Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 27-32.

A quantitative study of disfluencies in French broadcast interviews

Philippe Boula de Mareüil, Benoît Habert, Frédérique Bénard, Martine Adda-Decker, Claude Barras, Gilles Adda, Patrick Paroubek

LIMSI-CNRS, Orsay, France

Abstract

The reported study aims at increasing our understanding of spontaneous speech-related phenomena from sibling corpora of speech and orthographic transcriptions at various levels of elaboration. It makes use of 9 hours of French broadcast interview archives, involving 10 journalists and 10 personalities from political or civil society. First we considered press-oriented transcripts, where most of the so-called disfluencies are discarded. They were then aligned with automatic transcripts, by using the LIMSI speech recogniser. This facilitated the production of exact transcripts, where all audible phenomena in non-overlapping speech segments were transcribed manually. Four types of disfluencies were distinguished: discourse markers, filled pauses, repetitions and revisions, each of which accounts for about 2% of the corpus (8% in total). They were analysed by utterance", speaker and disfluency pattern types. Four question were raised. Where do disfluencies occur in the utterance? What is the influence of the speakers' status? And what are the most frequent disfuency patterns?

1. Introduction

This empirical study aims at expanding our knowledge of spontaneous speech-related phenomena from "sibling" resources of audio and written documents, such as available in TV archives or for parliament debates: instances of close, *bona fide* or trustworthy transcriptions, which are used for quotations, and even for legal purposes. The same content being both spoken and written, the comparison between either means of communication may then contribute to improve the modelling of so-called disfluencies (filled pauses, repetitions, etc.) in automatic speech recognition — to make transcriptions readily readable — and in speech processing, in applications like subtitling. Note that along this article we use the default term "disfluency" for the sake of convenience, regardless of its negative connotation which we do not assume. The mainstream terminology is questionable: if some phenomena may be described as production errors, as a "pollution" of the signal for automatic speech processing, others may help conceptualisation and contribute to fluency. Nonetheless, we dared not write "(dis)fluency" out of respect for the workshop title.

Written language and spoken language differ in many respects [6; 4; 9; 12; 3]. First, written language has a vocation for being persistent — it remains on a medium, unlike spoken words, which spring out in an ephemeral way and fly away. It is not tied to the same physiological constraints as spoken language, which uses the same organs as breathing. Whereas a lapse of time separates reading from writing, the simultaneity of spoken communication allows untimely interruptions and overlaps in the speech flow. In the former process, which is threefold, one can distinguish the writing activity (the dynamic nature of which may be traced in drafts), the written document (the result validated by the writer, which may also be anonymous) and the reading mechanism (a decoding step). In everyday conversation as in more supervised interviews, which are typically face to face, word tuning, talking and listening are synchronous: hence hesitations, repetitions and overlapped speech.

Some lexical and syntactic uses are well-established characteristics of written French: e.g. *car* ("for", "because"), inversion of verbs and subject pronouns in questions, while the drop of the *ne* in the discontinuous negation *ne… pas* is reminiscent of spoken language. These are only scattered examples; wide-coverage usage-based studies are better suited. Spoken corpora have known a considerable and unprecedented development for over a decade. The existence of large parallel spoken/written corpora now offers new prospects to answer the following question, which is crucial in linguistics: what characterises "spoken style" and "written style"? More generally, what is a "style"? — the "movement of the soul" according to Cicero, the "face of the soul" according to Seneca [5]. Labov [8] distinguishes between casual speech (in ordinary conversations) and careful speech (in an interview situation). Additional degrees of more or less colloquial speech could be considered. Yet, it is unnecessary here: our study definitely deals with careful speech, since it is based on TV shows in which journalists ask questions which politicians or representatives of civil society are bound to answer. We are thus faced with dialogues in which the interlocutors' roles are clearly established and accepted by both parts. They are at least partially prepared and they are public, since they are broadcast. The interaction is asymmetrical, since it is guided, and the situation is rather formal.

