Annotation and analysis of overlapping speech in political interviews

M. Adda-Decker[1], C. Barras[1,2], G. Adda[1], P. Paroubek[1], P. Boula de Mareüil[1], B. Habert[3]

[1]LIMSI-CNRS, [2] Univ Paris-Sud, [3] ICAR UMR 5191 & ENS LSH
LIMSI-CNRS BP 133, 91403 Orsay cedex FRANCE,
Univ Paris-Sud 91405 Orsay cedex FRANCE,
ICAR, 15, Parvis René Descartes BP 7000 69342 LYON cedex 07 FRANCE
Email: {madda,barras,gadda,paroubek,mareuil}@limsi.fr, benoit.habert@ens-lsh.fr

Abstract

Looking for a better understanding of spontaneous speech-related phenomena and to improve automatic speech recognition (ASR), we present here a study on the relationship between the occurrence of overlapping speech segments and disfluencies (filled pauses, repetitions, revisions) in political interviews. First we present our data, and our overlap annotation scheme. We detail our choice of overlapping tags and our definition of disfluencies; the observed ratios of the different overlapping tags are examined, as well as their correlation with of the speaker role and propose two measures to characterise speakers' interacting attitude: the attack/resist ratio and the attack density. We then study the relationship between the overlapping speech segments and the disfluencies in our corpus, before concluding on the perspectives that our experiments offer.

1. Introduction

Oral communication between several actors can be simplistically viewed as a sequence of single speaker turns. However overlapping speech, i.e. speech portions simultaneously involving more than one speaker, is very common in natural communication (Delmonte, 2005). Overlaps in speech may entail disfluencies (hesitations, repetitons, restarts) and are likely to contribute to speaker turn regulation. They definitely cause problems for automatic processing (Shriberg et al., 2001). This contribution focuses on overlapping speech phenomena in TV political interviews, where overlaps happen to occur, even though their overall ratio remains relatively low as compared to rates reported for conversational or meeting speech (Shriberg et al., 2001). An interview is an asymmetric interaction between speakers who have different statuses and complementary roles. It also differs from peer conversation on the competitive/cooperative axis (Grice, 1975; Schegloff, 2000). Symmetry and competitiveness increase the degree of interactivity, hence speech turn and overlap rates. Cooperativeness may consist of helping one's interlocutor speak, which is usually not necessary in political interviews. In broadcast political interviews, journalists often have to defend viewpoints, since they also speak to an audience (Bell, 1984). They happen to contradict and interrupt interviewees, thus favouring overlaps. The genre of the material studied here is probably less interactive than casual conversations, nonetheless it includes an interesting amount of speech overlaps. Whereas such speech overlaps have long been of major interest to Conversation Analysis (Schegloff et al., 1977), they only tend to become a hot topic for ASR. Speech overlaps are natural in spoken conversation, and the simplistic view of a sequence of separate speaker turns has to be improved in state-of-the-art ASR systems. As roles in these interviews are asymmetrical, it may be enlightening to analyse overlapping speech and disfluency measures with respect to the speaker's role in the communication context.

The questions addressed are the following: how to annotate overlapping speech for both automatic processing and more linguistically-oriented studies? Are there different types of overlapping speech and if so, can they be qualified as more or less intrusive. Do speaker roles impact overlap types? A further point of interest concerns the link between overlapping speech and disfluency occurrences. Do overlap types impact disfluency rates and types? Do disfluency rates significantly differ in active vs passive roles in the overlap situation?

In section 2. we present the speech corpus, before addressing the overlap segmentation and annotation issues in sections 3. and 4.. Section 5. presents results on speech overlaps and disfluencies and analyses them along different axes: intrusive vs non-intrusive overlaps, passive vs active overlaps. Finally section 6. gives a summary of the results obtained.

2. Corpus

The present work deals with broadcast interviews corresponding to a rather careful speech style. The corpus studied here is composed of 8 one-hour TV shows during which a major figure from either political or civil society is interviewed by 3 journalists and a chairman. The chairman watches over the schedule and may interrupt interviewees or interviewers to have them stick to previously determined topics and timing. This configuration favours speech overlaps and disfluencies among interlocutors.

