Speech Dysfluencies in Formal Context.
Analysis based on Spontaneous Speech Corpora*

Leonardo Campillos, Manuel Alcántara
Laboratorio de Lingüística Informática
Universidad Autónoma de Madrid (Spain)
leonardo.campillos@uam.es, manuel.alcantara@uam.es

Abstract

This paper examines dysfluencies in a corpus of Spanish spontaneous speech in formal contexts, and proposes a classification of dysfluencies in this register to show their relations with both linguistic and extra-linguistic factors. Main disfluency phenomena which are considered in this work are repeats, false starts, filled pauses and incomplete words. Results are compared with two previous studies for Spanish and with those of spontaneous English in formal contexts.

1. Introduction

Research in speech dysfluencies has recently gained interest in the Linguistics and Speech recognition research community, as the last ICAME workshop or the NIST competitive evaluations have shown. Even the Cognitive sciences have been interested in these phenomena; for instance, Psycholinguistic researchers (e.g. Ferreira et al., 2004) have presented a model of dysfluency processing. However, for Spanish language, only two previous studies carried out with the C-ORAL-ROM corpus (González Ledesma et al., 2004; Toledano et al., 2005) deal with the issue, so we can say it has not received enough attention till now. The first one proposed a comprehensive typology of phenomena affecting transcription tasks, which were classified into types and frequencies, and were compared for informal, formal, and media interactions. The goal of that work was not to study dysfluencies, but to measure their influence in the transcription difficulties. The second study was also focused on differences between informal, formal, and media interactions, but from an acoustic point of view. Instead of manual transcriptions, difficulty of automatic processing of different types of spontaneous speech was tested by performing acoustic-phonetic decoding by means of a recognizer on parts of the corpus.

In our study, dysfluencies have been analyzed and compared to those of spontaneous English in formal contexts from Biber et al. (1999) in order to show how dependent on the language these phenomena are. First, we explain the design and characteristics of the speech corpora analysed, and the methodology we have followed. Then we will discuss every type of disfluency (repeats, false starts, filled pauses and incomplete words) with samples obtained from our corpus, giving a definition and frequency counts for every kind of phenomena. We will conclude with some remarks about the future work.

Results are important not only for the linguistic insights they provide. They will also help improve current automatic speech recognition systems for spontaneous language since dysfluencies are one of the main problems in ASR (Huang et al., 2000: 857).

*Acknowledgments: this research has been partially supported by the Spanish Ministry of Education and Science under the grant TIN2007-67407-C03-02.

2. Corpora and methodology

122,000 words taken from the MAVIR and C-ORAL-ROM corpora have been analyzed. C-ORAL-ROM is a multilingual spoken corpus of four Romance languages: Italian, French, Spanish and Portuguese (see details in Cresti and Moneglia, 2005). Every subcorpus is composed of approximately 300.000 words and contains transcriptions from formal and informal communicative contexts. This study only analyses the formal in natural context files, without transcriptions from the media. The MAVIR corpus is a multimodal spontaneous spoken corpus which is still under transcription and contains 51972 words so far. Recordings are been taken from professional conferences on Speech Technologies and corporate presentations of companies of this market held at professional meetings in Madrid. Language data are mainly in Spanish, although transcriptions of international researchers in English are included as well.

Both corpora have been annotated with prosodic and linguistic tags (including more than 5000 hand-annotated dysfluencies), and they are made up of 33 documents classified depending on type of text: business, academic conference, law, political debate, professional explanation, preaching, political speech, teaching, round table and professional conference (see table 1).

	Type of text	File	Dialogic style	Speakers	Words per file	Words per type of text
C-ORAL-ROM (Formal in natural context)	Business	enatbu01	dialogue	2	3394	9680
		enatbu02	dialogue	2	3247	
		enatbu03	monologue	1	3039	
	Academic conference	enatco01	monologue	1	3148	12678
		enatco02	monologue	1	3115	
		enatco03	monologue	1	3246	
		enatco04	monologue	1	3259	
	Law	enatla02	monologue	1	3129	3129
	Political debate	enatpd01	conversation	5	3079	6321
		enatpd02	conversation	5	3242	
	Professional explanation	enatpe01	monologue	1	3174	12919
		enatpe02	conversation	2	3336	
		enatpe03	dialogue	2	3155	
		enatpe04	monologue	1	3254	
	Preaching	enatpr01	monologue	1	1054	7255
		enatpr02	monologue	1	1643	
		enatpr03	monologue	1	1785	
		enatpr04	monologue	1	327	
		enatpr05	monologue	1	675	
		enatpr06	monologue	1	1771	
	Political speech	enatps01	monologue	2	3120	6343
		enatps02	conversation	3	3223	
	Teaching	enatte01	dialogue	2	3124	13025
		enatte02	conversación	3	3369	
		enatte03	monologue	4	3248	
		enatte04	monologue	3	3284	
MAVIR	Round table	mavir02	conversation	7	13530	13530
	Professional conference	mavir03	monologue	1	6681	38442
		mavir04	monologue	4	9439	
		mavir06	monologue	3	4320	
		mavir07	monologue	2	3829	
		mavir08	monologue	1	2981	
		mavir09	monologue	1	11192	
TOTAL				68	123412	

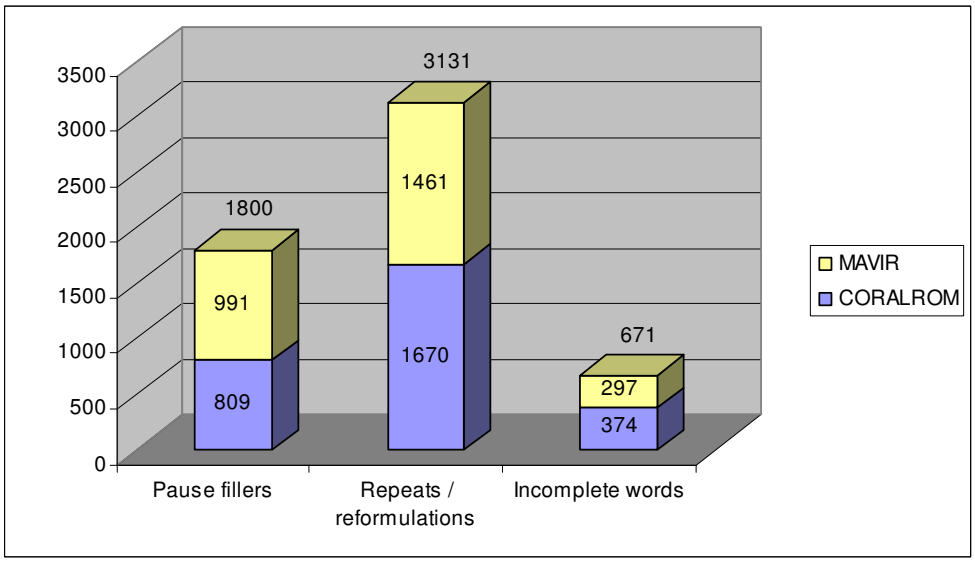
Table 1 – Distribution of the Spanish formal corpus analysed