We have been interested in a corpus composed of audio files, and press-oriented transcripts, provided by the French Institut National de l'Audiovisuel (INA). We then added precise orthographic transcripts. The information that speech bears is incomparably richer than the one conveyed by the corresponding transcript. In particular, the quite limited typographic means we have at our disposal (punctuation, expressive capitalisation, orthographic stretching) can hardly express attitudes and emotional states. More generally, prosody and voice quality are badly or not at all indicated by typography. But other speech phenomena can be transcribed orthographically: filled pauses, splutters, slips of the tongue, and self-repairs (revisions), repetitions (of function words especially) and all these "little words" typical of spontaneous speech (discourse markers) such as *enfin*, *bon*, *ben*, *eh bien* ("well"), *donc*, *alors* ("so"), etc. In our corpus, they were reported and labelled in precise transcriptions, but are often missing in press-oriented transcriptions, which tend to render the message linear. This allows us to measure the distance between the two, which is the main goal of this quantitative study. Where do disfluencies occur in the "utterance" (intuitively defined by non-linguist transcribers)? What is the impact of the speakers' status? And what are the most frequent disfuency patterns? These are questions we will attempt to answer.

2. Corpus and transcription guidelines

This study makes use of 9 hours of *L'Heure de Vérité* ("The Hour of Truth"), a French TV show recorded a dozen years ago. In each one-hour show, a major personality from either political or civil society (e.g. charities) is interviewed by at most 3 journalists and a chairman, who is the same in the 9 shows. The journalists prepare their questions (most of them are to be expected), and the answers are not casual speech (some of them are "caned" answers, prepared answers to obvious questions). On the other hand, the chairman who leads the debates, makes sure that beforehand determined topics are stuck to and watches over the schedule, often interrupts the interviewee and the current interviewer. This configuration favours disfluencies, and speech overlaps are frequent. Only part of the numerous disfluencies reveals information about the planning problem of the speaker; the rest corresponds to a "struggle for speech" between interlocutors, even though journalists do not "jump in" haphazardly [16; 14].

For each show, we have both the audio and a press-oriented transcript (TPress). The latter is intended to be rather close to the audio while keeping to implicit conventions: it lies somewhere in between written text and exact transcript. As a matter of fact, most disfluencies are discarded. We consequently produced an exact transcription (TExact) for the audio data, with all audible phenomena, and in particular disfluencies. Speech recognition was particularly helpful because it precludes the unconscious filtering of disfluencies. It is often difficult to distinguish hesitations, for instance, from the pronunciation of a final schwa. With the help of the LIMSI system, first in its standard version, then in an "informed" version (i.e. taking TPress into account in the lexicon and the language model), we generated an automatic transcription (TReco) [1]. We took advantage of a modified version of Transcriber (http://sf.net/projects/trans/) to align time-codes and to display coloured mismatch zones between TPress and TReco (about 15% of the archive corpus, which were a priori made up of disfluencies) [2]. The coupling of the "informed" TReco and the *bona fide* TPress can then be regarded as a transcription draft.

In order to label disfluencies, we followed the LDC (http://www.ldc.upenn.edu/Projects/MDE/,) metadata annotation guidelines, adopted in the Rich Transcription evaluations conducted by NIST [11]. We chose these conventions because they fit some of our purposes (i.e. providing readable transcriptions), and represent the result of a vast discussion. LDC metadata annotations cover fillers (filled pauses, discourse markers, explicit editing terms, asides and parentheticals), edit disfluencies (repetition, revisions, restarts and complex disfluencies), and sentence-like units (statement, question, backchannel and incomplete sentence).

With some adaptations to French and simplifications, we distinguished and annotated filled pauses (FP), discourse markers (DM), repetitions (RP) and revisions (RV). Transcribed as *euh*, FPs were labelled automatically.

DMs may have either a simple filler role or a real discourse structuring function. According to the Geneva school terminology of discourse linguistics, DMs may be consecutive (e.g. *alors*, *donc* "so"), counter-argumentative (e.g. *mais* "but") or re-evaluative (e.g. *enfin* "well") [15]. We are aware that discourse markers may have several functions and mean different things; they do not have exactly the same disfluency status as filled pauses. But it is not always straightforward to interpret their precise role. The conjunction *et* ("and"), in particular, may also be used to structure the dialogue, to begin speaking, to avoid a stigmatised *euh*, to link two utterances or to prevent from being interrupted. The same happens with idioms such as *je crois que* ("I believe that"), which may be mere habits or verbal tics for some speakers, and which are difficult to consistently annotate.