The audio corpus benefits from exact orthographic transcripts including specific annotations concerning discourse markers (DM) and disfluencies, namely filled pauses (FP), repetitions (RP) and revisions (RV) (Boula de Mareüil et

al, 2005) in line with the LDC annotation guidelines[1] and the French GARS conventions (Blanche-Benveniste, 1990). The Transcriber software (Barras et al., 2001)[2] has been customised to facilitate and speed up the manual annotation process through contextual menus and a coloured display of the various disfluency and overlap types. Specific overlap and disfluency annotation tags are embedded into the transcription files.

3. Overlap segmentation

In telephone conversations or meetings, overlaps are very frequent (with more than 10% of overlapped words (Shriberg et al., 2001)). It may hence be convenient to transcribe each speaker as a separate synchronised stream. On the opposite, broadcast news are very controlled, include a high proportion of monologues,and feature a very small amount of overlaps. For automatic speech transcription, it is usual to partition broadcast news data as a sequence of individual speaker turns, setting aside overlapping segments with a precise temporal anchoring (Barras et al., 2001), as their processing remains beyond the scope of state-of-the-art systems. Our corpus of political interviews is less controlled than broadcast news, and a crude segmentation of overlapping segments has the drawback of breaking the interaction stream. We thus chose to preserve the interaction structure, and to relax temporal synchronisation constraints at turn boundaries in the case of overlaps.
An overlap occurs when a first speaker (primary) keeps talking while a second speaker comes in. The more complex situation of more than two persons speaking simultaneously appeared to be negligible in our data. For overlap segmentation and transcription, two situations have been distinguished:

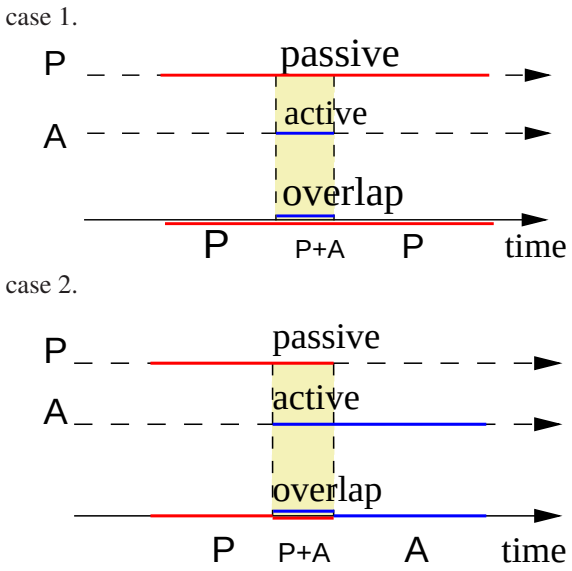
1. The overlap does not entail a speaker change: the primary speaker remains the same after the overlap.

2. The overlap results in a speaker change: the primary speaker stops and the second speaker becomes the primary speaker of the new turn.

Fig. 1 gives a schematic representation of these two cases, with the first two lines representing each speaker A and P as independent streams (along the time axis), and the last line represents the projection of all the speakers on a unique stream, with the overlap region clearly delimited on the time axis. The overlap speech (generated by simultaneous speakers) is marked as P+A. Overlap occurs when the primary speaker (P) is joined by the overlapping speaker (A), who is the active one with respect to the overlap situation, whereas the primary speaker has a passive role. For the second case the active overlap speaker (A) turns out to become the primary speaker (P') after the overlap. Fig. 2 gives examples of segmentation and transcription in both

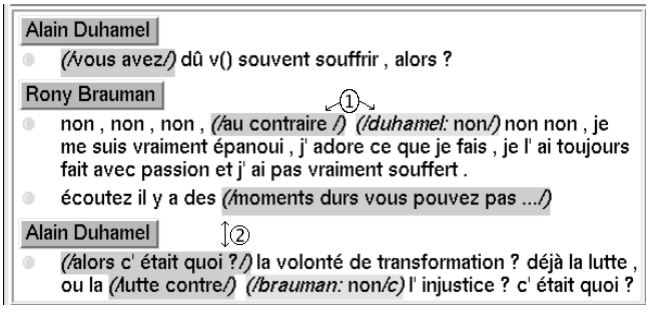
[1] http://www.ldc.upenn.edu/Projects/MDE/
[2] http://sf.net/projects/trans/