The total number of speakers is 68: 49 in C-ORAL-ROM and 19 in MAVIR corpus. As far as the dialogic style is concerned, the C-ORAL-ROM transcriptions from formal natural context are mainly monologues (17), but they are also dialogues between two speakers (4) and conversations among three speakers or more (5). Regarding the MAVIR corpus, it contains 6 monologues and one conversation. Some remarks about our classification of speaking style and the number of speakers in every file must be made. We have considered as a monologue fragments of speech produced by the same speaker, but in a file which has been classified as a monologue different speakers may appear. For instance, in the MAVIR corpus, three files (mavir04, mavir06 and mavir07) are transcriptions of juxtaposed monologues produced by different speakers who explain research projects one at a time. We didn't consider these transcriptions as conversations due to the lack of interaction among the speakers. The tags of 'dialogue' or 'conversation' should be reserved for a more interactive form of speech where participants react to each others and their utterances may even overlap. Notwithstanding that consideration, some files which were classified as conversations in C-ORAL-ROM could be considered as monologues. For example, in two lessons recorded at university (files enatte03 and enatte04), the main speaker is the professor who gives his lecture, although two students raise some questions in a specific moment. Even though the professor interacts with the students when he answers their questions, we could consider these speeches as interrupted monologues. For that reason, the classification of transcriptions regarding their dialogic style should be analyzed cautiously.

Linguistic and extra-linguistic factors to be considered have been chosen following the state-of-the-art literature of dysfluencies in other languages (Shriberg, 1994; Biber et al., 1999: 1052-1066). Linguistic factors are: word-related aspects of dysfluencies (e.g. which words are repeated and frequency of repetitions) and syntactic complexity (e.g. how syntactically complex are parts following a disfluency). Extralinguistic factors are: dialogic style, number of speakers, and speech domain. They altogether help us to explain differences in occurrence frequency of dysfluencies. Regarding the filled pauses, we have also carried out a descriptive acoustic analysis of *eh* and *ah*, using Praat software.

General findings in the analysis show that, in our formal Spanish corpora, frequency clearly varies going from one to 120 dysfluencies every 1,000 words. Out of 123412 words, there appeared 5602 dysfluencies, giving a mean dysfluency/word ratio of 0,0453. This means that approximately every 20 words a repetition, a reformulation, a filled pause or a fragmented word occurs (almost 5 dysfluencies every 100 words). This ratio in our formal corpus seems to be similarly correlated with data for the English language obtained by Ferreira (2004:723), who indicates that there will be at least 6 dysfluencies for every 100 words; and correlated also with research by Shriberg (2005: 2), although this researcher analysed three corpus whose characteristics are quite different from ours (man-machine interaction, telephonic dialogues and air travel planning dialogs; see Shriberg, 1994: 34).

The most frequent dysfluency phenomena in our formal spoken Spanish data are repetitions and reformulations altogether, followed by filled pauses, whereas fragmented words appear with the lowest frequency (see graph 1).



Graph 1 – Number of every type of dysfluency phenomena in formal spoken Spanish

3. Repeats and reformulations

3.1. General findings

Repeats and reformulations are altogether the most frequent type of disfluency phenomena in our data. Curiously, considering the POS categories of the words before and after the repeat or the reformulation, they are predominantly prepositions. This could be explained by the cognitive process of accessing the mental lexicon. If a lexical word must be retrieved and the speaker is still thinking on the best choice (or decides to revise what has been said), the word immediately before the core of the phrase (the lexical unit: a noun or a verb) is repeated or reformulated. This repeated word is a preposition or a determiner the majority of the times.

3.2. Repeats

In spontaneous speech, it is very usual to produce repetitions of the linguistic material: usually, one word, but there also occur repeats of sequences of short words. Typically, they are a strategy to keep the turn and continue speaking when thinking a term or the best way to convey an idea.

1) un mercado que es / difícil *de [/] de* conquistar (mavir02)
['a market that is / difficult to [/] to conquer']

2) el ministerio nos ha dicho *que [/] que* está estudiando el tema (enatbu02)
['the ministry has told us that [/] that they are studying the issue']

3) vivimos *de lo que [/] de lo que* vendemos (mavir02)
['we live on what [/] on what we sell']

Repeats often appear in two different turns of the same speaker who is interrupted:

4) *INM: hablaban en inglés / *cuando [/]*
*JOS: sí / sí //
*INM: *cuando* / entraron en el centro (enatpe03)

[*INM: 'they could speak English *when they [/]*
*JOS: yes yes //
*INM: *when they* entered the college']

Repeats can be complete or partial (Moneglia, 2005: 27), and the first segment is often a fragment of the repeated word (Biber et al., 1999: 1055):

5) el planteamiento de la estructura sigue *&exacta [/] exactamente* igual (enatbu02)
['the approach of the structure continues &exact [/] exactly the same']

6) la administración *es &inaccesi [/] es inaccesible* a nuestros proyectos (mavir02)
['the administration is &inaccess [/] inaccessible to our projects']

Regarding the other kinds of dysfluencies, repeats occur usually along with pause fillers:

7) *en &eh [/] en* el mundo / digamos los PDF es el [/] el formato estándar (mavir09)
['in er [/] in the world / let's say the PDF is the [/] standard format']

