Co-narration in French conversation storytelling: A quantitative insight

Roxane Bertrand, Robert Espesser

▶ **To cite this version:**

Roxane Bertrand, Robert Espesser. Co-narration in French conversation storytelling: A quantitative insight. Journal of Pragmatics, 2017, 111, pp.33 - 53. 10.1016/j.pragma.2017.02.001 . hal-01513024v1

HAL Id: hal-01513024

https://hal.science/hal-01513024v1

Submitted on 25 Feb 2022 (v1), last revised 6 May 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Abstract

This study addresses the issue of the interactional achievement of storytelling in French face-to-face conversations. Previous studies have described storytelling as a *joint activity* in which together, narrator and listener actively collaborate for a successful achievement. The main findings were either based on qualitative studies (*Conversational Analysis*) or established in experimental conditions that did not really fit the conversational context. Using the *Corpus of Interactional Data*, we argue that conversational storytelling can now be described in a more quantitative perspective in the scope of the emergent corpus-pragmatics approach. Turn, morpho-syntactic information and laughter are investigated. We test to what extent the temporal evolution of these components within speech production of each speaker throughout narratives provides evidence in favor of *joint activity*, i.e. *co-narration* systematically performed by both participants and resulting in a specific turn-taking organization.

Keywords

Conversational storytelling; co-narration; listening responses; turn-taking; morpho-syntactic category; French

Co-narration in French conversation storytelling: a quantitative insight

Roxane Bertrand & Robert Espesser

Aix Marseille Université, CNRS

roxane.bertrand@univ-amu.fr

1. Introduction

Since Sacks' pioneering works (1992, Lecture 2) in the *Conversation Analysis* framework (henceforth CA), different authors have shown that telling stories is a frequent activity in various interactional situations, among them ordinary and familiar conversations. Norrick (2007: 127) goes further saying that "storytelling is a fundamental mode of everyday conversation" which fulfills crucial functions including "sharing personal news, entertaining listeners, revealing attitudes, constructing identity, inviting counter-disclosure, and so on". More formally, the author defines storytelling as a shared activity resulting in a transfer of information from narrator to listener. However, despite its asymmetrical nature, narrative displays a co-construction involving all the participants. By a co-construction we mean "an interactionally collaborative achievement" as initially defined by Schegloff (1982) (see also Rühlemann 2015 for a review). How this co-construction is achieved and more specifically how the listener's activity unfolds throughout the narrative is the focus of this study. Whatever the terminology used, the notion of co-construction has been fruitfully analyzed from the *CA* perspective (Sacks, Schegloff & Jefferson 1974; Jefferson 1978; Goodwin 1979, 1981, 1984; Schegloff 1982, 1997; Norrick 2007, 2010, 2012; and so on). Furthermore, it is the core of the *collaborative model* (Clark 1996) in which dialog is considered a *joint activity* involving a mutual and constant coordination between participants. Sharing this assumption, Bavelas et al. (2000) have for the first time experimentally investigated listening responses and showed that *appropriate responses*, in other words expected responses depending on the current activity (as detailed in Section 2.2)- are required for successful storytelling. Our perspective is not one of analyzing the success of stories, but rather of showing that the speech production of both participants and more particularly the responses provided by listeners are indeed appropriate. Findings concerning appropriate responses in earlier works have been established either in experimental conditions or with natural data mainly involving qualitative data. We argue that storytelling can now be systematically described in a more quantitative way. Using the *Corpus of Interactional Data* (CID) (Bertrand et al. 2008) involving French face-to-face conversations, we aim to extend previous experimental results to a *conversational-style* corpus. Our approach subscribes to the emergent corpus-pragmatic research as presented in the first Handbook dedicated to this new perspective combining both *pragmatics* and *corpus linguistics* (Aijmer & Rühleman 2015). Indeed this approach provides quantitative and statistical analysis (usually addressed by *corpus linguistics*) while taking into account the context-dependence of pragmatic phenomena (usually addressed by *pragmatics*).

Corpus pragmatics also combines both a horizontal and a vertical methodology: "Given the dependence on context, pragmatic research has methodologically relied on the analysis of small numbers of texts where careful 'horizontal' reading is manageable, that is, where large and often whole texts are received and interpreted in the same temporal order in which they were produced and received – a methodology which, (…), contrasts sharply with the 'vertical' methodology prevalent in corpus linguistics" (Aijmer & Rühlemann 2015: 9), mainly consisting on frequencies analysis. Also, the different annotations performed on corpus make this type of analysis possible at each of the different linguistic levels annotated. For example, Part-of-Speech tags (POS) provide the morpho-syntactic information for each token that can be easily extracted (vertical reading). In the present study, turns, morpho-syntactic information and laughter are investigated. This vertical reading relates a horizontal one insofar as our work deals with listening responses and knowledge about their functions established in previous pragmatic analysis. More specifically, we test to what extent the temporal evolution of the components investigated within the speech production of each speaker throughout narrative provides evidence in favor of *joint activity*, i.e. *co-narration* systematically performed by both participants and resulting in a specific turn-taking organization.

2. Background

2.1 Expectations and rights in asymmetrical activity

Storytelling is seen as *asymmetrical activity* involving different discursive roles of the storyteller/narrator and listener. Both participants however actively participate and work together to construct the story (Jefferson 1978; Goodwin 1984; Norrick 2010, 2012; Bavelas et al. 2000; Guardiola & Bertrand 2013). While conversing, participants must respect the expectations that they have for the activity in which they are involved. Within the scope of these expectations, the storyteller must ensure that he/she can begin to tell the story and will not be limited in speaking time given that he/she needs several turns (or *Turn-Constructional Units*) to achieve what is called a *large project* (Selting 2000). Also the storyteller has to ensure the *tellability* of the story. Norrick (2007: 136) notes that "storytellers may worry about the scathing 'so what?'" which may be the response to a story with no clear purpose or significance. Beyond the content level of narratives, the author also highlights the importance of context in which narratives have to be produced. The relevance of stories indeed depends on, among other things, the circumstances or goal (informing, making people laugh) or, the relationship between the participants. Thus, while some familiar stories may be unoriginal because they are precisely shared by several family members who take delight in telling and retelling the same events, stories involving unfamiliar speakers have to present new or unexpected events in order to be appropriate. Boundaries of tellability are therefore subject to change and within their scope of expectations, speakers have also to define these boundaries. So, this question of tellability can still be viewed as a "joint construct" on behalf of both participants (Norrick 2007).

Like the storyteller, the listener is expected to adopt typical behavior. By accepting and becoming the receiver, the listener could be viewed either as "mute or invisible" or as "a speaker-in-waiting" (see Schober & Clark 1989 for a review of different models which aimed at describing storytelling from a monological perspective). Unlike this autonomous view of conversation, it is fairly well-known by now

that listeners as the story-recipient are actively involved and expected to provide appropriate responses (Jefferson 1978; Norrick 2008).

2.2 Listening responses in a joint activity

Storytelling is a *joint activity* in which the listener's role is just as important as the narrator's role in order that the storytelling is successful. Since 1970, Yngve has distinguished a parallel and subordinate channel by which the participant who is not speaking gives brief messages (such as *mh*, *uh*) to the main speaker without interrupting him/her. The author has suggested an interpretation consisting of a reciprocal effect by which backchannel is crucial in the regulation of dialog and more globally in the success of the communication. In this vein, Bavelas et al. (2000) have shown experimentally that expected responses in storytelling can be affected when listeners were distracted, and simultaneously that the story-telling was less efficient when listeners did not provide appropriate responses. Among these appropriate responses, the authors have proposed a distinction between *generic* and *specific responses* which fits the previous *continuers/assessments* dichotomy respectively (Schegloff 1982). *Generic responses* are often vocal items such as *mh* or gestural signals like *nods*. They are used to express attentiveness, interest and understanding of the current discourse. As the explicit marks of listening and comprehension processes, generic responses enable the narrator to track potential misunderstandings, whatever the story in which they are produced. On the contrary, *specific responses* are often expressed to comment about the current discourse. Thus they may include brief verbal utterances, hand gestures, winces, or particular tones of voice (Tomlinson & Fox Tree 2011). Whatever the kind of modality in which specific responses are produced, they involve a more evaluative function. In this way, they are typically adapted to each particular story unlike generic responses which do not convey narrative content.

In addition to their function, listening responses have been studied according to the specific environment in which they occur (Ward 1996; Ward & Tsukahara 2000; Koiso et al. 1998; Bertrand et al. 2007, Gravano & Hirschberg 2011; Beňuš et al. 2011), in relation to turn-taking organization for example (for a more exhaustive review see Tolins & Fox Tree 2014). More specifically, it has been shown that they occur at different stages in the narrative. *Generic responses* appear preferentially during the initial phases of narrative when the common ground is being established. Whereas, specific responses are more likely to appear later in stories once sufficient information has been provided (Goodwin 1984; Bavelas et al. 2000; Blöndal 2005; Stivers 2008; Guardiola & Bertrand 2013). Except Bavelas and colleagues who based their results on one single mean measure for each of the two halves of the stories, the other studies are based on qualitative observations. The present study aims to improve these previous results by providing a more precise estimation of where listening responses occur within the narrative. More generally, the examination of how feedback responses change over time enhances our understanding of their roles while contributing to the "overall structural organization" study, under-studied area in CA as noted by Robinson (2014).