Figure 1: Structure of the two overlapping cases.
①: no speaker change ; ②: speaker change.



cases. Overlapped words (P+A) are displayed sequentially (P,A) on a coloured background in the transcription. For case 1, the first portion of words corresponds to the default (primary) speaker, followed by the portion from the overlapping speaker, explicitly named here, as there is no primary speaker change. Case 2 features overlapped words in sequential speaker-dependent streams, the coloured background marking lose overlap boundaries. The option of coloring overlapped words without adding precise time stamps is very convenient, as an accurate localisation of overlap starts and ends is far from being obvious.

Figure 2: Overlap segmentation examples (cases ① and ②) in the customised Transcriber annotation editor.



4. Overlap tagset and annotation

During a preliminary phase, different overlap annotation options were explored before deciding on the annotation scheme developed hereafter. Several dimensions were explored with a special interest in the correlation between overlap and disfluency production.

The preliminary phase consisted in labelling each overlapping segment (excluding turn boundaries) using three independent features: *elaborated contribution* (yes/no), mainly related to the length of the overlap; *agreement* with the primary speaker (yes/no/other); and *interruption* (yes/no).
For this phase, a single show was labelled independently by four annotators. Few instructions were given, since the goal was mainly to get an empirical view of the phenomenon. For the *elaborated* and *interruptive* labels, a full consensus between all the annotators was reached in over 80% of the cases; for the *agreement* label, the full consensus was much less frequent. However a majority of at least three (out of four) annotators was reached in over 90% of the cases. Overall, 5% of overlaps were labelled as interruptive, 17% as elaborated, 10% as an agreement and 2% as showing a disagreement with the speaker. Considering the labelling results achieved with our three features, some broad overlap categories emerged:

- standard back-channel ("hmm") was the most frequent, characterised as uninterruptive, not elaborated, most often neutral and a few times felt as an agreement,
- some overlaps provided precision or answer, labelled as uninterruptive, elaborated and mostly neutral or in agreement,
- interruptive overlaps formed a third, less frequent category.

In a second phase, it was preferred to use a set of predefined mutually exclusive categories. The turn boundaries were included in the annotation process. Speech overlaps were annotated with 4 tags: back-channel (**bck**), turn stealing (**tst**), anticipated turn taking (**att**), and complementary (**cmp**). Back-channels like "hmm"s indicate that the listener follows the speaker, understand him/her, agree with him/her (Cerrato and D'Imperio, 2003); they barely disturb the main speaker. On the opposite, turn stealings clearly interrupt the main speaker, even though the attempt may fail as with any other speech act. Anticipated turn taking corresponds to the case where the incoming speaker seems to perceive cues indicating that the main speaker has finished (requested information delivered, phrase or clause boundary, falling pitch, etc.). Finally, the complementary (**cmp**) tag label was introduced for overlaps which aim at complementing the main speaker's utterance: a possibly paraphrased repetition of the primary speaker's statement, an explicit agreement or disagreement, a short anticipated answer, a precision forwarded or required, not only on the content but also on the form of the exchange (schedule, approached topic), a witty remark or the continuation of the utterance. This complementary label, contrary to the turn stealing one, is assigned to self-sufficient comments or utterances: the entering speaker does not take the floor to develop an argument. This type of overlap may be favoured by the situational context: beyond the speakers actively involved in the show, someone may wish to provide additional information to the audience. Fig. 3 shows an example for each overlap tag.