8) toda la actividad / *que &mm [/] que* una empresa de nuestras características debe desarrollar
['all the activity / that um [/] that a company with our characteristics must develop'] (mavir02)

We should point out that repetitions may also perform a rhetoric function. Speakers usually consciously repeat a word or a phrase in their discourse in order to stress an idea or to make sure that the listeners understand what they have said. This specially happens in academic speech: professors and teachers very well know the pedagogic function of repeating a key concept. Our analysis has only concentrated in mechanic or unconscious repetitions of words or phrases, which are transcribed with the **[/]** mark. Conscious, rhetoric repetitions are not transcribed with this mark, as it occurs in this example, where the speaker repeats *sin embargo* ('however, nevertheless') to emphasize a contrast of two ideas:

es / un razonamiento / totalmente / solitario // *y sin embargo / y sin embargo* / la actividad intelectual puesta en marcha / la demostración / qué quiere decir?
['this is an utterly / solitary / reasoning // *and however / and however* / the intellectual activity which has been set off / the demonstration / what does it mean?] (enatco03)

The corpus findings we present below were automatically obtained comparing the words before and after the **[/]** mark. Due to this fact, rhetoric repetitions which could have been interpreted as unconscious repetitions by the transcribers could have been considered as mechanic repetitions in our automatic analysis. This is an issue derived from the interpretation of the data during the transcription process, and it is constantly present when dealing with spoken data. Although the intonation may help to disambiguate a mechanic repetition and a rhetoric repetition, the discurse function of a repetition is not always clear for the transcribers.

3.2.1. Corpus findings

Analysis shows that most repeated forms are function words (mainly prepositions, conjunctions, articles), with *de* ('of '), *en* ('in'), *el* ('the') and *y* ('and') on the top of the list. These are mainly the same words (in a different order) which are more the most frequent in Spanish (an abridged frequency word list is shown in Moreno Sandoval et al., 2005: 160). Besides, repetitions of sequences of words also occur mainly in functional words (pronouns, articles, prepositions). Lexical words such as nouns or adjectives are seldom repeated, and only repeats of copulative verbs (*es*, 'is'), auxiliary verbs (*hay*, 'there are') and modal verbs (*puede*, 'can'), which are the most frequently used, seem to abound in Spanish formal speech.

Word	Category	Repeats
1 *de* ('of')	preposition	253
2 *en* ('in')	preposition	113
3 *el* ('the')	article	82
4 *y* ('and')	conjunction	81
5 *que* ('that'/ 'who'/'which')	conjunction or relative	74
6 *a* ('to')	preposition	58
7 *la* ('the')	article	49
8 *un* ('a')	article	45
9 *no* ('no')	adverb	38
10 *del* ('of the')	preposition + article	28

Word	Category	Repeats
11 *se* ('himself'…, or mark of passive)	pronoun	26
12 *para* ('for')	preposition	23
13 *con* ('with')	preposition	21
14 *es* ('is')	verb	20
15 *los* ('the')	article	19
16 *o* ('or')	conjunction	18
17 *más* ('more')	adverb	17
18 *una* ('a')	article	14
19 *lo* ('it'/'the')	pronoun/article	12
20 *como* ('like'/'as')	conjunction	12

Table 2 – The twenty most frequent repeated words in formal spoken Spanish

Repetitions of prepositions (*en, de, con, para*) are frequent in our results from Spanish formal speech data; this contradicts the results for English (Biber et al., 1999: 1057). This could be partly explained by the fact that Spanish is a language poor in prepositions compared to English, so a preposition which is very repeated in Spanish could correspond to two or three prepositions in English, which would be repeated less frequently than in Spanish. In addition to that, English data show that subject pronouns are repeated more than other pronouns (e.g. accusatives), whereas our data reflect that subject pronouns (e.g. *yo,* 3 repeats) are less frequently repeated than other non-subject pronouns (especially, *se,* with 26 repeats; *me,* is repeated 4 times, but it is not quantitatively relevant). Two factors seem to explain these results: the first one, the fact that, unlike English, subject pronouns are not obligatory in Spanish; speakers tend to not pronounce them, so they are not repeated. The second factor could be related to the intrinsic characteristics of formal speech, especially in academic conferences: speakers generally use impersonal constructions (*haber que* + infinitive, passive with *se*) as an strategy to seem objective when introducing hypothesis, general trends, descriptions, etc., avoiding the use of the first-person pronouns. These results deserve a comparison with informal speech data in the future.

As it seems to occur in spoken English (Biber et al., 1999: 1059), repetitions tend to appear at the beginning of a noun, prepositional or verbal phrase: articles *el, la, los, un, una*, etc; prepositions like *de, en, con,* or *para;* pronouns like *se* of Spanish pronominal verbs or passive tense, etc. This would be related to the cognitive process of accessing the mental lexicon, because the repetition occurs before the lexical word (generally, nouns or verbs) is produced. Besides, repetitions also tend to appear between boundaries of phrases or utterances, such as conjunctions *y* and *o*. Regarding the conjunction or relative pronoun *que*, it may be a repetition before the head of the verbal phrase in some periphrasis, such as *tener que* ('have to') + infinitive, or a repetition at the beginning of a subordinate clause.

Even though the same item is mostly repeated only twice, speakers sometimes repeat the word more times. For instance, three repetitions of the same element seem to happen in many frequent functional words such as *el, de, en, y, que, no, se,* etc.; and four repetitions of the same word, which are less frequent, only seem to happen with very short words (*el, es, ni*). Actually, the length of the word seems to be important in repetitions. Most repeated words are short-length items: the ten most frequent repetitions are monosyllables (*de, en, y, el, que, a, un, la, no, del*). Although other repeated words are disyllables (*para, una, como, menos, este, pero*), they are less frequent. With regard to the repetitions of the same word more than twice, they are predominantly monosyllables: only the indefinite feminine article *una* is repeated three times in one utterance, whereas four times repetitions of disyllables do not occur at all.