Furthermore, Stivers (2008) investigated *vocal continuers* versus *nods* as cues of the *alignment/affiliation* between speakers. In her work, alignment is defined as an adaptation to the activity in progress, i.e. storytelling. Her findings showed that alignment can be achieved through generic responses (vocal

continuers) which help to show the construction of shared knowledge in the beginning of stories preferentially. Defined as a support of the teller's conveyed stance, affiliation is preferentially achieved by specific responses (nods in this case) that display an *evaluative* or *attitudinal* function concerning the events described or the teller's stance for example. In the same vein, Guardiola and Bertrand (2013) brought to light a specific verbal response, i.e. *echo reported speech* which occurs, in most cases, towards the end of the narratives. The authors showed that when the storyteller had given enough information, the listener could then produce this type of response by which he/she could display alignment and affiliation in orienting the response respectively, to the current activity and the expected stance (2013: 14) while simultaneously taking the other's perspective. By reversing the usual asymmetric role of storyteller and listener, the use of this response made the listener a true *co-narrator* (Bavelas et al. 2002: 568). For the authors who aimed at defining the notion of *interactional convergence*, only sequences involving such specific responses could promote a highly convergent sequence. Based on this type of results, the current work is an illustration of the complementarity of both qualitative and quantitative approaches.

2.3 Formal phases of storytelling

Storytelling is considered a *sequential activity* that can be reflected by formal properties. According to Sacks (1972), stories are seen as "sequenced objects" that are very well articulated within the context in which they are produced (cited in Jefferson 1978: 219). Some studies have investigated how stories are triggered and what this means in formal terms. Jefferson (1978) for instance focused on the formal properties of particular phases of narrative such as its beginning and its end. Especially significant for us, Labov and Waletzky (1966) proposed a typology of formal phases within stories. Among other things, these formal phases reflect the different steps in the treatment of information including establishing common ground: as a storyteller, the main speaker has to make several elements of information available to the listener and these elements have to follow a precise sequential order. Respecting this latter via the different phases would contribute to identifying the story in itself by tracking its progress.

According to Norrick (2007: 129), the narrative phases can be described through the questions that they address, namely:

> *Abstract*: answers the question "what is it about?"
> *Orientation*: answers the questions "who, what, when, where?"
> *Complicating action*: consists of sequentially ordered narrative clauses
> *Evaluation*: answers the question "so what?"
> *Resolution*: answers the question "what finally happened?"
> *Coda*: puts off any further questions about what happened or why it mattered."

Not all these phases are necessarily present in every story. The *orientation* presents spatial and temporal elements of the story as well as presenting the characters. The *complication* is related to the actions or events that lead to the *apex* (or *climax*) of the story. The apex is the culminant point after which the *evaluation*, mainly consisting of commentaries about this latter, can be produced. Orientation, complication (involving the apex) and evaluation phases can be sufficient to make a story suitable. Even if all the phases are not required, expectations in terms of story structure are very high. When these

different steps are respected the listener is guided in his/her understanding which in turn allows him/her to provide appropriate responses. Studies dealing with listening activity mention that recipient action differs according to the previous phases such as orientation and the end of the story, as well as being during or after the apex (Goodwin 1984). In the same way, Norrick (2007) claims that evaluation phases appear as more conducive for the emergence of 'co-telling' (co-narration) sequences involving specific responses. In this study, we attempt to show to what extent a quantitative procedure allows us to recover some of these phases.

After this general overview, Section 3 presents the corpus and the method used. Section 4 reports results and Section 5 discusses how the latter, not only highlight the importance of paying attention to listeners but support the vision whereby narratives can be seen locally as a joint activity.

3. Corpus and Method

3.1 Corpus of Interactional Data (CID): experimental design

The *Corpus of Interactional Data* is an audio-video recording of French spontaneous face-to-face conversations (8 pairs of speakers, 8 hours, about 115.000 words). The corpus was recorded in an anechoic room. Each speaker was equipped with a microphone headset enabling the recording of both speakers' voices on two different sound tracks to allow for a fine-grained analysis at the phonetic and prosodic levels as well as the study of overlapping speech. The CID involved familiar speakers who were asked to talk about either unusual situations (3 dyads) or conflictual professional situations (5 dyads) in which they were involved.

3.2 Levels of annotation

Using Praat (Boersma & Weenink 2009) the speech signal was pre-segmented into Inter-Pausal Unit (henceforth *IPU*), defined as speech blocks surrounded by at least 200ms silent pauses; this duration is well-suited to French speech. This indexation makes localization in the corpus easier and facilitates the manual orthographic transcription. It also limits the propagation of errors during the automatic phoneme alignment. More generally, the annotation process (elaborated within the framework of the *OTIM* project, Blache et al. 2009) used the set of IPUs as input. By using the same formal annotation scheme, multiple annotations were then performed at the different linguistic levels (Blache et al. 2010). Precise synchronization between these levels enabled us to study the relationship between them.

3.3 Extraction of parameters

A set of 147 narratives extracted from the CID and involving all the speakers was investigated.
All the narratives in the CID were labeled manually by two annotators (one naïve annotator and checked by one expert). Within each dyad both speakers could alternately be narrator or listener. For each narrative of one speaker, the parameters of the speech production of this speaker (i.e. the narrator) were measured. Within the same time interval the parameters of the production of the other speaker (i.e. the listener) were also measured. Hence by construction the data somehow reflected the link between the two speakers in each dyad.

3.3.1 Inter-Pausal Unit (IPU)

In addition to being advantageous for speech transcription, the IPU is also a relevant parameter for measuring speech production of speakers in dialog (Koiso et al. 1998, Gravano & Hirschberg 2010 among others). The identification of the IPU is easier than other types of units such as intonational or syntactic units. Due to its objective nature, the IPU can also be automatically segmented while requiring neither a time-consuming manual annotation nor several annotators to ensure inter-annotator reliability. Thus, although IPUs are not completely equivalent to turn-taking (nor the intonational or syntactic boundaries), here they will be considered a turn-unit. Henceforth, the terms IPU and turn will be used indifferently.

3.3.2 Morpho-syntactic tagging

Tokens synchronized with speech signal were based on the orthographic transcription. The part-of-speech (POS) tags were then automatically identified using *MarsaTag* (Rauzy et al 2014). *MarsaTag* is a stochastic parser for written French which has been adapted to account for the specificities of spoken French. Among other outputs, it provides a morpho-syntactic category for each POS token. It is worth noting the excellent performance of this tagger that has reached an F-measure of 0.974 (version 2011).
For the current analysis, eight main categories were retained (verbs, nouns, adjectives, adverbs, prepositions, conjunctions, pronouns and auxiliary verbs). Another category that did not fit into the scope of morpho-syntactic categories was added. This category called *Interjection* is due to the conversational nature of the CID involving different oral phenomena (genuine interjections such as *bouh (booh)*, discursive markers such as *quoi (what)* and backchannel items such as *mh*). Although this category concerns less than 40 forms of the lexicon, it is of major importance in the overall treatment since it accounted for almost 10% of the tokens in the whole corpus.
To characterize speech production of speakers in dialog, and more particularly the listener's production, morpho-syntactic categories were considered according to their semantic load. As we said above (2.2), generic responses may occur in any type of narrative because of their very general character and their weak semantic load while specific responses cannot occur anywhere because they are more closely connected to the narrative. This suggests that specific responses could be more complex or semantically richer than generic ones. We know that verbs, nouns, adjectives and adverbs are characterized by a higher semantic load than prepositions, conjunctions, pronouns and auxiliary verbs, resulting in the well-established distinction between *content/lexical* versus *grammatical/function words* respectively. Grammatical words mainly work at a syntactic level by linking together different linguistic units (chunks, clauses, sentences, utterances). In this way, they are viewed as poorly informative although they indeed contribute to the construction of meaning with content words. Thus we chose to look at grammatical and content words separately although we expect that they will show a parallel evolution. We also dealt with the category of interjections separately. As we said above, this category is very heterogeneous and can therefore raise some difficulties: for example, whereas the non-lexical token *mh* clearly belongs to the interjection category and is very predominantly associated with generic responses, some other tokens can be more ambiguous. We address this peculiarity by distinguishing genuine interjections from generic ones (that we will define later). We claim that these three formal categories - content words, grammatical words and interjections - provide a suitable description of the richness of speech reflecting the distinction

between generic versus specific responses. We intended to adopt here a formally-based approach dealing with the morpho-syntactic categories in contrast with the meaning-based approach adopted by Bavelas and colleagues (2000) and which requires a high level of interpretation from naïve judges. Using such a formally-based approach, which requires neither interpretation nor time-consuming annotation, would allow us to refine our knowledge on listening responses which have until now been mainly based either on functional or positional criteria.