Figure 3: *Examples of the different overlap types, producing case 1 (***bck**, **cmp***) and case 2 (***tst**, **att***) overlaps.*

bck: backchannel
A: it is simply /the fact/ /B: hmm/ that...

cmp: complementary
A: I have a last question /about/ /B: very short/ about your...

tst: turn stealing
A: and in /this case.../
B: /I want to/ come back...

att: anticipated turn taking
A: and this leads to humanitarian /action?/
B: /well I/ think

Several options may be taken for labeling speech overlaps. Multiple annotators were felt necessary for this rather subjective, yet time-consuming task. A preliminary annotation with 5 annotators was carried out on two shows. The reference annotation then resulted from a consensus of individual annotations followed by a final adjudication phase for disputed labels. Table 1 presents the label distribution of the different annotators, for each category in the final annotation. It confirms the intermediate nature of the complementary label, and shows a rather high confusion value of 24% between **att** and **tst**. Compared to the reference, the 5 annotations show an inter-annotation agreement Kappa measure (Carletta, 1996) between 0.7 and 0.8, which decreases to a 0.6–0.7 interval when only a **tst** vs **att** binary choice is considered. Each of the remaining six shows was

Table 1: Overlap label distribution from 5 annotators relative to the final annotation of one show.

		annotator labels (%)			
label	count	**bck**	**cmp**	**tst**	**att**
bck	63	91.1	8.0	1.0	0.0
cmp	50	9.2	75.8	15.0	0.0
tst	107	0.4	3.6	89.2	6.8
att	26	0.0	0.0	**24.0**	76.0

processed by one annotator and passed over to a colleague for verification. Corrections involved between 3% and 6% of the labels. This can be taken as an estimate of the residual disagreement rate, but it also reflects the problem of assigning a unique label when two categories apply.
The work of segmenting and annotating overlapping speech highlighted that differences between overlap tags happen

to be subtle and may give rise to diverging interpretations. For example, some **att** events can be seen as **tst**. Even "hmm"s may have additional communicative functions of complementing or signaling that someone is eager to jump in: during a long lasting speaker turn, progressive transitions from back-channeling "hmm"s to complementary or turn stealing items are common. Yet, the established annotation scheme and the resulting annotations enables us to study the distribution of overlaps, their types as well as their link with disfluencies and speaker roles.

5. Analysis of overlapping speech

Although speech overlaps occur frequently (on average 3-4 overlaps per minute), their global duration remains low (below 5% of the data). Overlaps, averaging 2.5 words per segment, are very short compared to single speaker turns (30 words on average). In the following, we first introduce some keys to the overlap analysis, including the concepts of homogeneous speech regions, intrusive and non-intrusive overlaps, active and passive overlap speakers, before quantifying speech production in various overlap conditions. Such measures allow us to check e.g. whether passive speakers tend to slow down or not, whether they "resist" when facing an incoming, competing speaker. A more detailed analysis then addresses various disfluency phenomena and discourse markers. An increase of disfluencies on behalf of the primary speaker may correlate with intrusive overlaps, whereas non-intrusive overlaps are supposed to keep the contribution of disfluent speech close to average rates measured on non-overlapping speech.

5.1. Homogeneous speech regions

In the manual audio transcripts segment boundaries generally occur either at phrase boundaries or at speaker changes. Speech overlaps are indicated by appropriate XML tags in the transcripts and may or not entail a speaker change (case 1 vs case 2 in Fig. 1). To allow global measures of overlapping speech in the data, we consider the projection of the overlapping speaker turns on the time axis (see Fig. 1), as derived from the overlap segmentation and annotation. We then define as **H-region** a maximum length segment keeping **homogeneous** speaker characteristics. Thus a H-region spans one or several contiguous transcription segments corresponding to fixed speaker conditions. If there were no overlaps, the number of H-regions would equal the number of speaker turns. Table 2 gives a synthetic overview of the corpus in terms of H-regions. Single speaker H-regions represent 65% of the H-regions with slightly more than 95% of the words (counting the primary stream only: words from the secondary stream are not included). Mono-speaker H-regions reach an average of 30 words. The remaining 35% of H-regions correspond to short duration overlapping speech.

Table 2: Number of H-regions (with rate in H-region set), of words and average region-length in single-speaker and overlapping speech (counts on primary speakers only).

set	#H-reg (%H-reg)	#wrd	av.lg.
all	4k (100)	83.0k	20.7
mono-speaker	2.6k (65)	79.3k (95%)	30.0
overlap	1.4k (35)	3.7k (5%)	2.7

5.2. Intrusive/non-intrusive overlaps

The distinction between intrusive and non-intrusive overlaps may rely on prosody, but also on their localisation with respect to potential segmentation points (sentence, clause or phrase ends for instance). Disfluencies on behalf of the passive primary speaker may also contribute to qualify speech overlaps as intrusive.