Interestingly enough, as it tends to happen in English (Biber et al., 1999: 1057), in our findings the definite article (*el, la*) is more repeated than the indefinite article (*un, una*). Besides, in our data, repeats of the masculine forms are more frequent than the corresponding feminine articles: *el* is repeated more than *la*; *los,* more than *las* (9 repeats); *un*, more than *una*, etc. This trend of repetition seems to happen as well with some demonstratives: *ese,* 6 repeats, and *esa*, 3 repeats, but not with *este* (5 repeats) and *esta* (6 repeats). In spite of these results, the scarcity of our data doesn't make it possible to draw definitive conclusions. The reason why the masculine form is more frequently repeated may reside in the use of the masculine form to express the generic meaning in Spanish (besides the masculine meaning); that is, to express both the masculine and feminine with noun such as *el gato* ('the cat', male and female), *los amigos* ('the friends', both girls and boys), etc. Also, when considering the repetitions of *a* and *an*, Biber et al. (1999: 1059) suggest that the marked particle (*an*) is less frequent; in our data, we could apply this reasoning when looking at the gender of the repeated word: the feminine (marked) form of every particle seems to be less frequently used (and therefore, repeated) than the masculine counterpart (non-marked). This hypothesis would need to be revised with more data from informal speech.

3.3. Reformulations

Reformulations have been also called 'restarts' or 'false starts' (Moneglia, 2005: 27), 'retrace-and-repair sequences' or 'retractings' (Biber et al., 1999: 1062)[1]. Reformulations reflect a hesitation between two alternatives when the speaker is trying to express himself and repairs a segment of speech, which is abandoned and loses informational value (Moneglia, 2005: 27). The incompletion may be of just a word, a phrase or the beginning of an utterance. In the specialized literature (Shriberg, 1994: 7-8; Jurafsky and Martin, 2008: 418-419), the term 'reparandum' refers to this sequence, whereas the replacing fragment is called 'repair':

9) un / interesante debate / que también ~~vas~~ [/] vamos a tratar en el curso (enatte01)
reparandum **repair**

['an / interesting debate / which you are also going to [/] we are going to discuss]

10) esto facilita ~~la acción~~ [/] el éxito de la acción (enatbu03)
reparandum **repair**

['this facilitates the action [/] the success of the action']

In the C-ORAL-ROM and MAVIR corpus, restarts have been annotated in three levels: false starts at the word level, false starts such that the linguistic material is partially repeated and false starts such that the linguistic material is not repeated (Moneglia and Cresti, 2005: 27).

3.3.1. Corpus findings

As it happens in English language (Biber et al., 1999: 1062), in our formal spoken Spanish data reformulations appear along with other kinds of dysfluencies, especially pause fillers and fragmented words (sometimes, only a phoneme is deleted an it could be considered as a repetition as well):

11) un tipo de aproximación totalmente diferente // basada *en términos &eh [/] en herramientas automáticas* (mavir04)
['a kind of a completely different approximation // based in terms er [/] in automatic tools']

12) ¿dónde está *&espa [/] &eh* Italia?
['where is &spa [/] er Italy?'] (mavir04)

13) el porcentaje de reactiva *que &t [/] que [/] que &t [/] que tenga* en su consumo (enatbu02)
['the porcentage of reactive energy *which he &h [/] which he &h [/] which he has* in his consume']

There are frequent reformulations where a discourse marker is inserted, especially conversational markers:

14) en ese desajuste / *de [/] en definitiva / de* comprensión y comunicación / (enatbu03)
['in that maladjustment *of [/] in short / of* understanding and communication']

15) encuentran / &eh financiación *en el [/] pues en la* Generalitat // y / montan la empresa (mavir02)
['they find financing *from the [/] well from the* Catalan Parliament // and / they set up the business']

16) un módulo / de extracción / de resúmenes / &eh / *que [/] que bueno / que* genera un resultado mejor
['A summary / extraction / module / er / that [/] that well [/] that generates a better result'] (mavir06)

17) eso es el sistema genérico / *como [/] vamos* / el sistema de la arquitectura general
['That is the generic system / *how [/] well* / the system of the general architecture'] (mavir06)

4. Filled pauses

Filled pauses are "intermediary moments between heavier, information-laden articulatory gestures" (Joue and Collier, 2006: 146), and they tend "to consist primarily of vowels" because they "require less effort to articulate than consonants" (idem). These sounds in filled pauses have also been called 'pause fillers' or 'hesitators' (Biber et al., 1999: 1053). There exist three different types of filled pauses in formal Spanish: *eh, ah* and *mm*:

18) ellos saben / hacer su negocio / saben / *&eh* / cómo / puede mejorar su producto (enatbu03)
['They know / how to do their business / they know / *er* / how / their product can improve']

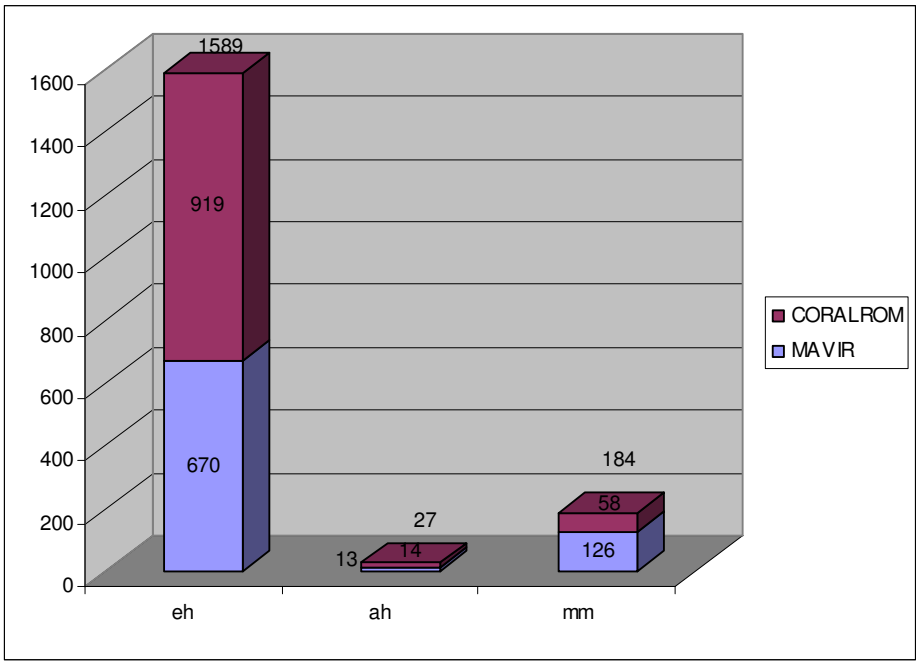
19) vamos ahora a ver / qué es lo que estamos ofreciendo a los clientes / de forma / &ah / directa / no ?
['now we are going to see / what we are providing to the clients / in a uh / direct / way / aren't we?'] (enatbu03)

