

Listener's responses during storytelling in French conversation

Mathilde Guardiola¹, Roxane Bertrand¹, Robert Espesser¹, Stéphane Rauzy¹

¹Laboratoire Parole et Langage, Université Aix-Marseille, Aix-en-Provence, France

Mathilde.guardiola@lpl-aix.fr, roxane.bertrand@lpl-aix.fr,

Robert.espesser@lpl-aix.fr, stephane.rauzy@lpl-aix.fr

Abstract

This study concerns the evolution of listeners' production during narratives in a French conversational corpus. Using the method of *Conversational analysis* in a first part of the study, we show that listeners use different discursive devices throughout the narrative. In a second part, we attempt to estimate this behavior in a systematic way by measuring the richness of morpho-syntactic categories. We confirm the presence of specific discursive devices as repetition or reported speech produced by listeners in the end of the narrative while only a slight tendency is observed concerning the increasing density of the richer morphosyntactic categories.

Index Terms: back-channel, conversation analysis, storytelling, French, convergence.

1. Introduction

Interactional co-construction in human-human conversation is minimally based on the production of simple backchannel signals, i.e. short utterances produced by the listener to signal sustained attention to the speaker while this latter is talking [1]. This cooperative process is a necessary requirement for a successful interaction [2]. We claim that in some points of the interaction, co-construction can progressively evolve until a real alignment, especially thanks to the production of responses specifically adapted to this point of the interaction. Ratification of the response by the main speaker (by repeating or integrating it in his own discourse) can lead to a particular sequence of high interactional convergence [3], but this is out of the scope.

Our study focuses on storytelling which is a very frequent activity in conversation. Many studies have shown the role of backchannel signals in the turn-taking organization such as [4] [5] among others. However, to our knowledge, a few studies have shown the role of listener in story-telling [6], and more specifically in a systematic way such as [7]. In this latter study, the authors have shown in experimental conditions that listeners become co-narrators and then improve the quality of the story by using adapted responses. These responses are what they called *generic* and *specific* responses. The generic correspond to the simple backchannel signals (such as mh, yeah, ok, and so on) while the specific responses are precisely specific to the current narrative and cannot be produced in another context. To produce specific responses, the listeners need to have enough information about the situation described. This means that they depend on the state of shared knowledge. This shared knowledge increasing throughout the narrative, listeners produce then more specific responses throughout the narrative.

This present study aims to confirm these results in more conversational data in French. In a first stage, we conduct a sequentially analysis inspired from the *Conversational Analysis* framework [8] about the forms and functions of the listener responses throughout the narrative. More particularly, we observe the typical responses produced around the end of the narrative. In a second stage, we attempt to show in a more systematic way that these specific responses are indeed produced later than generic ones. We hypothesize that the different types of responses throughout the narratives could be reflected by measuring the richness of the morphosyntactic categories produced by listeners.

2. Corpus & Methodology

In this study, we considered a subset of the *Corpus of Interactional Data* (CID) [9], i.e. two one-hour long French-speaking dialogues, involving two male participants for the first one (AG-YM) and two female for the other one (AB-CM). In these interactions, participants were told to tell unusual stories. This consign provokes storytelling as a privileged activity in the corpus. Although an experimental setting (record in an anechoic sound-proof room to have a high quality of speech), the interactions are spontaneous (unprepared speech). The participants were not given *a priori* a particular role in the interaction, and they manage themselves the turn-taking organization. So that interactions seem to be very similar to an ordinary conversation.

The Corpus of Interactional Data has been annotated in a multi-level perspective (OTIM Project [10]). All the annotations have been aligned on the signal. Among others, narratives have been identified. In the present preliminary study, we use only morphosyntactic [11] and narrative levels.

We conduct a study combining a double approach (qualitative and quantitative analysis). On the one hand, qualitative analysis consists of a sequential analysis of the interaction. On the other hand, quantitative analysis is in line with corpus linguistics and computational approach [12] for a similar approach in prosody and feedback. We attempt to measure the production of responses of the listener during the story-telling, in a more systematic way, by using a weight that reflects the richness of morphosyntactic categories.