The back-channel (**bck**) label typically corresponds to a very short non-intrusive overlap, meant to encourage a fluid interaction. Complementary (**cmp**) overlaps do not aim at a speaker turn, and may be felt as non-intrusive by their author. However, both their length and informational content are likely to disturb the primary speaker and thus to generate disfluencies in the speech flow. They are hence considered as intrusive here. Turn-stealing (**tst**) is clearly intrusive. Anticipated turn-taking (**att**) is a non-intrusive form of overlap, occuring slightly in advance of a commonly agreed speaker change. The message of the primary speaker, even though not yet completely uttered, has already been received by the audience.

5.3. Active/passive overlaps

Overlapping speech can be analysed by comparing productions from active overlap speakers to those of the primary speaker who is considered as passive with respect to the overlap situation. Table 3 shows overlapping segment counts, their frequency, word counts and mean length for intrusive (**cmp**, **tst**) and non-intrusive (**bck**, **att**) overlaps. Overall, non-intrusive overlaps are significantly shorter than intrusive overlaps. Figures for active and passive overlapping speech are quite comparable. The highest production is measured for active turn stealings, if not in number of occurrences, at least in number of words. In this challenging situation, active speakers tend to be speedier than passive competitors.

5.3.1. Attack/resist ratio

Before investigating whether disfluency productions significantly differ in active (overlapping) vs passive (overlapped) roles, we dedicate some lines to examine speakers' oral productions in both situations.

For each speaker, we computed the number of words he/she produces in a mono-speaker condition M, as a primary, i.e. passively overlapped speaker P, and as a secondary i.e. actively overlapping speaker A. A then measures the

Table 3: Overlap segment counts, frequency, word counts and mean length for passive (P) and active (A) roles, for **bck**, **cmp**, **tst** and **att**.

category		segment count	freq. /min.	words #	words %	mean length
bck	P	461	1.2	719	0.8	1.6
	A			550	0.6	1.2
att	P	168	0.4	345	0.4	2.1
	A			391	0.5	2.3
cmp	P	278	0.7	955	1.1	3.4
	A			974	1.1	3.5
tst	P	438	1.1	1447	1.7	3.3
	A			1658	1.9	3.8

speaker's interactivity or aggressivity, whereas P measures his/her resistance towards interruption. From the balance between A and P an attack/resist ratio can be defined as follows:

$$R = \frac{A - P}{A + P} \quad (1)$$

This ratio, when positive, indicates more active overlap than resistance to overlapping speakers. Negative values correspond to speakers who tend to keep the floor rather than to jump in. This ratio can be complemented with an attack density measure, defined as the ratio of active overlaps and all the words uttered by the speaker:

$$D = (\frac{100 \times A}{M + P + A}) \quad (2)$$

D measures the overall frequency of active overlaps for a given speaker. Table 4 displays R and D values, first computed overall, next separating journalists and interviewees in order to check whether the asymmetry of their respective roles entails an asymmetry in the measured R and D values. Measures per speaker are added: for interviewees we indicate whether they are French politicians (PF), from civil society (CF), or whether they are non-native French speakers (PI: international politicians). The average R including all speakers is close to 0, which means that the number of words uttered either as primary or secondary speakers does not vary much. R ratios and D densities differ between journalists and interviewees. For journalists, positive R values reflect a higher proportion of active overlaps, reflecting their role. Interviewees are characterised by negative R ratios and relatively lower overlap densities. As for the R ratio, the D density is highest for journalists, who are in charge of the successful progress of the interview, whereas it remains close to zero for interviewees, especially for non-native speakers (e.g. IntPI1). Interviewees are relatively passive with respect to the interview timing and program.

The introduced R and D measures thus highlight differences between journalists and interviewees reflecting their roles in the ongoing interview. Restricting R and D to the subset of intrusive overlap segments (**tst**, **cmp**) results in identical tendencies (not shown) for the two speaker classes.