20) es algo / *&mm* / muy interesante
['It is something *um* very interesting'] (enatps01)

Phonetically, it is frequent to add a nasal consonant sound to the vocalic pause filler, so *eh* may be pronounced as [e:] or [e:m], and *ah,* as [a:] or [a:m]. Actually, pause fillers tend to correspond to the more frequent (or characteristic) vocalic sounds of every language (Candea et al., 2005; Gil Fernández, 2007: 299)[2]. For instance, French speakers use the [œ:] sound; English speakers, the schwa vowel [ə:]; and the Spanish speakers, the [e] and [a] vowels (in this order of frequency, as shown in the graph 2). In fact, the [e] and [a] vowels are the most frequent vocalic sounds in Spanish (Moreno Sandoval et al., 2008: 1099). This feature of frequency seems to be related to the cognitive process involved in the production of speech when it is necessary to pause to think a word or a syntactic structure. There occurs a cognitive overload because of the action of two simultaneous processes: accessing the mental lexicon, and articulating a sound to keep the turn of speaking. As the retrieval of the searched word is the process which requires more cognitive load, in the process to keep the turn it seems logical to use the vocalic phoneme which is less marked or the consonant sound which requires the smaller articulatory effort (as is the case of [m:]).

4.1. Corpus findings

The analysis carried out on our corpus retrieved 1800 filled pauses in the 123412 words, which means a disfluency/word ratio of 0,0145. That is, on average, a pause filler appears every ten words in formal Spanish. Among the three types of pause fillers in Spanish, the most frequent item seems to be *eh* (1589 out of 1800), whereas *ah* appears with the lowest frequency (27), as can be seen in graph 2.



Graph 2 - Number of different types of filled pauses in the formal Spanish corpus.

Biber et al. (1999: 1054) remark that in English filled pauses appear at lesser or medial syntactic boundaries. In a future analysis, this could be compared to what happens in Spanish.

4.2. Speaker and subregister variation of filled pauses

Another aspect of filled pauses is the frequency variation due to every speaker's speaking style. In fact, although the majority of the texts from the preaching and academic speech present a low disfluency-per-word ratio, there is not a clear correlation between the subregister (academic conference, business, round tables or professional explanation) or the dialogic style (monologue, dialogue or conversation) and the frequency of filled pauses. The speaker's particular speaking style seems to be the factor which is more influential, but others may also affect the frequency of production of hesitation phenomena: the difficulty concerning the degree of complexity of the contents explained in every discourse, the conviction with which these are presented, the anxiety or calm induced by the communicative situation, etc.

4.3. Acoustic analysis of filled pauses

Modelling filled pauses in a way that they would not be erroneously recognized is an important issue in automatic recognition of spontaneous speech, as they tend to be included in the grammar module along with full lexical words (Jurafsky and Martin, 2008: 418). So *eh* may be mistaken with the conjunction *e* (used as a variant of *y* before a syllable beginning with *i-* or *hi-*) or *ah* may be interpreted as the preposition *a*; and vice versa.

In order to find out the acoustic cues which could be taken into account in the disambiguation of both words, some vocalic filled pauses have been analysed and compared to the homophone sounds pronounced by the same speaker. We have to point out that only a descriptive analysis has been carried out; in a deeper analysis, we should perform a normalization of the data. For the acoustic analysis, spectrograms and information about pitch and vocalic length were obtained with *Praat* software. First, we show the comparison between *eh* and the copulative conjunction *e*. The analysed sounds appeared in the following utterance:

> hay una persona / con / *&eh* / potencialidades / y con limitaciones / que / va / a filtrar / y a determinar / de forma directa / *e* inmediata / cuál es el alcance / del sistema de información que está manejando (enatbu03)

> ['there is a person / with / &eh / potentialities / and with limitations / who / is going / to filter / and to determine / in a direct / and immediate way / which is the scope / of the information system which he is using']

The fragments corresponding to the pause filler *eh* and the *e* sound for the copulative conjunction are shown on the spectrograms below (figs. 3 and 4 respectively). As can be seen, the length of the conjunction *e* is shorter (10 miliseconds) than the length of the *eh* (62 miliseconds) pronounced by one speaker in the same utterance. Secondly, in the sound of the copulative *e,* the coarticulation effect which affects the dynamic of the vocalic formants (moving to the position for the next sound) is an important feature to distinguish this sound from the pause filler *eh* (whose formants dynamic seems more stable). Besides, the *eh* pause filler appears between two silences, unlike the conjunction *e*, which coarticulated with the following vowel. Finally, regarding the pitch patterns, they seem to agree with the results obtained for English hesitated vowels by Daly-Kelly (1995: 1025): the value of the mean pitch of *eh* is lower (89.42 Hz, in fig. 3) than the mean pitch of *e* (168 Hz, in fig. 4).