RPs cover:

- repetitions of words (possibly truncated and/or interrupted by another speaker), where the left-most term(s) are marked up (e.g. *(RV le) le* "the the");
- emphatic repetitions, strengthening a statement;
- discontinuous repetitions (after parentheticals).

Note that only really "disfluent" repetitions were considered in the LDC metadata annotation guidelines, excluding emphatic or distant repetitions.

Finally, RVs involve word fragments, words or short chunks that are abandoned, without necessarily being corrected. Unfinished sentences, resulting from an interruption by another speaker, do not fall into this category. Nevertheless, albeit rare, complex cases exist, where RVs can include DMs, which in turn can include RPs and FPs. Disfluencies are particularly numerous at the borderline of overlapped speech sequences and within them. Only "clean" speech was so far systematically marked, because overlapped speech is quite difficult to handle. After discarding 24 minutes of overlapped speech, we have an amount of 7:18 of speech (88,056 words): 30 minutes by interviewee, 51 minutes for the chairman, 25 minutes for 3 recurrent journalists, 5 minutes for the other 6 journalists.

For disfluency annotation, a customised version of Transcriber was used. It allowed a quick annotation through contextual menus and a coloured display of the various disfluency types, similar to what LDC proposed for their own annotation scheme. The new disfluency annotation tags were embedded into the initial XML transcription files.

- Example of annotation:
  (RP ça veut dire) ça veut dire , par exemple , que quand on gagne le SMIC , **(DM eh ben)** bien évidemment non , on perdra pas *(RP de de)* de revenus parce qu' on peut pas , (FP euh) quand on gagne le SMIC , ***(RV perdre un)*** perdre son revenu.
- Press-oriented counterpart:
  Ça veut dire, par exemple, que quand on gagne le SMIC, bien évidemment non, on ne perdra pas de revenus parce que l'on ne peut pas, quand on gagne le SMIC, perdre son revenu.
- English translation:
  It means, for example, that when you earn minimum wage, of course not, you won't loose incomes because you can't, when you earn minimum wage.

3. Results

Our annotation enabled us to classify the words involved in disfluencies into DMs (2.5%), FPs (1.9%), RPs (2.3%) and RVs (2.2%). The proportions, computed with respect to the total number of words in the corpus, are relatively well balanced. It turns out in Table 1 that interviewers produce more filled pauses and repetitions, while interviewees produce more discourse markers and revisions. A test of comparison of two proportions reveals that each difference is significant with $\alpha = 0.05$:

$(p_1 - p_2) / \sqrt{p(1-p)(1/n_1 + 1/n_2)}$ > 1.96 in absolute value

where

n_1 is the number of words uttered by journalists,
n_2 is the number of words uttered by interviewees,
p_1 is the proportion of disfluent words uttered by journalists,
p_2 is the proportion of disfluent words uttered by interviewees,
$p = (n_1 p_1 + n_2 p_2) / (n_1 + n_2)$.

This difference may be due to the difficulties journalists meet, when they try to interrupt their interlocutor, while interviewees try to build a real argument.

3.1. Overall distribution

The occurrences of disfluencies can be studied as a function of the "utterance" length, the speaker's status (role, authoritativeness) and their context-dependency (whether some sequences of disfluencies are more prone to appear together, in a given order).

Table 1: Interviewees' and interviewers' disfluencies (DM = discourse marker; FP = filled pause; RP = repetition; RV = revision).

Speaker	words	%DM	%FP	%RP	%RV	%dis.
Brauman	8,174	1.5	1.2	2.2	3.6	8.4
de Robien	7,589	4.7	1.8	1.0	1.9	9.4
Delors	7,462	3.2	0.6	3.1	3.4	10.4
Voynet	7,177	4.0	2.5	1.8	1.7	10.0
Pasqua	5,385	1.4	0.9	1.6	1.5	5.4
Diouf	4,809	0.4	0.8	1.9	2.2	5.6
Brittan	4,806	4.4	4.3	5.8	3.5	18.0
Pinay	4,006	1.6	0.7	3.3	3.3	8.8
Chevènement	3,842	3.8	2.8	1.4	0.9	8.9
Lamassourre	2,729	0.6	1.1	0.9	0.6	3.2
Total inter-viewees	55,979	**2.8**	**1.6**	**2.1**	**2.2**	**9.1**
de Virieu	10,184	1.6	2.3	1.8	1.2	7.0
Duhamel	7,175	1.8	2.2	2.8	1.2	8.0
Colombani	4,818	2.2	1.3	2.3	2.9	8.7
du Roy	3,706	1.3	4.3	2.5	2.7	10.8
Diop	1,904	1.8	2.0	1.9	1.4	7.5
Tesson	1,270	2.4	2.0	2.5	5.8	12.7
Giesbert	886	5.6	2.2	4.6	3.1	16.0
Laffon	809	1.9	2.6	2.2	0.6	7.3
d'Orcival	743	3.2	1.1	2.2	2.4	8.9
English	622	3.7	3.5	4.2	2.3	13.7
Total inter-viewers	32,117	**2.0**	**2.4**	**2.4**	**1.9**	**8.7**
Total	88,056	**2.5**	**1.9**	**2.3**	**2.2**	**8.9**