3. Qualitative analysis: generic and specific responses to narrative

A precise sequential analysis has highlighted that during a narrative the listener produces responses which have different functions in the interaction (continuers, confirmation request, assessment among others). These appropriated responses can be

generic or specific [13]. Generic responses can be produced in any narrative (“*ouais/yeah*”, “*mh*”, laughter, etc.). Specific ones are specific for the current narrative, they consist in several devices: question, other-repetition, reformulation, comments and completions of narrative. Within a study on interactional convergence in general, we showed that some particular phenomena (such as direct reported speech and other-repetition) appear mostly in the middle or end of the narrative, when a certain common ground is already established [14].

Most of the specific responses appear in the latest phases of the narrative, often provoked by the apex (culminative point) of the narrative, according to the formal model of narrative based on [15].

Among specific responses, other-repetitions are used by the listener in order to show his participation to the interaction. More particularly, in example 1, this repetition has a savoring function as defined by [16]. By repeating, the listener shows his appreciation of what has been previously said by the narrator:

Example 1.

AP et je galérais un peu sur la sur le bouchon
 AP et si j'étais là je bloquais un peu sur la table
 AP et je vois **une activité animale** sur la table
 LJ @ **une activité animale**
 AP tu sais j'ai vu enfin
 AP dans mon champ visuel y a eu quelque chose tu vois ça s'est mis à bouger oh

*AP and it was hard with the covert
 AP and if i was like this I was looking at the table
 AP and I see an animal activity on the table
 LJ @ an animal activity
 AP you know I saw
 AP in my field of vision there was something you see it started to move oh*

In example 2, AB is the narrator and the listener CM produces a complex back-channel [17]: "oh fuck, excellent". Since the shared knowledge is already constituted, and the narrative is almost finished, CM is able to produce a very specific response to the narrative, showing her understanding of the situation described. She can even produce direct reported speech as a completion of the narrative (in echo [14]), and takes punctually the place of main speaker:

Example 2.

AB il était à moitié allongé par terre a(v)ec sa jambe comme ça en disant **oh j'ai mal j'ai mal j'ai mal**
 CM oh p(u)tain excellent
 AB on a dit on s'en fout on se barre et tout a(l)ors il a quand même réussi
 CM **tu peux crever**

*AB he was half lying on the floor with his leg like this saying oh it hurts it hurts it hurts +
 CM oh fuck excellent
 AB we said we don't care we leave and so on so he still managed to
 CM you can die*

This type of responses, (a subtype of completion) has not yet been described as a reported speech in other studies [14]. We would consider them as complex specific back-channels since they are strongly adapted to the precise situation.

On the contrary, at the beginning of the narrative, listeners can however produce a type of response similar to what [7] consider as a specific response. This latter corresponds mostly to a confirmation or clarification request (about a character, a place or an event involved in the narrative) that precisely helps participants to build the common ground. Example 3 is an illustration of this case:

Example 3.

CM on était complètement euh complètement
 CM désynchronisés en fait ouais c'est ça de euh
 AB et c'était quoi c'était un bar avec euh un un écran ou
 CM c'était un bar mais tu sais un truc vachement moderne alors que tu es dans euh un patelin euh complètement euh

*CM we were totally er totally desynchronized in fact yeah that's it er
 AB and what was it it was a pub with a a screen or
 CM it was a bar but you know a very modern stuff whereas it was in er a small village er totally*

4. Quantitative analysis: morphosyntactic richness of responses

We analyzed two dialogs (2 female speakers, 2 males speakers), with a total of 50 narratives. Each participant produced between 9 and 16 narratives.

Figure 1 shows the narratives durations according to their rank of apparition for each speaker.