3.3.3 Laughter

The last component investigated was laughter. Laughter was labeled manually throughout the whole corpus by two annotators since it remains difficult to annotate automatically. Since laughter cannot be assigned a morpho-syntactic value, we studied it separately. Moreover, and even if one same component could be either generic or specific, laughter is certainly the most multifunctional one: laughter "could be polite or appreciative generic responses; it could also be specific to the narrator's own amusement, or it could be maintaining the dialogue (metacommunicative)" (Bavelas et al. 2000: 946). In all cases, it reflects the story's tellability and a form of success. In other studies however, laughter can also be otherwise identified as a sign of awkwardness for example (Chafe 2007).

3.4 Measurements

Before examining the different parameters involved in the evolution of a speaker's production we had to address the issue of the variation in length of the narratives.

The narrative duration was normalized between 0 and 1 to account for the variation of durations. For example, the apex, which typically occurs around the mid-point of the narrative, thereby corresponds to a normalized time value about 0.5. This time normalization can be seen as an extension of a piece-wise procedure (1st vs 2nd half of the narrative, for example). The normalization allows a fine-grained description of time evolution. Unlike splitting the time into a number of categories, maintaining time as a continuous variable enhances statistical power. In summary, we considered that the variation of duration of the narratives could be accounted for by linear scaling.

In order to quantify the evolution of the production of both listener and narrator, we needed to estimate some parameters on a given interval throughout the narrative. At least two main types of method could be used: a window-based method, which is commonly used in signal processing, and which consists in repeatedly shifting a fixed-length window by a fixed-length interval throughout the narrative (Guardiola et al. 2012, for an example of using this method), or an utterance-based or turn-based method taking into account the dynamical aspects of interaction (Levitan & Hirschberg 2011). Indeed the turn-based method is more suitable for tracking the length variations of the production of both speakers (for example short production for listener when he/she produces feedback items versus more long turns for narrator). You will recall that we used IPU as a unit of turn. This method also avoids randomly cutting the IPU or conversely merging several IPUs, generating imprecise measurements. For example, let us suppose the fixed time window is greater than an IPU: given 3 tokens in this window, 4 IPU patterns are possible:

- 3 IPUs containing 1 token each
- an IPU containing 1 token followed by a second IPU containing 2 tokens

- the reverse of the previous pattern
- a single IPU containing 3 tokens

Given the focus of this study (evolution of speech production during narrative), these four situations have to be distinguished since all other things being equal, they each reveal a different turn organization. Obviously, the window-based method could not distinguish these four patterns.

The following parameters were used:

- rate of occurrence of IPUs
- Token count by category, within each IPU: token count was preferred to IPU duration because it better reflected the quantity of production of each speaker. Indeed IPU duration is more sensitive to segmental and prosodic effects (syllabic duration, supra-lengthening and speech rate). Moreover, such a fine-grained token count by category allowed us to account for phenomena of compensation between categories.
- Token proportion by category within each IPU: the proportion of each category of tokens pertains to the semantic richness of the speaker production. It completes the information given by the simple token counts (above). The comparison of trends of count cannot statistically assess the change in proportion.
- Probability of occurrence of an IPU containing at least one laughter: this probability is a valuable measurement of the laughing activity because the great majority of IPUs included only one or no laughter.

Thus all the parameters were analyzed using the IPU (turn)-based method, with the exception of the rate of IPU which does not require such a fine-grained method and was estimated using a window-based method.

4. Results

4.1 Descriptive Statistics

4.1.1 Narrative distribution and duration by speaker

The CID contained 147 narratives. The total cumulated duration of the narratives (12,607 s, i.e. roughly 3 and a half hours) accounted for 43% of the total duration of the CID (29,292 s, i.e. roughly 8 hours). Table 1 shows descriptive statistics of the narratives by speaker for the 8 dyads. All durations are in seconds. IQR stands for interquartile range. Dyad total duration ratio is the ratio of the cumulated duration of narratives between the two speakers of each dyad.

Speaker	Narrative count	Total duration (s)	Median duration (s)	IQR duration (s)	Dyad total duration ratio
AB	7	1261	145	143	1.41
CM	6	895	131	70	
AG	14	879	34	63	1.38
YM	13	635	41	29	
BX	6	355	54	26	1.14
MG	5	310	53	22	

ML	12	998	73	69	1.02
IM	13	1022	84	64	
MB	17	1721	56	56	1.93
AC	8	890	77	76	
AP	10	558	39	34	1.05
LJ	8	532	57	59	
EB	9	648	68	23	1.67
SR	11	1082	77	61	
LL	2	166	83	36	3.94
NH	6	654	122	71	
Total	147	12607	66	69	

Table 1: Descriptive statistics of narrative by speaker.

Two dialogs were very unbalanced regarding the production of narrative (AC-MB and NH-LL in which LL only produced two narratives in total).

Otherwise, a very long narrative (538s, speaker MB) exhibiting many reiterations of successive formal phases was considered an outlier and was removed from the analysis.

4.1.2 IPU distribution and duration by speaker and discursive roles

The narratives included 5631 IPUs, accounting for 41% of the 13,621 IPUs in the CID.

After removing the previous very long narrative, 5375 IPUs remained. Errors related to the different labeling processes (automatic and manual) needed data screening: (a) there were misalignments between the boundaries of IPUs and narratives, accounting for about 2.9% of the 5375 IPUs. Removing these spurious IPUs did not lead to unexpected modifications of the temporal distributions of IPUs within the narrative, for both narrator and listener; (b) very long IPUs (greater than 8s) occurred and were due to the weakness of the automatic detection of IPU, based on the presence of a silent pause greater than 200ms. It is unlikely that there is no breath taking within a duration exceeding 8s[1]. Not surprisingly, these IPUs were mainly found in the narrator's production, and accounted for about 1.5% of the set of IPUs obtained after the previous phase of screening. Removing these IPUs did not modify the pattern of the temporal distributions of IPUs within the narrative. In brief, the stability of the temporal distribution of IPUs regarding the data screening a) and b) supports the hypothesis that all these dubious IPUs appeared randomly distributed throughout the narratives. Therefore the effect of the data screening is limited to a small decrease (for a total of 4.3% of the 5375 initial IPUs) of the effective number of IPUs for listener and narrator. This impact will be commented in the statistical analysis section 4.2.

[1]For French language, lowering this 200 ms threshold would lead to many more errors due to the confusion of pause with the closure part of unvoiced consonants, or with constrictives produced with a very low energy.

Figure 1 shows the IPU counts and durations by speaker and discursive role. The right panel shows boxplots of the IPU durations (the height of the box indicates the interquartile range, the line in the box indicates the median, the dots indicate the remaining outliers following the standard boxplot definition).

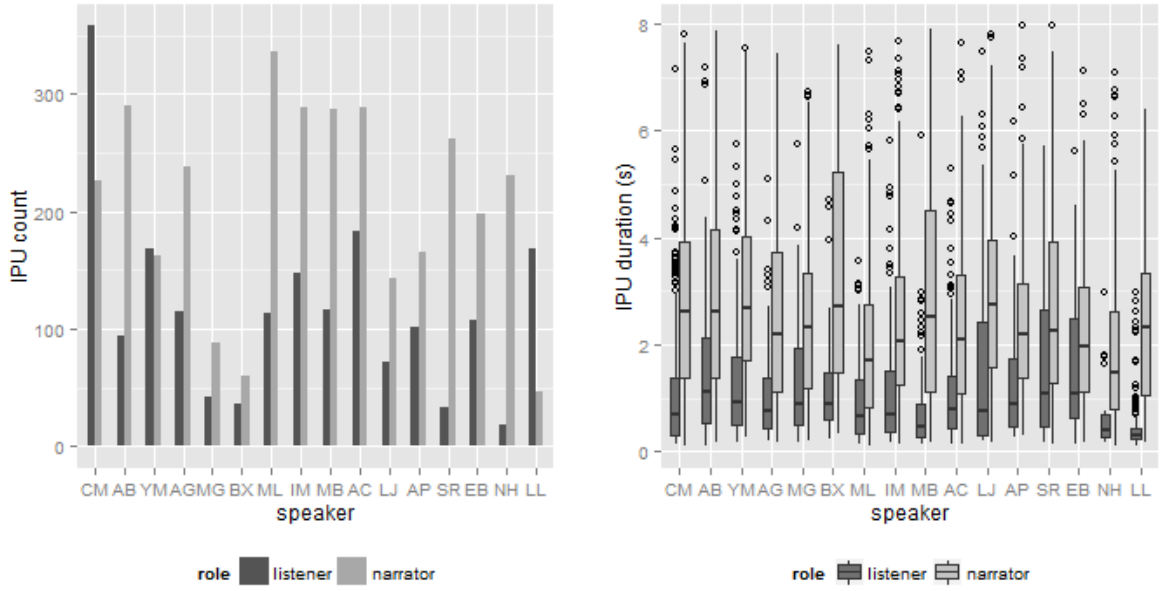

Figure 1: IPU count (left) and IPU duration (right) by speaker and role (after data screening)

Unsurprisingly, most of the speakers produced a much larger number of IPUs as narrator than as listener, except YM (quasi equal), CM and LL. The latter can be explained by the very small sample size of stories (only 2). We also notice that BX and MG produced less IPUs than the other speakers, both as narrators and as listeners. Concerning the IPU duration, the IPUs produced by the narrators were always longer than those produced by the listeners.

