manage their speech.

References

- ALBANO LEONI, F., MATURI, P. (2002). *Manuale di fonetica*. Roma: Carocci.
- ALFANO, I. (2008). Strutture sillabiche ed accentuali in italiano e in spagnolo. In *Testi e Linguaggi*, 2. Roma: Carocci, 18-36.
- 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>

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>

BELZ, M. (2021). *Die phonetik von äh und ähm: Akustische variation von füllpartikeln im deutschen*. Berlin: Springer Nature.

BETZ, S. (2020). Hesitations in Spoken Dialogue Systems. PhD Dissertation, University of Bielefeld.

BETZ, S., BRYHADYR, N., KOSMALA, L. & SCHETTINO, L. (2021). A Crosslinguistic Study on the Interplay of Fillers and Silences. In ROSE, R., EKLUND, R. (Eds.), *Proceedings of DiSS 2021, The 10th Workshop on Disfluency in Spontaneous Speech*, Paris, France, 25-27 August 2021, 47-52.

BETZ, S., EKLUND, R. & WAGNER, P. (2017). Prolongation in German. In ROSE, R., EKLUND, R. (Eds.), *Proceedings of DiSS 2017, The 8th Workshop on Disfluency in Spontaneous Speech*, Stockholm, Sweden, 18-19 August 2017, 13-16.

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.

BISHOP, J., KUO, G. & KIM, B. (2020). Phonology, phonetics, and signal-extrinsic factors in the perception of prosodic prominence: Evidence from rapid prosody transcription. In *Journal of Phonetics*, 82, 100977.

BORTFELD, H., LEON, S.D., BLOOM, J.E., SCHOBBER, 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>

BRAUN, A. & ROSIN, A. (2015). On the speaker specificity of hesitation markers. In MARIA WOLTERS, M., LIVINGSTONE J., BEATTIE, B., SMITH, R., MACMAHON, M., STUART-SMITH, J., SCOBIE, J., M. (Ed.), *Proceedings of the International Congress of Phonetic Sciences (ICPhS 2015)*, Glasgow, UK, 10-14 August 2015.

CANDEA, M., VASILESCU I. & ADDA-DECKER, M. (2005). Inter- and intra-language acoustic analysis of autonomous fillers. In *Proceedings of the 4th Workshop on Disfluency in Spontaneous Speech (DiSS 2005)*, Aix-en-Provence, France, 10-12 September 2005, 47-51.

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 PICCARDI, D., ARDOLINO, F. & CALAMAI, S. (Eds.), *XV Convegno Nazionale AISV. Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale*. Studi AISV 6. Milano: Officinaventuno, 205-231.

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., 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>

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.

DE LEEUW, E. (2007). Hesitation Markers in English, German, and Dutch. In *Journal of Germanic Linguistics*, 19(2), 85-114.

- DEN, Y. (2003). Some strategies in prolonging speech segments in spontaneous speech. In EKLUND, R. (Ed.) *Proceedings of the ISCA Tutorial and Research Workshop on Disfluency in Spontaneous Speech (DiSS 2003)*, Göteborg, Sweden, 5-8 September 2003, 87-90.
- DI NAPOLI, J. (2020). Filled pauses and prolongations in Roman Italian task-oriented dialogue. In *Proceedings of the Laughter and Other Non-Verbal Vocalisations Workshop*, Bielefeld, 5 October 2020, 24-27.
- EKLUND, R. (2004). Disfluency in Swedish human-human and human-machine travel booking dialogues. PhD Dissertation, Linköping University: Electronic Press.
- GIANNINI, A. (2003a). Hesitation phenomena in spontaneous Italian. In SOLÉ, M.J., RECASENS, D. & ROMERO, J. (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, Barcelona, Spain, 3-9 August 2003, 2653-2656.
- GIANNINI, A. (2003b). Vocalizzazioni e prolungamenti vocalici. In MAGNO CALDOGNETTO E., COSI P. (Eds.), *Voce, canto, parlato. Studi in onore di Franco Ferrero*. Padova: Unipress, 163-172.
- 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>
- 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>
- ISHIHARA, S. & KINOSHITA, Y. (2010). Filler words as a speaker classification feature. In TABAIN, M., FLETCHER, J., GRAYDEN, J., HAYEK, J. & BUTCHER, A (Eds.), *Proceedings of the 13th Australasian International Conference on Speech Science and Technology*, Melbourne, Australia, 14-16 December 2010, 34-37.
- 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. (2018). Should 'uh' and 'um' be categorized as markers of disfluency? the use of fillers in a challenging conversational context. In DEGAND, L., GILQUIN, G., MEURANT, L. & SIMON, A., C. (Eds.), *Fluency and Disfluency across Languages and Language Varieties. Corpora and Language in Use*. Louvain-la-Neuve: Presses universitaires de Louvain, 67-89.
- 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>
- LEDGEWAY, A. (2016). The dialects of southern Italy. In LEDGEWAY, A. & MAIDEN, M. (Eds.), *The Oxford guide to the Romance languages*, Vol. 1, 245-269, Oxford: Oxford University Press.
- LEE, T.-L., HE, Y.-F., HUANG, Y.-J., TSENG, S.-C. & EKLUND, R. (2004). Prolongation in spontaneous Mandarin. In *Proceedings of Interspeech 2004*, Jeju Island, Korea, 4-8 October 2004, 2181-2184.
- 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*. Chichester: John Wiley & Sons, 445-474. <https://doi.org/10.1002/9781118584156.ch20>