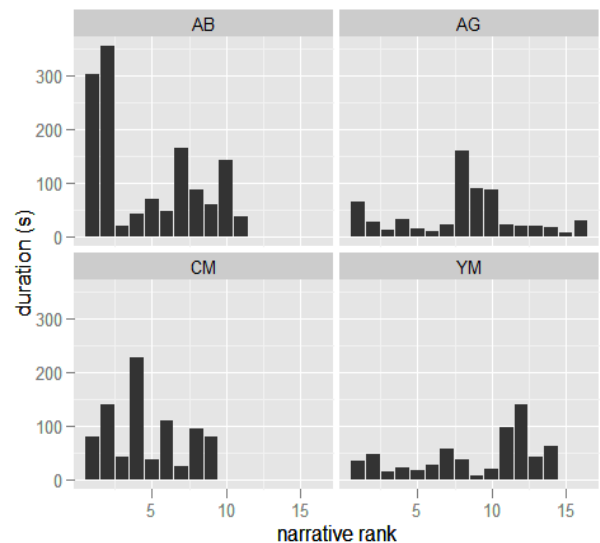


Figure 1: duration of narratives for each speaker

The quantitative analysis allows checking the systematicity of our manual observations. For quantitative analysis, we consider every response from the participant during the progress of the narrative. For this purpose, we take into account every token produced by the listener during a narrative by a main speaker. Time was normalized for the total length of each narrative. So that time is expressed in fractions of the total length of the narrative in which the token appears in. Consequently, we can compare the temporal localization of tokens across the narratives. A token consists in a word or a vocal signal (excluding laughter, since we cannot assign a morpho-syntactic category to laughter, which is a meta-communicative phenomenon [6]). There were 2534 tokens.

We consider the other types of tokens' morphosyntactic category. The CID morphosyntactic annotations were obtained in two steps. In a first stage, the enriched orthographic transcription (adopted in the OTIM project) has been filtered of information that we cannot assign a morpho-syntactic category to, such as laughter or disfluencies, in order to form the input for a modified version of the syntactic parser for written French text StP1 [11]. StP1 has been modified in order to account for the specificities of speech analysis.

Two levels of hierarchy were introduced in the syntactic treatment, corresponding to the strong punctuation marks (such like final point, exclamation mark) and weak or soft punctuation marks (such like comma) that can be found in written text. Lexical entries have also been modified for words playing specific function in speech in interaction, such as vocal back-channel, discourse markers, etc. We found convenient to label these tokens as interjection.

In the second stage the output of the parser was manually corrected for the totality of the Corpus of Interactional Data. Morphosyntactic information is given following the Multext tagset features which contain ten main categories (determiner, adjective, noun, pronoun, preposition, conjunction, auxiliary, verb, adverb and interjection). To account for the morphosyntactic richness, we assigned a "weight" to each token, depending on its grammatical category.

Table 1. Number of tokens by category for 2 dialogs.

Weight	Morpho-syntactic category
Low (973)	Interjection (973)
Medium (797)	Auxiliary (37)
	Conjunction (126)
	Determiner (132)
	Preposition (109)
	Pronoun (393)
High (764)	Adjective (82)
	Adverb (200)
	Noun (167)
	Verb (315)

We assume that "low" category corresponds to simplest responses, whereas "high" categories are used in the morphologically richest responses.

In a first step, figure 2 shows the production of tokens for each weight by the 4 pooled speakers, as narrators (top) and as listeners (low), according to the normalized narrative time.