For almost all speakers, the longer the utterance, the lower the percentage of disfluent words, as is apparent in Figure 1, where average rates for "utterances" of less than 12 words and more than 16 words are displayed. This observation is most likely due to the speech communication situation. As established by [17], disfluencies occur at the beginning rather than at the end of utterances (see Figure 2). We interpret this fact by a higher difficulty to start a rather than to continue a formulation.

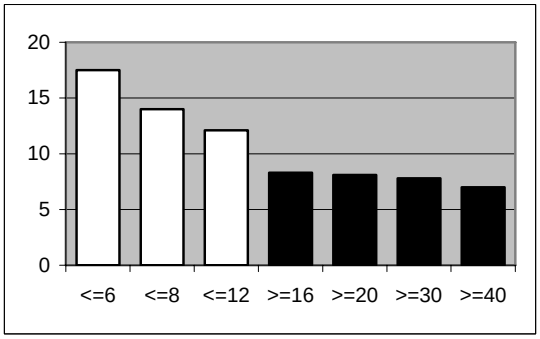

Figure 1: Percentage of disfluent words as a function of the "utterance" length (in words).

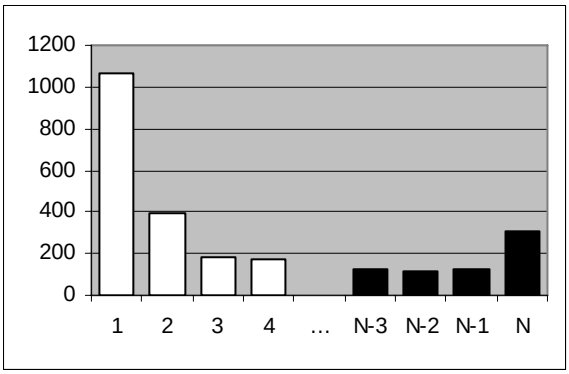

Figure 2: Distribution of disfluencies as a function of their position in the "utterance", from 1 (1st word) to N (last word).

In addition to the disfluency location, the content and context of appearance of DMs, FPs, RVs and RPs were investigated. The 1,568 FP occurrences all correspond to *euh* or its variants. As far as the other types of disfluencies are concerned (DMs, RPs and RVs), their distribution in terms of lexical items or idioms seems to follow Zipf's law (see Figure 3). Further details will be presented in the next subsections.

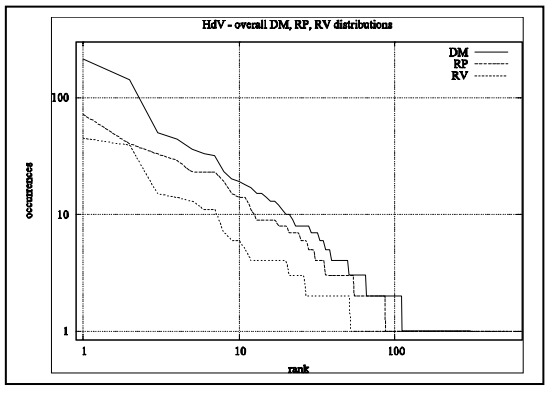

Figure 3: Zipf's distribution of lexical items or idioms involved in the DM, RP and RV categories.