- LLISTERRI, J., MACHUCA, M.J. & RÍOS, A. (2019A). Caracterización del hablante con fines judiciales: fenómenos fónicos propios del habla espontánea. e-AESLA. In *Revista digital de lingüística aplicada*, 5, 265–278.
- LLISTERRI, J., MACHUCA, M.J. & RÍOS, A. (2019B). VILE-P: un corpus para el estudio prosódico de la variación inter e intralocutor. En LAHOZ-BENGOCHEA, J.M. & PÉREZ RAMÓN, R.J. (Eds.), *Subsidia: Tools and Resources for Speech Sciences / Subsidia: herramientas y recursos para las ciencias del habla*, Universidad de Málaga, 117-123. <https://hdl.handle.net/10630/18177>
- LLISTERRI, J., MACHUCA, M.J. & RÍOS, A. (2022). La función de las hesitaciones en la identificación del hablante. In BLECUA, B., CICRES, J., ESPEJEL, M. & MACHUCA, M.J. (Eds.), *Propuestas en fonética experimental: enfoques metodológicos y nuevas tecnologías*, Universitat de Girona, 160-164. <http://hdl.handle.net/10256/20770>
- LOPORCARO, M. (2009). *Profilo lingüístico dei dialetti italiani*. Roma-Bari: Laterza.
- LOPORCARO, M. (2016). L'Italia Dialettale. In LUBELLO, S. (Ed.), *Manuale di linguistica italiana*. Berlin: De Gruyter, 275-300.
- MACHUCA, M.J., LLISTERRI, J. & RÍOS, A. (2015). Las pausas sonoras y los alargamientos en español: un estudio preliminar. In *Normas. Revista de Estudios Lingüísticos Hispánicos*, 5, 81-96.
- MCDougall, 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 *Proceedings of the 8th Annual Conference of the International Speech Communication Association*. Antwerp, Belgium, 27-31 August 2007, 2820-2824 <https://doi.org/10.21437/Interspeech.2007-695>
- 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>
- PÉAN, V., WILLIAMS S. & ESKENAZI, M. (1993). The design and recording of ICY, a corpus for the study of intraspeaker variability and the characterization of speaking styles. In *Proceedings of Eurospeech 2003*, Geneva, Switzerland, 1-4 September 2003, 627-630.
- PELLEGRINI, G.B. (1977). *Carta dei dialetti d'Italia*. Pisa: Pacini.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- REBOLLO, L. (1997). Pausas y ritmo en la lengua oral. Didáctica de la pronunciación. In MORENO, F., GIL, M. & ALONSO, K. (Eds.), *El español como lengua extranjera: del pasado al futuro. Actas del VIII Congreso Internacional de la Asociación para la Enseñanza del Español como Lengua Extranjera*, Alcalá de Henares, Spain, 17-20 September 1997, 667-676.

- SAVY, R., CUTUGNO, F. (2009). Diatopic, diamesic and diaphasic variations in spoken Italian. In MAHLBERG, M., GONZÁLEZ-DÍAZ, V. & SMITH, C. (Eds.), *Proceedings of the 5th Corpus Linguistics Conference (CL2009)*, Liverpool, UK, 20-23 July 2009, 20-23.
- SCHEGLOFF, E.A. (2010). Some other “uh (m)”s. In *Discourse Processes*, 47(2), 130-174. <https://doi.org/10.1080/01638530903223380>
- SCHETTINO, L., BETZ, S., CUTUGNO, F. & WAGNER, P. (2021). Hesitations and individual variability in Italian tourist guides’ speech. In BERNADASCI, C., DIPINO, D., GARASSINO, D., NEGRINELLI, S. & SCHMID, S. (Eds.), *Speaker Individuality in Phonetics and Speech Sciences: Speech Technology and Forensic Applications*. Studi AISV 8, Milano: Officinaventuno, 243-262.
- SCHETTINO, L. (2022). The role of disfluencies in Italian discourse. Modelling and speech synthesis applications. PhD Dissertation, University of Salerno.
- 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>
- SILBER-VAROD, V., GÓSY, M. & LERNER, A. (2021). Is it a filler or a pause? A quantitative analysis of filled pauses in Hebrew. In KARPOV, A., POTAPOVA, R. (Eds.), *Proceedings of the 23rd International Conference on Speech and Computer*, SPECOM 2021, St. Petersburg, Russia, 27-30 September 2021, 638-648.
- 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.
- 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. (2019). From pause to word: Uh, um and er in written American English. In *English Language & Linguistics*, 23(1), 105-130.
- 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.
- VOGHERA, M. (2017). *Dal parlato alla grammatica*. Roma: Carocci.