Table 4: Attack/resist ratio R and attack density D (overall, for journalists, interviewees and detailed per speaker).

set	R	D
all	**0.0**	**4.0**
journal.	**0.3**	**8.0**
interv.	**-0.3**	**2.2**
Chairman	**0.2**	**6.6**
journal.	R	D
Journ1	0.3	10.8
Journ2	0.5	6.7
Journ3	0.1	4.3

interv.	R	D
IntPF0	-0.1	2.4
IntPF1	0.0	3.6
IntPF2	-0.2	3.4
IntPF3	-0.6	1.0
IntCF1	-0.4	2.9
IntCF2	-0.7	1.2
IntPI1	-0.4	0.7
IntPI2	-0.1	2.3

5.4. Disfluencies

Overlapping speech tends to increase disfluency rates as compared to overall rates measured in single speaker regions. As seen earlier overlap regions are essentially of short duration with about three words on average.

The first three lines of Table 5 show disfluency and discourse marker (DM) rates in mono-speaker and overlapping speech. Disfluencies comprise filled pauses (FP), restarts and revisions (RV) as well as repetitions (RP). Measures exhibit an important increase of disfluencies in overlap regions. More disfluencies are produced by the active speakers than by the passive (primary) speakers. Higher rates of repetitions and discourse markers are measured on active speaker segments in comparison to their passive counterparts. These high rates should not be explained by overlapping speech alone: concerning non-overlapping regions, it has been shown that disfluency rates globally follow a "declension line" over time with high figures for repetitions and discourse markers at the very beginning of speech turns, followed by quickly dropping rates for more turn-internal positions (Adda et al., 2007). Overlaps often correspond to speaker turns (**tst**, **att**): active overlap speakers tend to be in a position which favours disfluencies even for non-overlapping speech. On the contrary, passive overlap speakers of **att** overlaps are in turn-end position. Active overlap speakers produce more discourse markers than passive overlap speakers, suggesting that discourse markers contribute to speaker turn negotiation or turn starting.

The following lines of Table 5 give separate figures both for intrusive vs non-intrusive and passive vs active conditions. Very few disfluencies are measured for the non-intrusive condition. Concerning DMs, the highest rate are achieved by non-intrusive overlaps (**att** here).

To get a more statistically-informed view of the presented disfluency results, Fig. 4 makes use of a box-and-whiskers representation. To do so, we considered the population of speakers, while the points ($\circ$ $\bullet$ $\times$) for the different speaker

Table 5: DM and disfluency ratios for non-overlapping (**mono-speaker**) and overlapping (**over**) speech. For the latter, passive (P) and active (A) conditions are compared. Figures are given globally and separately for non-intrusive (**non-intr**: bck, att) and intrusive (**intr**: cmp, tst) overlap types.

category		% DM	% disfluencies FP	RV	RP	All
mono-speaker		2.4	2.0	2.5	2.5	6.9
over	P	2.1	1.6	2.3	7.2	11.1
	A	5.9	0.5	3.0	11.0	14.5
non-intr	P	2.4	1.6	2.0	1.3	4.9
	A	7.2	0.6	0.9	5.2	6.7
intr	P	2.0	1.6	2.5	9.5	13.6
	A	5.4	0.4	3.8	13.0	17.2

categories (Journalists, Interviewees, Chairman) are produced by cumulating the occurrences for each indivual of a category. As previously, disfluency rates are given for the different types of regions: non-overlapping (**non-over**) and overlapping (**over**). The latter are then analysed with respect to both passive and active roles in overlaps, corresponding respectively to the primary and overlapping (secondary) speakers. Overlap types are examined in intrusive (**intr**) and non-intrusive (**non-intr**) conditions. Disfluency rates are lower for non-intrusive (**bck**, **att**) segments as compared to intrusive overlaps. In passive conditions, they are even lower than the average disfluency rate in non-overlapping speech: back-channels are known to be poorly disfluentogenous and the primary speaker of an **att** overlap is by definition at the end of his/her turn.