4.1.3 Token and Laughter

The narratives included 51,173 items, i.e. tokens and laughter, accounting for 44% of the 115,656 items in the CID.

The previous data screening (i.e. the very long narrative and dubious IPUs) removed 6842 items, accounting for 13% of 51,173 original items.

Figure 2 shows the counts of the remaining items by speaker and role.

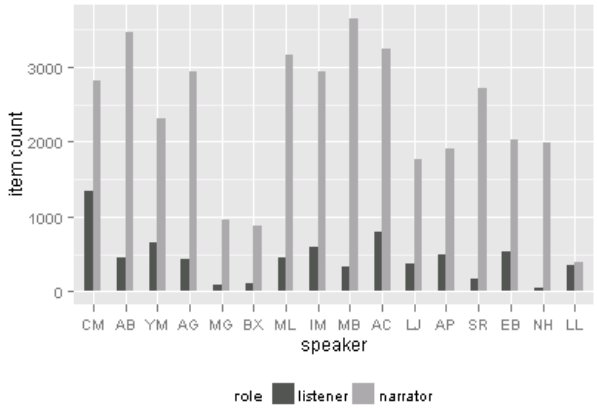

Figure 2: Token count and laughter count by speaker according to the discursive role (after data screening)

Again, figure 2 shows the expected results, i.e. narrators produce many more tokens than listeners. As for the IPU, LL, BX and MG remain the speakers who produced the least tokens, both as narrators and listeners. NH produced very few tokens as a listener since she was in this discursive role only twice.

Figure 3 shows the distribution of POS by category (interjection, grammatical word, content word) and laughter.

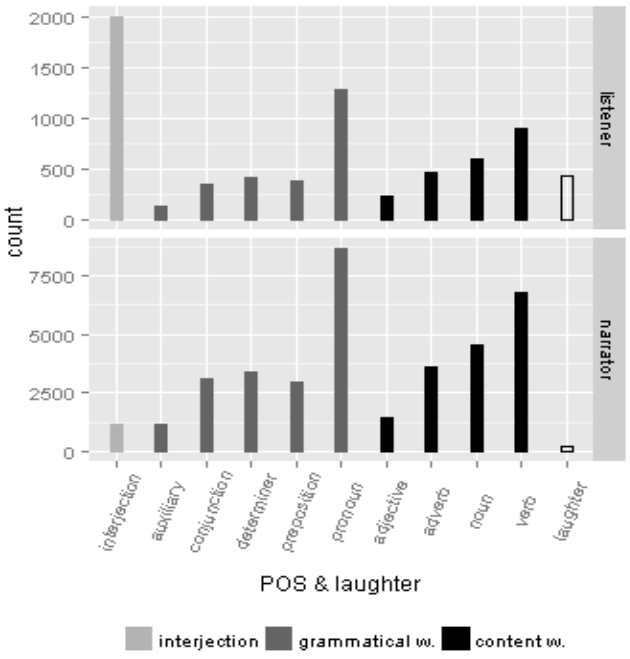

Figure 3: POS count (by category) and laughter count according to the discursive role

We observe that the POS distribution exhibits the same pattern for both listeners and narrators (while taking into account, of course, the different scales), except for interjection and laughter which are much more strongly associated with the listener role.

Table 2 summaries the counts by role and category.

category	Listener count	Narrator count
interjection	1999	1140
grammatical word	2593	19335
content word	2192	16416
laughter	438	218

Table 2: Token count (by category) and laughter count according to the discursive role

4.2 Statistical models

Generalized linear mixed models of the R package (R, 2014) were used (package lme4, Bates & al 2014). A mixed model (LME) incorporates both fixed effects, which are parameters associated with certain repeatable levels of experimental factors, and random effects, which are associated with individual experimental units sampled at random from a population (Pinheiro & Bates, 2000; Quené & Van den Bergh, 2004). Hence LMEs take into account the correlation of observations within the same experimental unit. Moreover LMEs can handle unbalanced data. Fixed effects account for central trends in the data, whereas random effects account for trends at the grouping levels. In this study, the random effects account for grouping by speakers and by narratives within speaker.

4.2.1 Evolution of the rate of IPUs

Note that IPUs rates were slightly reduced by the data screening described in 4.1.2.

The time normalization was realized by splitting the narrative duration into 10 equal bins (time deciles). The value 10 ensured a granularity small enough to describe the temporal evolution throughout the narrative. On the other hand, it was large enough to decrease the risk of having the IPU cut: the decile of the smallest narrative was longer than most of the IPUs in the corpus.

A Poisson mixed model was fitted, with the IPU count within each decile as the dependent variable. The predictors were the time decile rank (from 1 to 10), the discursive role (listener, narrator) and their interaction. The logarithm of the actual duration of the decile was added as an offset to the model to avoid evident bias due to the varying durations of the deciles (it was likely that the longer the decile, the greater the IPU count). Therefore the model estimated an IPU rate. Two random intercepts accounted for variability across the 16 speakers and across the 291 combinations of narrative:role:speak. Random slopes (relative to the time decile rank) did not improve the model. 2180 observations were available.

As expected, the narrator's IPU rate was always greater than that of the listener.

Concerning the time slope, results showed that it was not significant for the narrator ($\beta=0.0077$, $z=0.128$, $P=0.90$); on the contrary, the time slope for listeners was significantly greater than for narrators, because of the significant interaction of role with decile rank ($\beta=0.549$, $z=5.37$, $p<0.001$). The narrator's IPU rate remained constant about 0.3/s, whereas the listener's rate increased from 0.1 (first time decile of the narrative) to 0.27 (last decile) (Figure 4).

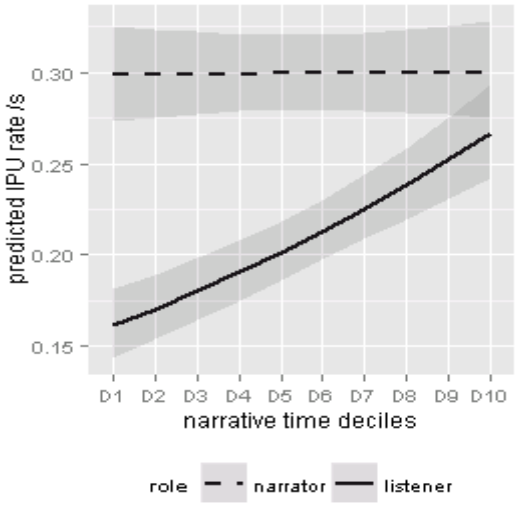

Figure 4: Predicted IPU rates by role according to the narrative time deciles (from the fixed effects only). The grey bands show the 95% confidence interval

4.2.2 Tokens analysis

Unlike IPU rates, note that the counts and probabilities of tokens studied in this section were not impacted by the data screening (4.1.2) because there were computed by IPU.

Examination of the raw data reflected the inherent distinction between listener and narrator. The differences were so evident that the statistical test of the factor role and its interaction with other predictor(s) provided very little (no new) information. Therefore we chose to fit a separate model for each discursive role for the sake of simplicity**.** Furthermore, the token count and the token proportion of a given category in each IPU throughout the narrative duration could be possibly best modeled with a quadratic function of time. Therefore for each discursive role, two models were estimated for each of the three categories of token (one model for count and one for proportion). Each model had the same structure: the dependent variable was the count or the proportion of token (for the category) within each IPU in the narrative. The time predictors were the linear and quadratic forms of the normalized start time (hereafter Nt) of the IPU in the narrative.

As polynomial regression has high collinearity between predictors, we performed a sequential comparison of the nested models obtained with the sequential addition of each time predictor; this procedure is more appropriate in such a case to test the significance of predictors. A likelihood ratio test (LRT) was used to test the differences between models. In addition to the classic P-values, P-values were also computed using a parametric bootstrap procedure (package pbkrtest, Halekoh & Højsgaard 2014) and noted as P_{boot}. There were 400 simulations, which lead to a lower limit of P_{boot} of 0.0025.

The observations were grouped by narratives, which were nested in speakers. Therefore a random intercept was added for each of these grouping levels, i.e. for the 16 speakers and for the 144 combinations speaker:narrative in the listener case on the one hand and for the 146 combinations in the narrator case on the other.

For each of these levels, random slopes relative to time were also added when the likelihood ratio test (LRT) showed that the data were best described by it. These cases are indicated in the results. All the random terms were uncorrelated (to avoid numerical issues due to overfitting).
1640 observations were available for the listener situation, 3299 observations were available for the narrator situation.

Time evolution of token count within IPUs

Figure 5 (left side) shows the token count predicted by the models. The detailed outputs of the comparisons of the nested models are annexed herewith (see appendix A).

Content words category

Two negative binomial models were fitted to account for overdispersion in the data.
As expected, overall the listener produced less content words than the narrator.
For the listener, only the linear term was significant. The count monotonically increased from a minimum of 0.9 token/IPU at the beginning of the narrative to a maximum of 1.56 at the end.
For the narrator, both terms were significant. The function reached a maximum of 5.5 token/IPU at Nt=0.42 and then decreased to 3.6 at the end of the narrative.