3.2. Discourse markers

In DMs, which represent more than a quarter of all disfluencies, we find expected conjunctions, adverbs and

interjections (see Table 2): *et* ("and"), *alors* ("so"), etc. Either the former or the latter is the most frequent or the second most frequent DM for each speaker; but the other item in the leading pair is quite variable. For instance, the conjunction *mais* ("but") is the 4th most frequent DM, while it never appears in the leading pair of any speaker; and inversely, the phrase *je crois que* ("I believe that") is not so frequent, but appears in the heading pair of two speakers (see Tables 3 and 5). Interestingly, these two speakers are interviewees: not only are DMs speaker-specific, they also depend on the interviewer/interviewee position. Journalists are more inclined to use impersonal fillers (e.g. the interjection *hein* for Virieu). As for the interviewees who produce many DMs, they resort to a wide range of different expressions.

Table 2: DM morphosyntactic classes — *donc* ("therefore") is counted among conjunctions even though its distributional behaviour distinguishes it from other traditional conjunctions.

Category	%DM	Example
"conjunction"	27	*et, mais, donc*
adverb	25	*alors, enfin, d'ailleurs*
complex	17	*oui eh bien écoutez*
verb "phrase"	17	*je crois que*
interjection	12	*eh bien, ben, hein, bon*
pronoun	2	*moi*

To sum it up, we can distinguish three broad types of discourse markers: structuring (e.g. *alors*), position (e.g. *je crois que*) and interaction (e.g. *hein*). Each subtype represents about one third of all DMs, even though the latter — which shows that we are answering an interlocutor, that we try to convince him or that we agree with him — are somewhat fewer. As for the position subtype, its overuse by interviewees is particularly obvious with a speaker like Voynet.

Table 3: The 2 most frequent DMs for each speaker and their ratio within the DM class for the speaker.

Interviewees			Interviewers		
Brauman	*ben, et*	37%	Colombani	*alors, et*	33%
Brittan	*et, je crois que*	66%	de Virieu	*alors, hein*	39%
Chevènement	*et, hein*	29%	Diop	*alors, donc*	64%
Delors	*et, je pense que*	24%	Duhamel	*et, alors*	47%
de Robien	*et, eh bien*	49%	du Roy	*alors, et*	61%
Diouf	*et, moi*	50%	Tesson	*et, moi*	37%
Lamassourre	*et, je crois*	38%	English	*alors, et*	33%
Pasqua	*et, alors*	33%	Giesbert	*alors, bon*	43%
Pinay	*et, moi*	35%	Langelier	*et, alors*	62%
Voynet	*je crois que, alors*	45%	d'Orcival	*alors, et*	50%

3.3. Filled pauses

FPs can be found almost anywhere. More precisely, 35% of FPs occur at a sentence boundary indicated by a full stop (14%) or at a major phrase boundary indicated by a comma (21%), with respect to the TPress punctuation. For the remaining 957 FPs, Table 4 gives the distribution of the most frequent left and right contexts, considered independently. Even in the middle of a sentence, FPs frequently precede a determiner or a preposition; they rather follow a conjunction or a preposition. This asymmetry suggests that filled pauses (at least transcribed as *euh*) are avoided within noun phrases, especially between a determiner and a noun. In this situation, other mechanisms such as final lengthening or repetitions are preferred.

Table 4: FPs' most frequent left and right contexts; RVs' most frequent right contexts.

FP Left context		FP Right context		RV right context	
word	# (%)	word	# (%)	word	# (%)
que	40 (4.2)	*de*	53 (5.5)	*d'*	34 (4.7)
et	27 (2.8)	*la*	41 (4.3)	*l'*	30 (4.1)
pour	26 (2.7)	*des*	38 (4.0)	*la*	29 (4.0)
de	21 (2.2)	*les*	33 (3.4)	*vous*	25 (3.4)
avec	19 (2.0)	*l'*	26 (2.7)	*de*	23 (3.2)
à	13 (1.4)	*le*	23 (2.4)	*on*	21 (2.9)
qui	12 (1.3)	*un*	21 (2.2)	*le*	19 (2.6)