Concerning the relation with the speaker's role, we can see in Fig. 4 that although overall disfluency rates are almost the same for Journalists, Interviewees, and the Chairman, condition-dependent rates in overlapping speech are quite different. In non-intrusive segments, Interviewees have higher disfluency rates; for intrusive segments the situation is dissymmetric for passive and active conditions: in the passive case, Journalists have higher rates, while for the active condition, rates are comparable. Possible explanations include an exchange of standard roles (active overlap for Journalists and passive overlap for Interviewees). The Chairman achieves lower disfluency rates in all conditions.

Fig. 5 gives the same box-and-whisker representation by overlap types: **att**, **bck**, **tst** and **cmp**. It reveals that disfluency rates remain very low for **bck** overlaps, both for passive and active overlap speakers. Minimal rates are observed for the passive **att** condition. By contrast, active speakers happen to become very disfluent during **att** overlaps. This can also be related to turn-start positions. Whereas **tst** favours the production of disfluencies by active speakers, passive (primary) speakers become dramatically disfluent on **cmp** segments which correspond to overlapping comments from the entering speaker (see subsection 5.2.).

Figure 4: Disfluency rates for the different segment types (non-overlapping, overlapping intrusive/non-intrusive: passive and active roles), and the different speaker roles. **int** and **non-int** respectively mean intrusive and non-intrusive segments

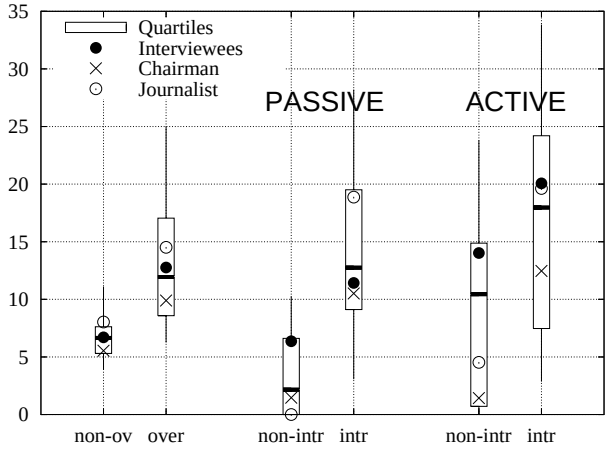
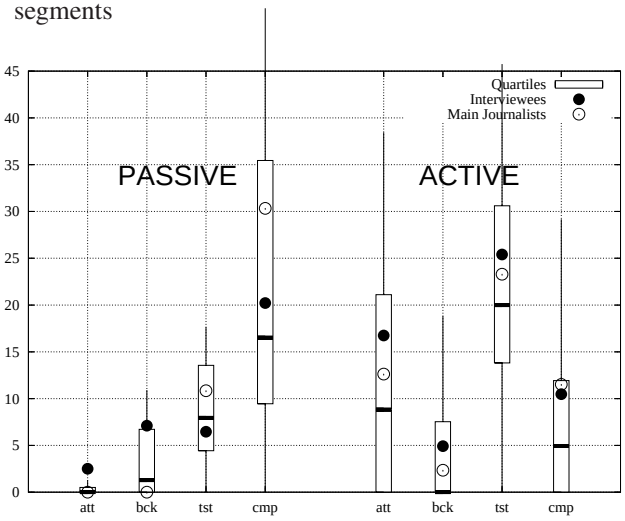


Figure 5: Disfluency rates for the different overlapping segment types: **att**, **bck**, **tst**, **cmp** for passive and active overlap speakers, with details for the different speaker roles. **int** and **non-int** respectively mean intrusive and non-intrusive segments



6. Discussion and perspectives

The choice of working on broadcast political interviews for studying disfluencies and overlap speech segments revealed to be quite productive. Using a bottom-up approach we converged after a few iterations (Adda-Decker et al., 2003; Boula de Mareüil et al, 2005; Adda et al., 2007) on a set of annotation tags for both overlap segments and disfluencies, and exhibited correlations between the occurrences of

these events in relation with the speaker role (i.e. interviewer/interviewee).