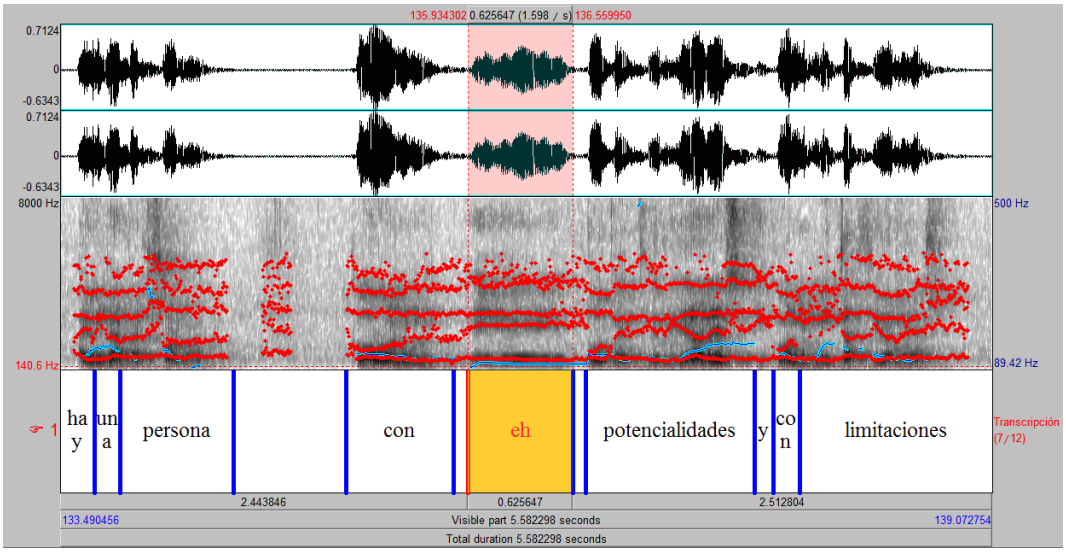


Fig. 3 – Acoustic analysis of the fragment from the utterance: *hay una persona / con / **&eh** / potencialidades / y con limitaciones* ['there is a person / with / &eh / potentialities / and with limitations']. The pitch contour appears in blue, and the mean value (89.42 Hz), in the right side.

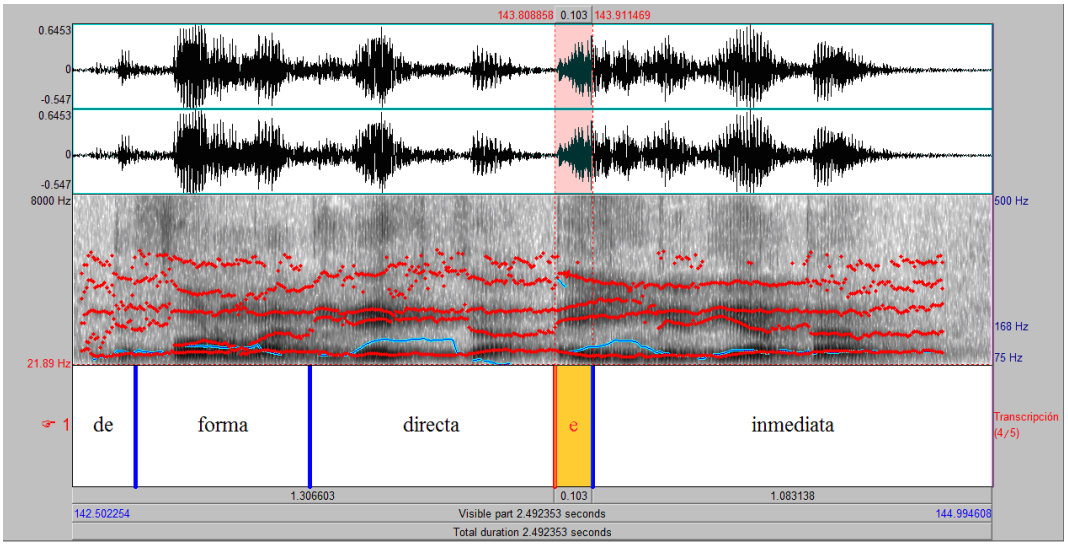


Fig. 4 - Acoustic analysis of the fragment from the utterance: *de forma directa / e inmediata* ['in a direct / and immediate way']. The pitch contour appears in blue, and the mean value (168 Hz), in the right side.

As far as the pause filler *ah* is concerned, we compared it with the homophone sound *a* (preposition). Spectrograms of two sentences from the same speaker (taken from C-ORAL-ROM corpus) show similar acoustic patterns as those for *eh* and *e*: the length of the pause filler *ah* (47 milliseconds, fig. 5) is longer than the sound corresponding to the preposition *a* (26 milliseconds, fig. 6), even if this sound does not coarticulate with another sound and appears between pauses. The dynamic of the corresponding vocalic formants behaves alike the *e* conjunction and *eh* pause filler. Regarding the F0 values, *a* and *ah* show a difference similar to that from *e* and *eh*: the mean pitch of the *a* vocalic sound is slightly higher (143.06 Hz, fig. 6) than the mean pitch of *ah* (81.19 Hz, fig. 5).

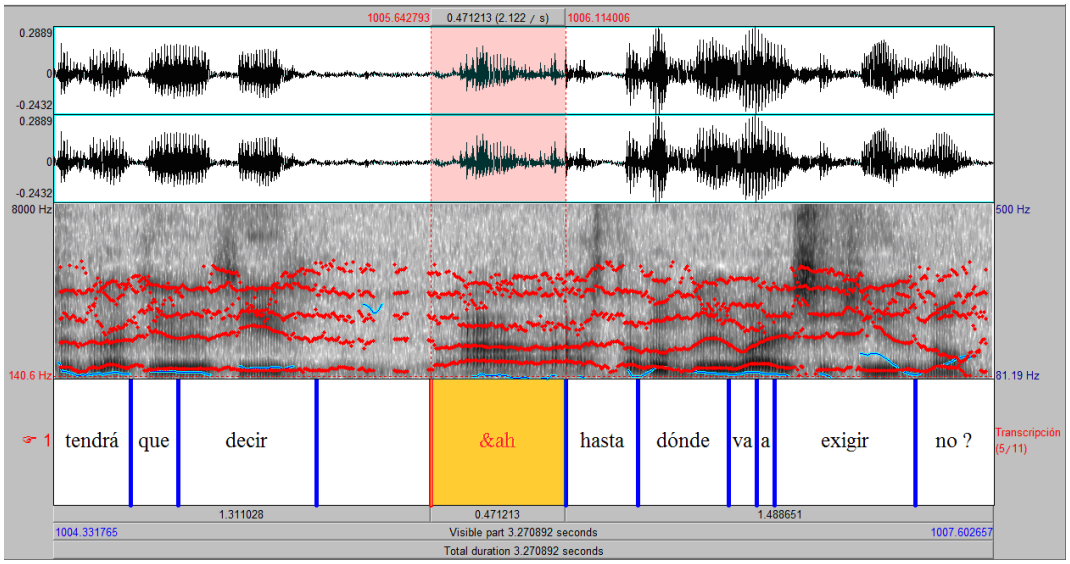


Fig. 5 – Acoustic analysis of the fragment of the utterance: *tendrá que decir / &ah / hasta dónde / va a exigir / no ?* ['(the administration) will have to say uh to what extent they are going to demand (the taxes), won't they?] (enatbu02). The pitch contour appears in blue, and the mean pitch of *ah* (81.19 Hz), in the right side of the figure.