3.4. Repetitions and revisions

RPs and RVs exhibit some features in common: first, they both involve 1 or 2 words on average, and there is a high correlation (0.8) among speakers between their numbers of RP and RV occurrences. Speakers who produce many repetitions also tend to make many revisions. Second, if we look at the most frequent RPs and RVs, we can only see monosyllabic function words: *de* ("of", 72 RPs + 45 RVs), *le* ("the/him", 40 RPs + 39 RVs), etc. For all speakers, in the first two places and in the same order, we have very frequent French words. The form *le* is by far more often a determiner than a pronoun, even though nothing prevents a subject pronoun such as *je* ("I") from being one of the most repeated or revised words (see [7]). Most words are shared between RPs and RVs in Table 5, which is not surprising according to the following interpretation: in the process which consists of looking for words, a bootstrap word such as the masculine singular article *le* in French (or *the* pronounced as [i:] in English) may be repeated if it agrees grammatically with what follows, and may be corrected otherwise. The fact that there are more masculine nouns than feminine nouns in French (16k vs. 12k in the BDLEX dictionary [13]) does not seem to be sufficient to explain why *le* outweighs *la* in both RPs and RVs. By contrast, the conjunction *et* ("and") hardly lends itself to revisions, and we only find it among RPs.

Inspection of the right part of Table 4 shows that the most frequent words that follow RV-labeled words are *d'* ("of") and *l'* ("the"): precisely the shortened forms of the most frequently revised words. This means that the most frequent repairs are of the form *de d'*, before a word beginning with a vowel. We then have *la* (more frequent than *le*), which is in keeping with what we have just seen in the previous paragraph. Next, the presence of *vous* ("vous") or *on* ("we") is striking, since these personal pronouns are absent from Table 5: they really represent syntactic breaks, following abandoned phrases. The part-of-speech mismatch between the reperandum and the repair could be an objective criterion to label restarts, which we consider as RV subtypes. Levelt's [10:499] assertion, according to which "speakers tend to preserve the original syntax in the repair", deserves to be quantified.

Table 5: Most frequent words involved in disfluencies (DMs, RPs and RVs) — numbers of occurrences and percentages of the disfluency type they represent.

DM			RP			RV		
word	#	(%)	word	#	(%)	word	#	(%)
et	214	(9.8)	*de*	72	(4.3)	*de*	45	(2.2)
alors	141	(6.5)	*le*	40	(2.4)	*le*	39	(1.9)
je crois que	50	(2.3)	*et*	33	(2.0)	*à*	15	(0.7)
mais	44	(2.0)	*je*	29	(1.7)	*que*	14	(0.7)
donc	36	(1.6)	*un*	23	(1.4)	*la*	13	(0.6)
eh bien	33	(1.5)	*à*	23	(1.4)	*les*	11	(0.5)
hein	32	(1.5)	*les*	23	(1.4)	*je*	11	(0.5)

Content words may also be involved in repetitions and revisions, and are more affected by truncation phenomena than are function words. This is unsurprising, since they are far more often polysyllabic. In our annotation scheme, truncation phenomena are split into RPs and RVs, but they only represent 0.4% of our corpus.

3.5. Disfluency patterns

So far, we have considered what happens within and around disfluency markups. To finish with, let us regard disfluencies as single events. Apart from isolated disfluencies which dominate**,** the most frequent patterns of immediately consecutive disfluency labels are RV FP (53 occurrences), DM FP (47 occurrences), FP RV (46 occurrences), FP RP (45 occurrences) and DM RP (33 occurrences). Once more, a certain asymmetry is noteworthy in their order of appearance, chiefly between DM FP (47 occurrences) and FP DM (21 occurrences). More generally, the patterns DM + other disfluency appear twice as much as the patterns other disfluency + DM (98 vs. 50 occurrences). A possible explanation is that DMs, are often used to start a message (133 occurrences in position 1 vs. 65 occurrences in position 2, in disfluent zones which involve at least two disfluency labels) owing to their structuring and filler role. Once an interruption point is met, discourse markers are less employed in the editing term and the subsequent repair. Longer patterns exist, such as FP DM RP (5 occurrences), even though they are fewer: e.g. nous nous trouvons confrontés `(FP euh)` , **`(DM disons)`** , *`(RP à des)`* à des incohérences ("we are confronted to, let's say, to inconsistencies"). More data is required to extensively address disfluency patterns.

4. Conclusion and future work

Aligning press-oriented and automatic transcriptions diminishes the cost of an exact transcription, and enables the use of natural language processing (NLP) tools. It allowed us to examine a large speech corpus: several hours of French broadcast interviews. Four types of spontaneous speech-specific phenomena were analysed: discourse markers, filled pauses, repetitions and revisions (accounting for 8% of the corpus). They were sorted out by "utterance", speaker and pattern types.