Grammatical words category

Two negative binomial models were fitted to account for overdispersion.
As expected, overall the listener produced less grammatical words than the narrator.
For the listener, only the linear term was significant. The count monotonically increased from a minimum of 1.06 token/IPU at the beginning of the narrative to a maximum of 1.78 at the end.
For the narrator, both terms were significant. The count reached a maximum of 6.7 token/IPU at Nt=0.45 and then decreased to 4 at the end of the narrative.
Content words and grammatical words exhibit the same tendency, these latter being slightly more numerous.

Interjections category

Poisson mixed models were used. Random slopes were necessary for each model.
As expected, overall the listener produced more interjections than the narrator.
For the listener, the quadratic term was significant; the interjection count reached a maximum of 1.38 token/IPU at Nt=0.38 and then decreased to 1.1 at the end of the narrative.
For the narrator only the linear term was significant. The count monotonically increased from 0.2 token/IPU at the beginning of the narrative to 0.44 at the end.

Time evolution of the proportion of token by category within IPUs

For each category, two logit mixed models were fitted.
Figure 5 (right side) shows the proportions predicted by the models. The detailed outputs of the comparisons between the nested models are annexed herewith (appendix B).

Content words category

As expected, overall the proportion of content words within an IPU was lower for the listener than for the narrator.

For the listener, both quadratic and linear terms were significant. The proportion reached a minimum of 0.29 (at Nt=0.37) and then increased to a maximum of 0.37 at the end of the narrative.

For the narrator, only the linear term was significant. The proportion monotonically decreased over the narrative (from 0.46 to 0.43).

Grammatical words category

As expected, overall the proportion of grammatical words within an IPU was lower for the listener than for the narrator.

For the listener, only the linear term was significant. The proportion monotonically increased from 0.33 to 0.40.

For the narrator, both terms were significant. The proportion reached a maximum of 0.54 (at Nt=0.46) and then decreased to 0.49 at the end.

Interjections category

Random slopes were necessary for each model.

As expected, overall the proportion of interjections within an IPU was greater for the listener than for the narrator.

For the listener, both linear and quadratic terms were significant. The proportion reached a maximum of 0.36 (at Nt=0.32) and then decreased to 0.22 at the end of the narrative.

For the narrator, the quadratic term was significant. The proportion monotonically increased from 0.02 (at Nt=0.28) to 0.066 at the end of the narrative.

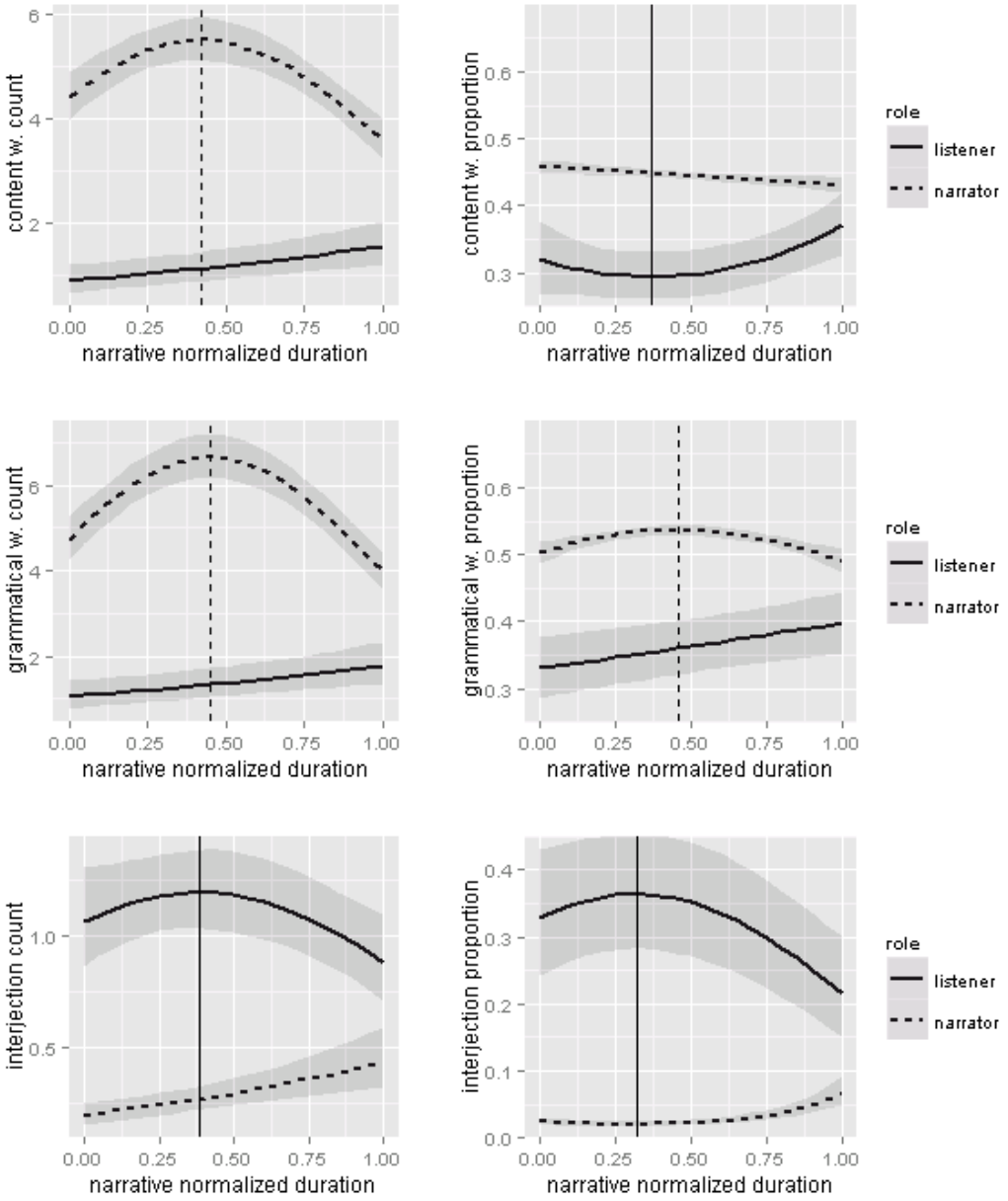

Figure 5: Predicted token count within IPU (left), token proportion within IPU (right) by discursive role and category of POS throughout the narrative normalized time (from the fixed effects only). The vertical lines indicate the relevant vortices. The grey bands indicate the 95% confidence interval

In brief, for the narrator, the token counts within the IPUs changed considerably over the narrative duration. The main changes were the presence of a maximum of grammatical and content words just before the mid-time of the narrative. On the contrary, the token proportions changed to a lesser extent. For the listener, the changes in token counts were globally less marked, but their combinations resulted in greater changes of token proportions (compared to the narrator): the richness of the IPUs increased over the narrative duration.

The speech production by discursive role is reflected by the total number of tokens, obtained by summing the predicted token counts of the three examined categories (figure 6). As expected, the total number of

tokens is much greater for the narrator than for the listener. For the listener, the total number of tokens in an IPU monotonically increased throughout the narrative duration. For the narrator, the total number of tokens reached a maximum just before the mid-time of the narrative and then decreased to a minimum at the end of the narrative.

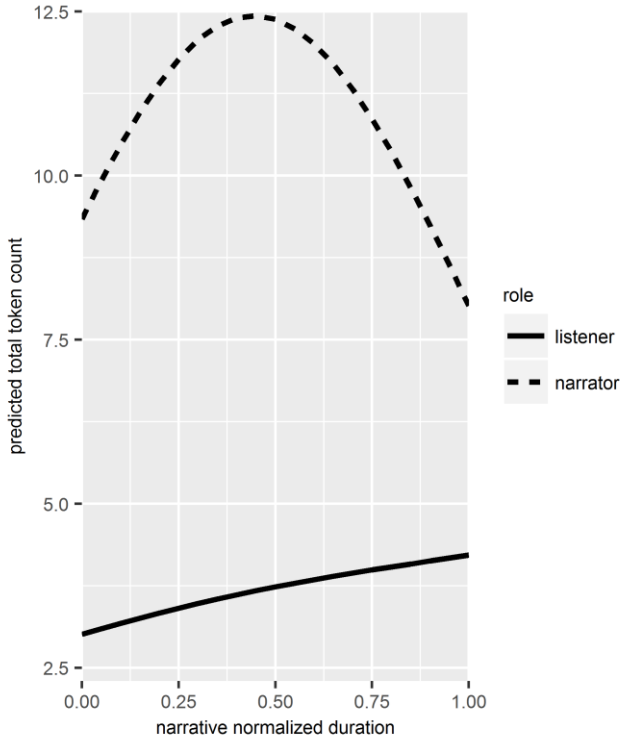

Figure 6: Total number of tokens in IPU by discursive role

Interjection type for listeners

The category of interjections is composed of heterogeneous phenomena such as discursive markers, feedback/generic responses (see 3.3.2). So we attempted to better define what type of phenomena made up this category. Most of the interjections consisted of the items *ouais (yeah), mh* and *ah*, accounting for 80 % of the 1999 interjections, respectively 40.8%, 28%, 11.2% (figure 7).