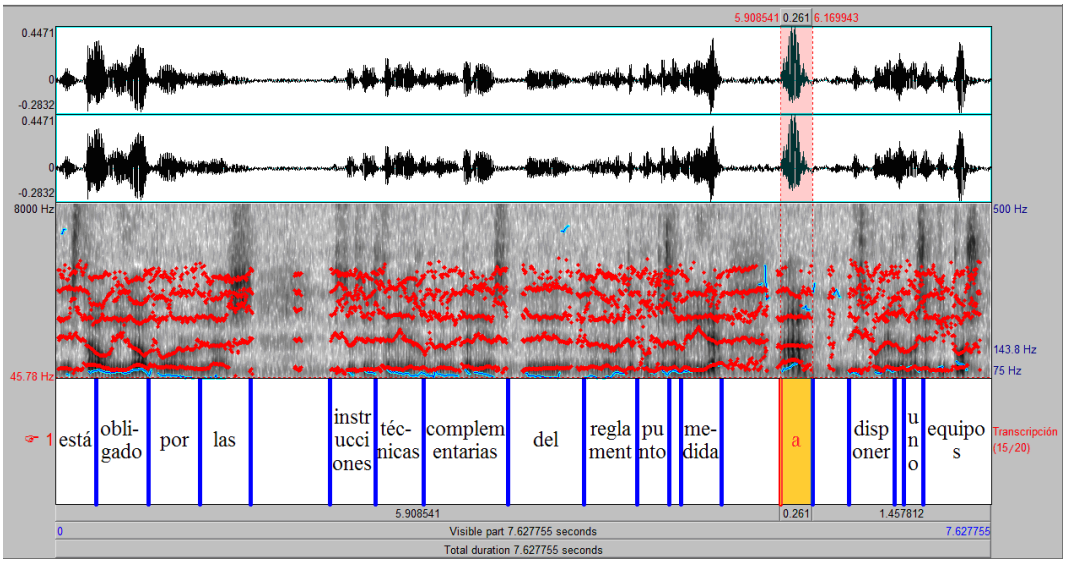


Fig. 6 – Acoustic analysis of the fragment of the utterance: *está obligado / por las / instrucciones técnicas complementarias / del / reglamento puntos de medida / a / disponer de unos equipos* (enatbu02). The pitch contour appears in blue, and the mean pitch of the *a* vocalic sound (143.06 Hz), in the right side of the figure.

Thus, as far as the vocalic filled pauses are concerned, empirical data from spontaneous formal Spanish seem to follow the results obtained for English by Daly-Kelly (1995: 1025) and Shriberg (1994: 175), and the research for other languages (Candea et al., 2005)[3]. Concerning the vocalic length, Shriberg states (1994: 175): "vowels in FPs have much longer durations than the same vowels in other contexts. Duration, then is a simple cue that could be used by speech recognition systems in discriminating vowels in filled pauses from the same vowels elsewhere." Regarding the pitch, the F0 values of the vocalic sounds from pause fillers (*ah, eh*) tend to be lower than those corresponding to the homophonous words (*a, e*), maybe due to a lesser articulatory effort. Curiously, the pitch value of the pause fillers is similar to that of digressions in Spanish, whose tone register is lower than the mean pitch value of the utterance where they appear (Gil Fernández, 2007: 395). So we could say *ah, mmm* and *eh* behave as a digressive element (but without lexical meaning) used to take time to think in the following word or concept. In any case, all of these acoustic features (length, vocalic formant dynamics, and prosody) should deserve a more detailed statistical research in the future.

5. Incomplete words

Spontaneous speech is filled with truncated words due to the way that orality is produced: speakers are thinking while they are speaking, and they often retrace and self-correct when a segment is already uttered.

21) trabajamos mucho para *&Sudamer* [/] Sudamérica (enatpe01)
['we work a lot for South &Am [/] South America']

22) cuando es un *&produ* [/] un proyecto importante // tiene una dimensión económica / fuerte (mavir02)
['when it is a &produ [/] an important Project // it has a strong economic dimension']

In our analysis, there appeared 670 truncated words (out of 123412), giving a ratio of 0,0054, which means that one incomplete word occurs every 200 words. Incomplete words are generally short. The most frequent fragmented words are monosyllabic, but they can also be disyllabic.

6. Discussion

Dysfluencies are not linguistic elements, but when we measure them we can discover some general trends of their nature, though they tend to be produced without a systematic behaviour. In fact, our corpus data from different disfluency phenomena show that frequencies are not related among them. For example, documents with high rate of repeats do not necessarily have a high rate of filled pauses.

As far as the dialogic style is concerned, the frequency of dysfluences in our data is not specifically related with monologues, dialogues or conversations, although we should compare our data with informal speech in a future work.

Regarding the type of text, in our data the number of dysfluences is neither related with any specific type of specific speech, excepting preaching and sermons. Indeed, preachers produce the lowest number of pause fillers, repeats and reformulations or fragmented words. A similar low frequency of dysfluencies is observed in political speech. In general trends, preachers and politicians are professional speakers and this fact is reflected in our data.

Actually, the number of speech dysfluencies seems to be determined mostly by each speaker's inherent speaking style. For instance, the speaker recorded while giving an explanation about some legal issues (file enatla02) produced a great number of repeats, reformulations and pause fillers, but it seems unrealistic to consider that the topic determine these phenomena. Indeed, in her study about speech dysfluencies in English, Shriberg (1994) differentiated between speakers who tend to repeat fragments of speech (repeaters) and those who tend to delete (deleters).