Despite the size of our corpus, the conclusions we draw should be related to its genre, that of broadcast interviews, and would benefit from a comparison with conversational speech. With this end in view, the probabilities of discourse markers such as *je crois que*, *je pense que* ("I think that") were considered and compared to what is obtained in other corpora of fine-grained transcriptions in French — Broadcast News (3.6M words) and Telephone Conversational Speech (1M words). We notice that for interviewees we are close to the value estimated in conversational speech, whereas for journalists we are even below the value estimated in BN.

The corpus *bona fide* and fine-grained transcriptions were enriched with morphosyntactic tags. This information will be used in future work. Also, a prosodic study is planned, on function word final lengthening (phonemes longer than 300 ms) and speech rate in fixed expressions such as *je crois que* ("I believe that"): the latter is indeed quite frequent in almost all interviewees, but it seems to have been labelled as a discourse marker only when it is pronounced quickly.

In the near future, we also plan to study the relationship between disfluencies and turn taking, their position within sentence-like units (SUs) as well as the influence that struggle for speech has on disfluencies. Finally, this type of analysis would arguably improve by being related to the study of eye movements and body gestures, since we have video recordings at our disposal.

5. Acknowledgements

We are indebted to the INA Research and Experimentation Directorate (http://www.ina.fr/) for the *L'Heure de vérité* corpus (audio and video files) and its *bona fide* transcription. INA plays the role of public archive for audio and video resources in France.

6. References

[1] Adda-Decker, Martine *et al*. 2003. A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models. *Proc. DISS*, 5–8 September 2003, Göteborg, pp. 67–70.

[2] Barras, Claude *et al*. 2004. Automatic audio and manual transcripts alignment, time-code transfer and selection of exact transcripts. *Proc. LREC*, 24–30 May 2004, Lisbon, pp. 877–880.

[3] Biber, Douglas. 1995. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge, Cambridge University Press.

[4] Blanche-Benveniste Claire *et al.* 1990. *Le français parlé, études grammaticales*. Paris, Éditions du CNRS.

[5] Fónagy, Ivan. 1983. *La vive voix. Essais de psycho-phonétique*. Paris, Payot.

[6] Hagège, Claude. 1985. *L'homme de parole*. Paris, Fayard.

[7] Henry, Sandrine & Bertille Pallaud. 2003. Word fragments and repeats in spontaneous spoken French. *Proc. DISS*, 5–8 September 2003, Göteborg, pp. 77–80.

[8] Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, University of Pennsylvania Press.

[9] Léon, Pierre. 1993. *Précis de phonostylistique.* Paris, Fernand Nathan.

[10] Levelt, Willem J. M. 1989. *Speaking. From Intention to Articulation.* Cambridge, Massachusetts: MIT Press.

[11] Liu, Yang *et al*. 2005. Structural Metadata Research in the EARS Program, *Proc. IEEE ICASSP*. 18–23 March 2005, Philadelphia, pp. 957–960.

[12] Morel, Marie-Annick & Laurent Danon-Boileau. 1998. *Grammaire de l'intonation. L'exemple du francais*, Paris, Éditions Ophrys.

[13] Pérennou, Guy & Martine de Calmès. 1987. *BDLEX, base de données lexicales du français écrit et parlé*. Toulouse, Travaux du laboratoire CERFIA.

[14] Plauche, Madeleine & Elizabeth Shriberg. 1999. Data-Driven Subclassification of Disfluent Repetitions Based

on Prosodic Features. *Proc. ICPhS*, 1–7 August 1999, San Francisco, vol. 2, pp. 1513–1516.

[15] Roulet, Eddy *et al*. 1991. *L'articulation du discours en français contemporain*. Berne, Peter Lang.

[16] Shriberg, Elizabeth. 1994. *Preliminaries to a theory of speech disfluencies*. Ph.D. thesis, University of Berkeley, California.

[17] Shriberg, Elizabeth. 2001. To "Errrr" is Human: Ecology and Acoustics of Speech Disfluencies. *Journal of the International Phonetic Association* **31**(1), pp. 153–169.