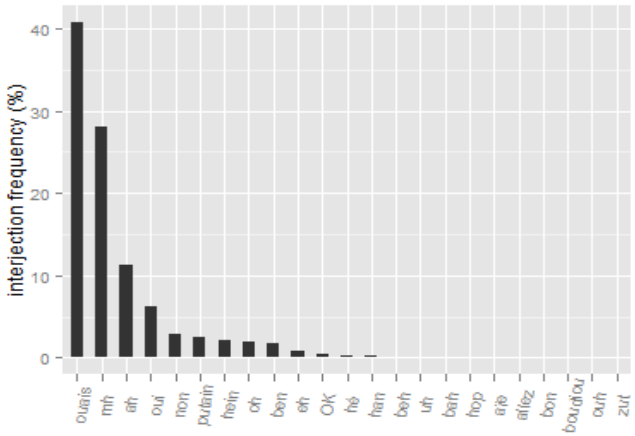

Figure 7: Frequency distribution of the listener's interjections by orthographic form

Given the distribution of the listeners' interjections, we split them into generic versus genuine interjections. We use the term genuine interjections to qualify somewhat primitive expressions of feeling or to characterize items in terms of "mental states" while generic ones are more specifically linked to listening activities and related to the turn-taking system (Norrick 2015: 251-252). Otherwise, although generic items can have other functions than *continuer* -such as *acknowledgment* or *surprise* (see Gardner for details about "mm")-, their role as simple continuer (or generic response) is largely predominant. Moreover, distinguishing these functions would require a specific annotation including for example a prosodic level not yet available and out of the scope here.

Examination of the data showed that a global model with the type of token as factor (generic, genuine) would provide no more information than two separate models and that the evolution of the total number of tokens in each IPU appeared possibly best modeled with a quadratic function of time. Therefore a separate Poisson mixed model was fitted for each type of interjection, with the number of token of the relevant type in the IPU as a dependent variable; the predictors were the linear and the quadratic form of the normalized start time of the IPU. 1640 observations were available. For each model, random intercepts were added to account for the variability across the 144 combinations speaker:narrative and the 16 speakers. Adding random slopes (for the linear and quadratic terms) improved the model of the generic responses.

Figure 8 shows the predicted counts by both models. For the genuine interjections, the comparisons of the nested models showed that both linear ($P=0.23$, $P_{boot}=0.24$) and quadratic ($P=0.1$, $P_{boot}=0.097$) terms were not significant. The number of genuine interjections in an IPU remained constant throughout the narrative duration, at around 0.22 token/IPU.

For the generic responses, the comparisons of the nested models showed that both linear ($P=0.011$, $P_{boot}=0.017$) and quadratic ($P=0.0002$, $P_{boot}=0.0024$) terms were significant. The generic responses count reached a maximum of 0.94 at $Nt=0.38$ and then decreased to 0.56 token/IPU.

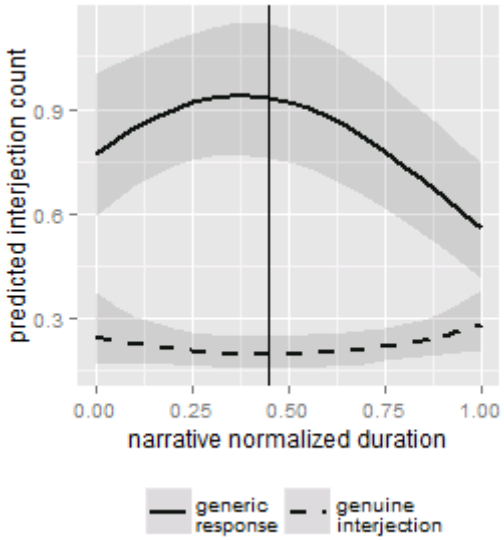

Figure 8: Predicted generic response and genuine interjection counts within listener's IPUs throughout the narrative normalized time (from the fixed effects only). The vertical line indicates the relevant maximum. The grey bands indicate the 95% confidence interval

The evolution of the generic responses count is indeed responsible for the evolution/change of the interjections category.

4.2.3 Time evolution of the occurrence probability of IPU containing laughter

As for the tokens, examination of the data showed that a global model with role as factor would provide no more information than two separate models. Furthermore, the evolution of the probability of an IPU containing laughter appeared to be best modeled with a quadratic function of time. Therefore a separate logit mixed model was fitted for each discursive role, with the presence/absence of laughter in the IPU as dependent variable; the predictors were the linear and the quadratic form of the normalized start time of the IPU in the narrative. 1878 observations were available for the listener, 3315 for the narrator. Random intercepts were added to account for the variability across the 145 combinations speaker:narrative and the 16 speakers. Adding random slopes did not improve the models. To test the significance of the predictors, we proceeded as for the token analysis (i.e. sequential comparisons of nested models and parametric bootstrap procedure).

Figure 9 shows the probabilities predicted by the models.

As expected, overall the probability that an IPU contained laughter was much greater for the listener than for the narrator.

For the listener, the comparisons of the models showed that the quadratic component was significant ($P<0.0001$, $P_{boot}=0.005$). The probability reached a maximum of 0.23 at Nt=0.58.

For the narrator, the comparisons showed that both quadratic and linear component were significant ($P<0.0001$, $P_{boot}<0.0025$, $P=0.012$, $P_{boot}=0.005$, respectively). The probability reached a maximum of 0.058 at Nt=0.73.

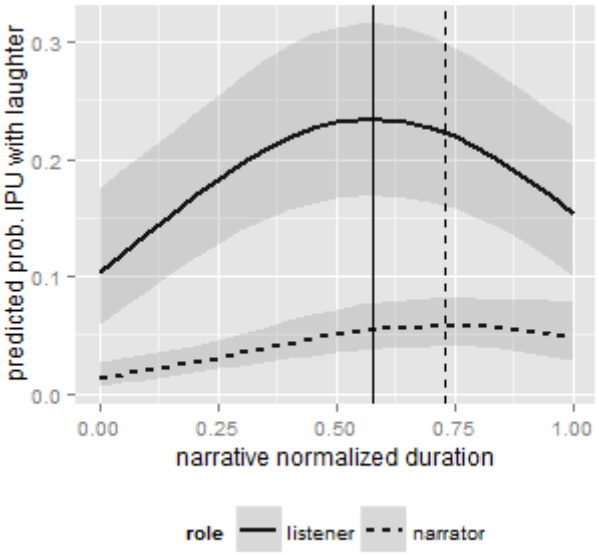

Figure 9: Predicted probability of IPU containing laughter *by role, throughout the narrative normalized duration (from the fixed effects only). The vertical lines indicate the maxima. The grey bands show the 95% confidence interval.*

The estimated probability of IPU with laughter was based on a model opposing IPU with laughter to IPU without laughter, in other words it was the proportion of IPU with laughter in the narrative. Therefore it was not impacted by the data screening as described in 4.1.2.

4.2.4 Individual characteristics

Examining individual characteristics was relevant due to the relatively small number of speakers (16). The individual trends can be obtained by direct examination of the random terms estimated by the model, and/or examination of the predicted values of the dependant variable at the speaker level.
Firstly we examined the individual characteristics concerning the production of tokens within IPUs, then the production of IPU with laughter.

Tokens proportion

We focused on proportion of category of token within the IPUs which are relevant for describing individual characteristics concisely. Only the interjection and content word proportion are shown, since the evolution of the grammatical word proportion was somehow similar to that of content word.
The related models had more complex random terms therefore the individual trends were best described with plots of explicit estimated proportions. Figure 10 shows the estimated proportions by speaker and discursive role.