Besides, as the number of speakers in each type of text (business, academic conference, etc.) is not balanced in our data, we can not make reliable comparisons among them. The same thing happens with the dialogic styles: we need to analyse more data gathered from conversations and dialogues, because monologues predominate in our corpus.

Nevertheless, there exist some general trends in disfluency behaviour. Fragmented words are mainly monosyllabic, repeats are mostly function words, and pause fillers present specific acoustic features (most frequently vocalic sounds, longer, lower fundamental frequency, etc.).

7. Conclusions

A descriptive analysis of dysfluency phenomena in formal spoken Spanish has been explained. Although our results show that frequencies are not related, further research should be carried out to compare these data with transcriptions of speech from informal communicative situations (conversations, monologues…) to get a deeper understanding of speech production across registers. Apart from that, we need to gather more data from more speakers of every type of text (academic, business, law…) in order to find out if the frequency of pause fillers, repeats, reformulations and fragmented words could be in correlation with different text domains or dialogic styles (monologue, dialogue…). Apart from that, a thorough acoustic analysis of speech dysfluencies would be worthwhile for ASR tasks. These results from native data then could be compared to those obtained from other speech corpora (learner corpora, child language corpora, pathological speech corpora, etc.) so that a whole panorama of disfluency behaviour could be outlined in future.

8. Notes

[1] Some issues of classification and identification of retractings are exposed in Shriberg (1994: 14-15).
[2] Nevertheless, as the results obtained by Candea, Vasilescu and Adda-Decker (2005) show, Italian presents vocalic supports with [ə], "which is not part of the Italian vocalic system".
[3] These authors even claim that duration and pitch are universal criteria: "fillers are significantly longer than intra-lexical vocalic segments (…) and have a flat F0 contour".

9. References

9.1. Books and articles

Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Candea, M., Ioana Vasilescu, Martine Adda-Decker (2005). "Inter- and intra-language acoustic análisis of autonomous fillers". *Proceedings of DiSS'05, Disfluencies in Spontaneous Speech Workshop 2005*, Aix-en-Provence. Available at: http://hal.archives-ouvertes.fr/docs/00/32/19/14/PDF/candea-vasilescu-adda_diss05.pdf (accessed: 24 June 2009)

Cresti, E. and M. Moneglia (eds) (2005) *C-ORAL-ROM. Integrated Reference Corpora for Spoken Romance Languages.* Amsterdam: John Benjamins.

Daly-Kelly, N. A. (1995). "Linguistic and Acoustic Characteristics of Pause Intervals in Spontaneous Speech". *Proceedings of ESCA. EUROSPEECH'95. 4th European Conference on Speech Communication and Technology.* Madrid. September 1995.

Ferreira, F., E. F. Lau, and K. G. D. Bailey (2004) "Disfluencies, language comprehension, and Tree Adjoining Grammars". *Cognitive Sciences,* 28. pp. 721-749

Gil Fernández, J. (2007). *Fonética para profesores de español.* Madrid: Arco/Libros S.L.

González Ledesma, A., G. de la Madrid, M. Alcántara Plá, R. de la Torre, A. Moreno-Sandoval, (2004). "Orality and Difficulties in the Transcription of Spoken Corpora". *Proceedings of the Workshop on Compiling and Processing Spoken Language Corpora, LREC, 2004, Lisbon*. Available at: www.lllf.uam.es/ESP/publicaciones/oralid_final.pdf (accessed: 24 June 2009)

Huang, X., A. Acero, H.-W. Hon (2001). *Spoken Language Processing: A Guide to Theory, Algorithm and System Development*. NJ: Prentice Hall PTR.

Jurafsky, D., and J. H. Martin (2008). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition.* 2nd edition. Prentice Hall Series in Artificial Intelligence.

Joue, G., and N. Collier (2006). "Functional motivations for the sound patterns of English non-lexical interjections", in J. Davis, R. Jovanović Gorup, N. Stern (eds). *Advances in functional linguistics*. Studies in Functional and Structural Linguistics, 57. Amsterdam: John Benjamins. 143-161.

Moreno Sandoval, A., G. de la Madrid, M. Alcántara, A. González, J. M. Guirao, and R. de la Torre (2005) "The Spanish corpus". In E. Cresti and M. Moneglia (2005) (eds) *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages.* Studies in Corpus Linguistics, 15. John Benjamins.

Moreno Sandoval, A., D. T. Toledano, R. de la Torre, M. Garrote, and J. M. Guirao (2008). "Developing a Phonemic and Syllabic Frequency Inventory for Spontaneous Spoken Castilian Spanish and their Comparison to Text-Based Inventories". *Proceedings of the Language Resources and Evaluation Conference 2008*. Available at: www.lrec-conf.org/proceedings/lrec2008/pdf/283_paper.pdf (accessed: 24 June 2009)

Shriberg, E. (1994). *Preliminaries to a Theory of Speech Disfluencies.* U. Cal. Berkeley. Ph.D. Thesis.

Shriberg, E. (2005) "Spontaneous Speech: How People Really Talk and Why Engineers Should Care". *EUROSPEECH 2005 - INTERSPEECH 2005. Proceedings of the 9th European Conference on Speech Communication and Technology . 4-8 September, 2005.* 1781-1784. Lisbon, Portugal. Available at: www.speech.sri.com/papers/eurospeech2005-shriberg-keynote.pdf (accessed: 24 June 2009)

Toledano, D. T., A. Moreno, J. Colás, and J. Garrrido (2005). "Acoustic-phonetic decoding of different types of spontaneous speech in Spanish". *Proceedings of DiSS'05, Disfluencies in Spontaneous Speech Workshop 2005*, Aix-en-Provence. Available at: www.lllf.uam.es/ESP/publicaciones/Toledano-Moreno.pdf (accessed: 24 June 2009)

9.2. Software

Boersma, P. *Praat. Free software for Acoustic Analysis.* University of Amsterdam. Available at: www.fon.hum.uva.nl/praat/ (accessed: 24 June 2009)

9.3. Corpus

Cresti, E. and M. Moneglia (2005) (eds) *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages.* Studies in Corpus Linguistics, 15. John Benjamins.

MAVIR corpus. (2008) Distributed by the Fundación General de la Universidad Autónoma de Madrid (FGUAM).