In the annotation process, we chose to preserve the interaction structure, and to relax temporal synchronisation constraints at turn boundaries in the case of overlaps, in order to simplify the annotation task and to preserve legibility of the annotated material. The work of segmenting and annotating overlapping speech highlighted that differences between overlap tags happen to be subtle and may give rise to diverging interpretations. Yet, the established annotation scheme and the resulting annotations enables us to study the distribution of overlaps, their types as well as their link with disfluencies and speaker roles.

We observed that overlaps generate twice as many disfluencies as non-overlapping speech portions. In non-overlapping speech, each disfluency type (as well as discourse markers) accounts for about 2% of the corpus. The disfluency rate increase mainly concerns repetitions, in particular for active speakers in intrusive overlap situations such as turn stealings. More repetitions and discourse markers are observed for active speakers than for passive speakers, which can also be explained by the turn-start position. Previous studies showed that disfluencies and discourse markers occur at the beginning rather than at the end of utterances (Adda et al., 2007). Passive (primary) speakers become dramatically disfluent during complementary comments brought by their interlocutors. This corroborates the intrusive nature of these complementary overlaps which do not aim at a speaker change but may disturb the main speaker due to their length and informational content. By contrast, back-channels do not increase the disfluency rate of passive speakers. This rate is even lower than it is in non-overlapping speech.

Finally, interesting differences are observed between journalists and interviewees, whose roles are asymmetric. Even though their disfluency rates are on the whole comparable, journalists show higher disfluency rates when they are passive speakers in intrusive (turn stealing or complementary) overlap situations. In this case, there seems to be an exchange of standard roles (active interruption for journalists and passive overlaps for interviewees).

Enriched and more accurate models are necessary for both talk-in-interaction analysis and speech recognition (Delmonte, 2005; Schegloff et al., 1977). We think that drawing up a descriptive overlap inventory may contribute to the design of a pragmatics model and may be profitable to improve automatic conversational speech transcription, whose performance is still poor as compared to prepared speech recognition.

7. Acknowledgements

We are indebted to the INA Research & Experimentation Directorate (`http://www.ina.fr/`) for the *L'Heure de Vérité* corpus. This work was partly financed by the CAP DIGITAL Competitiveness Cluster project INFOM@GIC.

8. References

G. Adda et al. 2007. Speech overlap and interplay with disfluencies in political interviews. In *International Workshop on Paralinguistic Speech - between models and data, ParaLing 2007*, pages 41–46, Saarbrücken.

M. Adda-Decker, B. Habert, C. Barras, G. Adda, P. Boula de Mareüil, and P. Paroubek. 2003. A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models. In *Proceedings of the Disfluency In Spontaneous Speech (DiSS) Workshop*, Göteborg, September.

Claude Barras, Edouard Geoffrois, Zhibiao Wu, and Mark Liberman. 2001. Transcriber: development and use of a tool for assisting speech corpora production. *Speech Communication*, 33(1-2):5–22, January.

A. Bell. 1984. Language style as audience design. *Language in Society*, 13(2):145–204.

C. Blanche-Benveniste. 1990. Le français parlé, études grammaticales. *Éditions du CNRS, Paris*.

Ph. Boula de Mareüil et al. 2005. A quantitative study of disfluencies in french broadcast interviews. In *Proceedings of the Disfluency In Spontaneous Speech (DiSS) Workshop*, pages 27–32, Aix-en-Provence, September.

J. Carletta. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2):249–254.

Loredana Cerrato and Mariapaola D'Imperio. 2003. Duration and tonal characteristics of short expressions in Italian. In *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*, pages 1213–1216, Barcelona, August.

Rodolfo Delmonte. 2005. Modeling conversational styles in italian by means of overlaps. In *Proceedings of Disfluency In Spontaneous Speech (DiSS) Workshop*, pages 65–70, Aix-en-Provence, September.

H. P. Grice. 1975. Logic and conversation. *Syntax and Semantics*, 3:41–58.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53:361–382, Decembre.

E.A. Schegloff. 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29:1–63.

E. Shriberg, A. Stolcke, and D. Baron. 2001. Observations on overlap: Findings and implications for automatic processing of multi-party conversation. In *Proceedings of the European Conference on Speech Technology (EuroSpeech)*, pages 1359–1362, Aalborg, September.