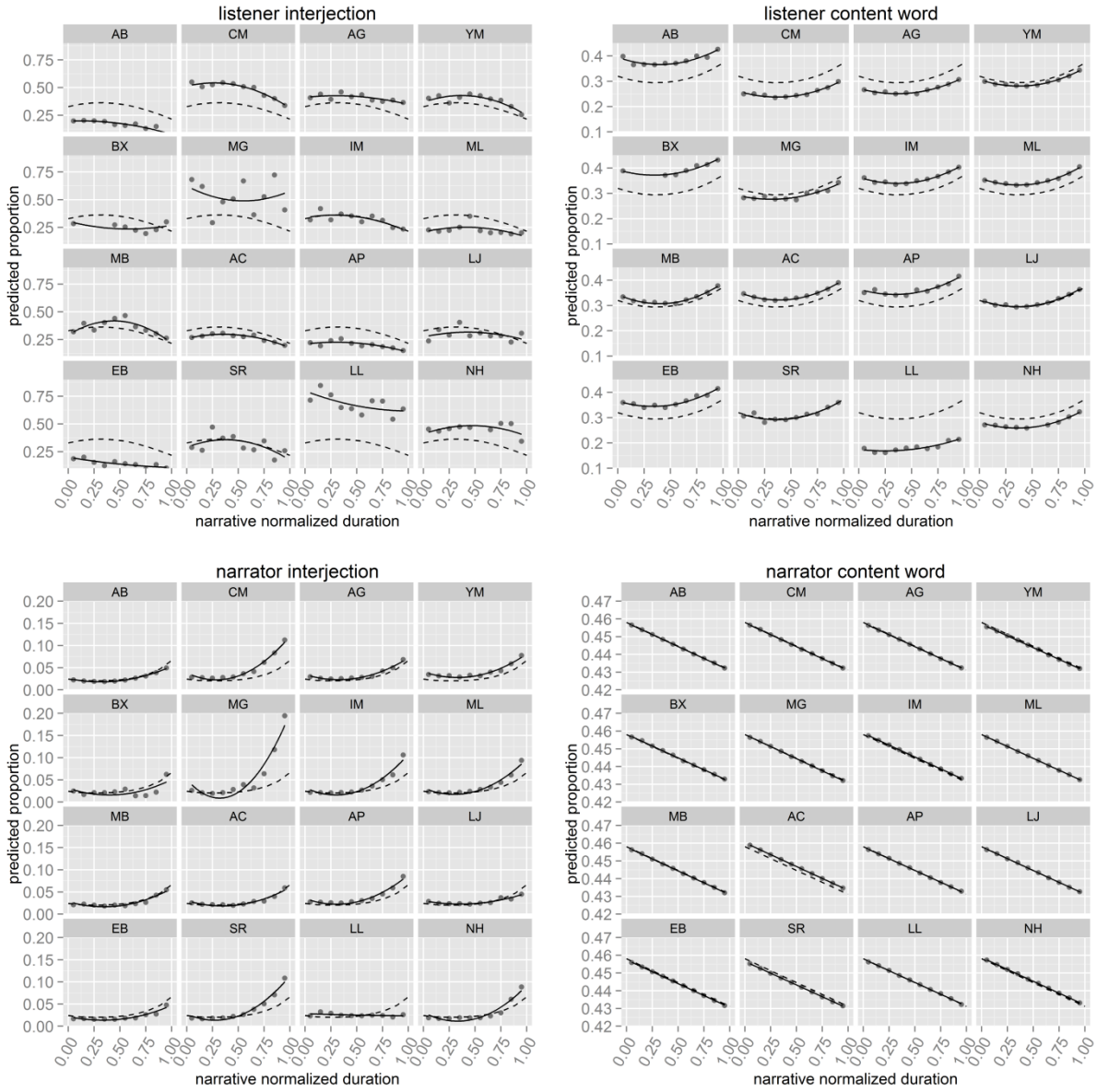

Figure 10: Predicted proportion of interjection (left) and content word (right) at the speaker level, as listener (top) and narrator (bottom). The dotted line indicates the predicted central trend (fixed effects only).

Overall, there was less variability between narrators than between listeners. There was also less variability between speakers for content words than for interjections (presence of parabola with open upward).

For many speakers in the listener role, the content word and interjection proportions were somehow symmetrical. Compared to the central trend, an excess of interjections corresponds to a deficit for content words. However, the speakers BX, MG and IM did not show this symmetry.

Listeners BX, MG and NH had interjection proportions which were not in line with the central trend, with no clear or even opposite trends. In addition, MG and NH had high mean interjection proportions.

As listeners, the following speakers were in line with the decreasing central trend but with an important offset:

CM had a high mean interjection proportion (about 0.15 above the mean central trend)

LL had the highest mean interjection proportion (about 0.35 above the mean central mean).

EB and AB were the listeners with the lowest proportions of interjection (mean level respectively 0.19 and 0.16 below the mean central trend)
As narrators, CM and MG showed a higher increase of interjection proportion at the end of the narrative.

Laughter

Laughter models had simpler random terms, hence the individual trends may be described by what are called caterpillar plots of the random intercepts. Figure 11 shows the random intercepts estimated by the IPU with laughter probability model, by speaker and discursive role.

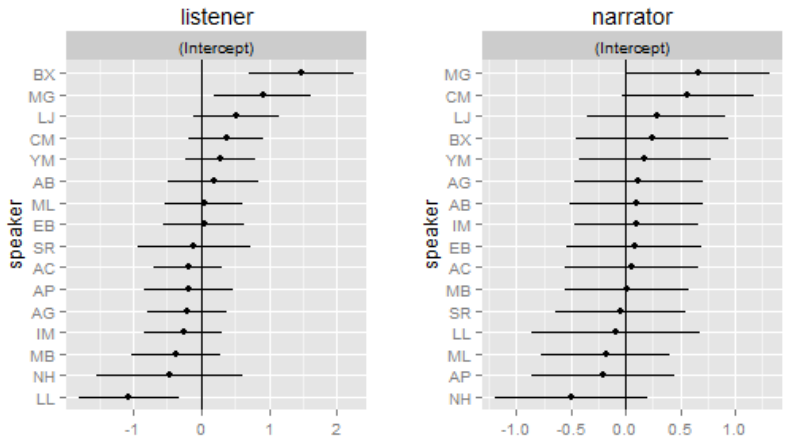

Figure 11: IPU with laughter probability model: speaker random intercepts. The vertical line at zero corresponds to the central trend.

A speaker with a positive (negative) intercept produces IPU with laughter more (less) frequently than indicated by the central trend, throughout the narrative. Not surprisingly, for both discursive roles, the speakers who were instructed to tell stories with unusual situations (i.e. AB-CM, BX-MG, AG-YM) have positive intercepts (except AG as listener).
As listeners, speakers BX and MG have a higher proportion of IPU with laughter whereas speakers LL and NH have the lowest.

5. Discussion

In this study, we investigated whether the speech production of 16 speakers in 147 French conversational narratives could be described by modeling the temporal evolution of IPU (turns), morpho-syntactic components and laughter. Our quantitative results provided evidence that narratives exhibit co-narration phases through the different responses provided by listeners.

Regarding listeners, we observed a marked increase in IPU/turn rate from the beginning of the story to the end as well as an increase in the global number of tokens by IPU (retained as a parameter for estimating the size of turn). Hence as a story progresses, listeners tend not only to speak more often but also at greater length. Our interpretation is that these longer and more frequent turns provide listeners with the opportunity to give *specific responses*. These *specific responses* largely contribute to the content and the dramatization of the narrative once sufficient information was provided. Regarding narrators, the

IPU/turn rate remains constant throughout the narrative while the global number of tokens peaks a maximum count around the middle of narratives. In other words, turns from narrators are produced at the same pace but they vary in size (being longer until the middle of narrative and then becoming shorter). This decrease of tokens for narrators and the increase throughout the story for listeners, reveals a more symmetrical turn-taking exchange in the latter part of stories. This result is the first evidence in favor of the narrative being conceived as a *joint action* in terms of speech production time.

This joint action is also highlighted by our results with the morpho-syntactic component within turns. After having distinguished between *content word*, *grammatical word* and *interjection*, we showed that these three categories enabled us to retrieve the *generic* versus *specific listening responses* distinction previously established by Bavelas et al. (2000). Closer examination of each category showed that content and grammatical word count increased linearly for listeners throughout the story (growing from less than one content word to almost 2 content words and from 1 grammatical word to almost 1.8 grammatical words). Moreover, the proportion of content and grammatical words both increased significantly throughout the narrative (due to more frequent IPUs and to more content and grammatical words within each IPU). This means that listeners not only produce more speech, but that their speech is richer. We can therefore not only confirm what past studies have shown concerning the occurrence of specific responses later in stories, but we have also localized them more precisely within the stories. Whereas previous results were only based on a single measure, our fine-grained temporal analysis provides a good estimation of the continuous temporal increase (evolution) throughout stories. Concerning interjections (with 80% consisted of the items *mh*, *ouais*, *ah*, which are equivalent to generic responses), results showed that their temporal evolution reaches a maximum just before the middle of the story and then decreases until the end. Unlike previous studies showing that they occurred in the first parts of stories (still based on one single measure), our results show more specifically what listeners do in different regions of stories. Firstly, our results showed that *generic responses* tend to be produced throughout the whole story. Secondly, they showed a significant increase before the mid-point of the stories. This is an interesting result directly in line with narrators' results. Regarding these latter, content and grammatical words changed in the same way, specifically there was a marked increase up until the middle of stories followed by a decrease until the end whereas interjections (although produced in a very small number) showed an opposite trend (slightly significant). We interpret these different results in the frame of Labovian phases (1966): just before the mid-point of stories, which we identified as the *apex*, the *orientation* and *complication* phases deal with spatial and temporal components of stories and descriptions of actions or events respectively. The increase in content and grammatical words for narrators could illustrate their greater involvement in the complication phase intended to gather as much information as possible prior to the culminant point (apex) of stories. In parallel the increase in generic responses for listeners is indeed explicit marks of their understanding and more generally of the establishment of common ground. However this finding also corroborates Norrick's claim (2010) that in the phase preceding the apex the narrator produces more words designed to attract listening responses (*response triggers*). Listening responses, in terms of frequency and quality, are then of great importance for the narrator who is about to achieve the culminant point of his/her story. Subsequently, the number of the narrator's content and grammatical words starts to decrease although they are still produced in a very high quantity, with the

narrator remaining the main speaker. Simultaneously, listeners produce less generic responses (decreasing interjections) but the responses become more specific. These results confirm that the apex and the following phases, notably the evaluation phase could be the most conducive to the production of specific responses (Norrick 2010). More generally, our results give a more formal and quantitative support to the proposal that the listener frequently but briefly become a *co-narrator* (Bavelas et al. 2002: 568) by using turns which are more frequent, longer and semantically richer.

The last component examined concerns IPUs containing laughter. Results showed parallel trends for both listeners and narrators: the listeners' laughter reaches a maximum at Nt=0.58 whereas narrators' laughter is more spread out throughout the narrative with a slight maximum at Nt=0.73. This finding confirms the importance of the maximum level in the middle of stories by illustrating that listeners tend to laugh just after the apex phase. Within the scope of expectation of CID's speakers (involving either unusual stories or stories about professional conflict), laughter can appear not only as an appropriate listening response but also as the *preferred response* (Levinson 1983 among others). Whatever the instruction, laughter gives an evidence of a story's tellability (Norrick 2007), and possibly of its success. Laughter reflects that one of the narrator's expectations (to make people laugh) has thus been fulfilled. Furthermore, we also observed that narrators laughed just after listeners, behavior which is the expected: it would be somewhat peculiar if narrators were the first to laugh at their own stories whereas it is highly appropriate for them to laugh along with their listeners.

The whole findings about turn-taking, morpho-syntactic information and laughter occurring in stories, confirm that both speakers contribute to the elaboration of narratives. We can consider that other cues such as prosodic and/or gestural cues could prove useful to enhance our results. However these cues still require a meaning-based approach involving a high level of interpretation (judges' or experts' annotation) that was out of the scope. Finally, the consistency of our results is in favor of the relevance of the procedure of time normalization allowing a more precise understanding of the evolution of production throughout the time narrative, recovering at least the orientation/complication, the apex and the evaluation phases.

Although our interest was in the general behavior of narrators versus listeners in storytelling, it was important to examine the individual characteristics of each speaker. Among the 16 participants of CID, a few of them exhibited deviations from the central trends. Listeners AB and EB had a higher proportion of content words than the central trend, resulting in a rich listening. On the contrary, CM was a very "generic" listener (more than the central trend). CM was also among the speakers who produced many IPUs with laughter. While most listeners exhibit symmetry between content word and interjection proportion, a few do not: while they were within the central trend in the first case, they had an extreme position in the second one (MG BX IM and LL). While MG as a listener produced many generic responses (like CM), her production of content words matched the central trend (unlike CM). Furthermore, the narrators' overall production was more constant than that of the listener: with regard to a standardized narrative activity, the listening activity allows greater inter-speaker variability. Finally, if we look at again CM and MG, as narrators they both had a higher proportion of interjections to the end of their stories. Hence these two speakers (already highlighted as very "generic" listeners) reversed the asymmetry of their interaction when

they were in narrator roles while at the same time becoming true listeners during the co-narration phases. The examination of the individual characteristics also showed three main atypical speakers. Due to her very small number of narratives (2!), LL mainly appeared as a listener. However she was a very "generic" listener (with a very high proportion of interjections) while she produced very few specific responses and little laughter. BX and MG form an atypical dialog in terms of durations of narratives (the shorter ones), a high proportion of interjections for MG whatever her role (narrator or listener) and a high production of laughter for both speakers. It is noteworthy that these two dialogs (LL-NH and BX-MG) brought together the speakers who were the least familiar with each other. We considered that the more intimate speakers would be characterized by the central trend. The task of the CID involved intimate stories. This type of story may be less compatible for speakers who are not so familiar with each other, revealing in the case of these two dialogs a lack of involvement from LL, a cue of anxiety or signs of awkwardness for BX and MG which was highlighted by the high proportion of laughter.

Whatever the individual characteristics, our perspective emphasizes the global collaborative nature of dialogue as a joint action, i.e. co-narration. The question now is how this co-narration deals with the notion of convergence in interaction. Convergence is usually defined as behavior that becomes more and more similar over the time. If we adopt this definition, Figure 5 showed that differences between the speakers tended to decrease throughout the narrative and could thus be interpreted as becoming more and more similar within the storytelling. However storytelling is an asymmetric activity, in which each of participants displays, as we demonstrated here, typical behavior that is intrinsically linked to each of the discursive role's rights and expectations resulting in distinct speech production patterns and turn-taking organization. More generally, the heterogeneity of entire conversations, specifically through the different types of activities involved, does not allow us to, automatically and *a priori*, measure a potential sequence of convergence that is defined solely by such a simple notion of similarity. We argued that interactional convergence sequences require at least, from listeners, alignment via generic responses and affiliation via specific responses (Guardiola & Bertrand 2013). Once this is done, specific responses can sometimes be produced with patterns similar to those occurring in the current discourse from narrators, and then create an interactional convergent sequence. Our present results can be seen as a prerequisite to allow further automatic studies about an interactional convergence sequence achievement. Larger narrative databases could be then investigated by focusing on certain sites (as longer IPUs involving an amount of content words from listener) that would be valuable candidates for the achievement of convergent sequences. Indeed, a potential prosodic or gestural similarity between participants on these sites could be then associated with a true convergence whereas the same similarity measured elsewhere in the narrative may only reveal the co-narration demonstrated here.

6. Conclusion

One of the purposes of face-to-face conversation is the accomplishment of activities such as *storytelling*. This accomplishment is the result of *joint actions* by all the participants (*interactional achievement*). Our findings show that these joint actions are reflected by a typical temporal evolution of speech production of both narrator and listener resulting in co-narration. This co-narration not only affects the nature of

turns, especially the type of listening responses, but also the turn-taking organization. So, the present study enabled us, not only to retrieve the distinction between generic and specific listening responses, but also to improve knowledge about co-narration by highlighting how and where it is systematically performed throughout narratives. It is the recent development of large conversational databases and automatic tools that makes these more systematic and quantitative analyses suitable (including cross-linguistic studies). More generally, we suggest that such analyses would then bring a valuable contribution to understanding the "overall structural organization" of other types of activities such as explanation, argumentation or negotiation occurring in conversation and requiring similar investigation.

Acknowledgments

This research was supported by the Agence Nationale de la Recherche (grant number ANR-12-JCJC-JSH2-006-01). We thank Lauren C. Ponisio for sharing its code to extend the R bootstrap functions to negative binomial models.

Appendix A

Content word	Estimate	Std.Error	P	P_{boot}
Listener				
Intercept	-0.108	0.147	-	-
Nt	0.551	0.139	< 0.001	≤ 0.0025
Nt^2	-	-	0.075	0.092
Narrator				
Intercept	1.482	0.053	-	-
Nt	1.062	0.200	< 0.0001	≤ 0.0025
Nt^2	-1.266	0.194	< 0.0001	≤ 0.0025
Grammatical word	Estimate	Std.Error	P	P_{boot}
Listener				
Intercept	0.054	0.153	-	-
Nt	0.523	0.149	< 0.0001	≤ 0.0025
Nt^2	-	-	0.158	0.150
Narrator				
Intercept	1.555	0.055	-	-
Nt	1.521	0.203	< 0.0001	≤ 0.0025
Nt^2	-1.692	0.198	< 0.0001	≤ 0.0025
Interjection	Estimate	Std.Error	P	P_{boot}
Listener				
Intercept	0.058	0.106	-	-
Nt	0.619	0.360	0.028	0.030
Nt^2	-0.804	0.323	0.012	0.012
Narrator				
Intercept	-1.621	0.118	-	-
Nt	0.795	0.178	< 0.0001	≤ 0.0025
Nt^2	-	-	0.550	0.590

Appendix A shows the sequential comparisons of the token count models by category and narrative role. Nt and Nt^2 were respectively the linear and quadratic form of the normalized start time of the IPU in the narrative. P_{boot} are P-values given by the bootstrap procedure. For example, the first three lines show that for the listener, adding Nt to the intercept-only model improved the description of the data, whereas adding Nt^2 to this second model offered no improvement.

Appendix B

Content word	Estimate	Std.Error	P	P_{boot}
Listener				
Intercept	-0.755	0.126	-	-
Nt	-0.651	0.426	0.002	≤ 0.0025
Nt^2	0.881	0.384	0.022	0.017
Narrator				
Intercept	-0.168	0.022	-	-
Nt	-0.109	0.038	0.005	0.010
Nt^2	-	-	0.120	0.110
Grammatical word	Estimate	Std.Error	P	P_{boot}
Listener				
Intercept	-0.709	0.105	-	-
Nt	0.294	0.094	0.002	≤ 0.0025
Nt^2	-	-	0.110	0.110
Narrator				
Intercept	0.011	0.034	-	-
Nt	0.579	0.156	0.250	0.270
Nt^2	-0.629	0.152	< 0.0001	≤ 0.0025
Interjection	Estimate	Std.Error	P	P_{boot}
Listener				
Intercept	-0.714	0.220	-	-
Nt	0.979	0.498	0.005	0.017
Nt^2	-1.553	0.467	< 0.0001	0.005
Narrator				
Intercept	-3.698	0.130	-	-
Nt	-1.337	0.467	0.186	0.187
Nt^2	2.389	0.460	< 0.0001	≤ 0.0025

Appendix B shows the sequential comparisons of the token proportion models by category and narrative role. Nt and Nt^2 were respectively the linear and quadratic form of the normalized start time of the IPU in the narrative. P_{boot} are P-values given by the bootstrap procedure. For example, the first three lines show that for the listener, adding Nt to the intercept-only model improved the description of the data, as did adding Nt^2 to this second model.