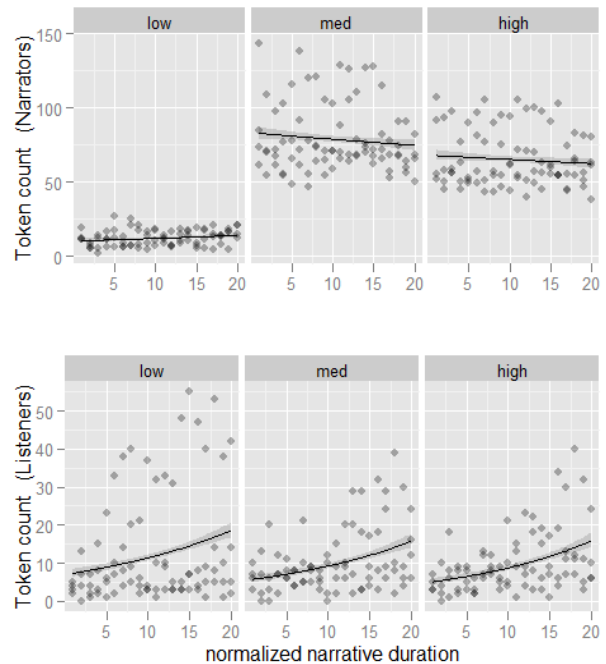


Figure 2: Production of tokens for each weight by the 4 speakers

As expected, the production was roughly stable for narrators with a high count for medium and high weight. On the contrary, the listeners' production increased throughout the narrative. This is in line with the previous section, this increasing reflecting the production of more and complex or specific responses.

Figure 3 presents the boxplots of the start time of every token for each listener, according to their syntactic weight. The distribution of start time seems to be ordered from inferior times to superior times according to the increasing syntactic weight.

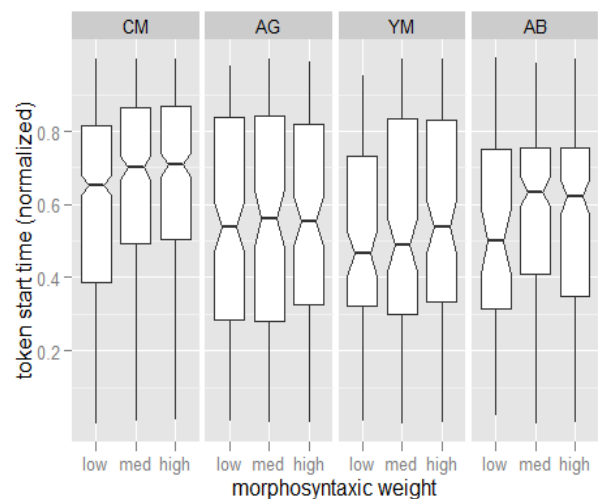


Figure 3: boxplots of start times of every token for each listener, according to their syntactic weight.

A preliminary statistical analysis was run, based on a linear mixed model (package lme4 [18], R software [19]). The dependant variable was the normalized start time of the tokens, the predictors were the syntactic weight (as a 3 level ordered factor), the speakers (as a 4 levels factor) and their interaction. A random intercept was added to account for the variability across the 50 combinations of speakers and narratives. There were 2534 tokens involved. The results show a small but significant time shift, i.e. the higher the morphosyntactic weight, the greater the start time.

Nevertheless, the results were not robust enough to assess firmly the effect. As expected, the shift effect was more robust when only two morphosyntactic weight categories are defined instead of three, (i.e. interjection vs. other).

The durations of the narratives seem to play an important role: the three longest narratives involved in the dialogue AB-CM accounted for 41% of the 1748 tokens of this dialogue; these 3 narratives appear responsible for the presence of the shift effect. For the other dialogue (AG-YM) the results were more robust. This effect is restricted to these 4 subjects, and the size effect is small. This weakness can be attributed to a/ the nature of the measurement, in which morphosyntactic category is a weak measure of the syntactic richness, b/ the automatic procedures of time alignment and morphosyntactic tagging leading to errors.

5. Discussion & Conclusion

This study suggests that the listener responses become more complex, in a discursive and morphosyntactic levels, throughout the narrative in French conversational dialogs. A qualitative sequential analysis shows that the listener produces discursively more complex responses at the end of the narrative. Conversely, in the first parts of this latter, the listeners produce almost simple responses as backchannels that increase towards a more specific type of responses such as reported speech or other-repetitions (among others). A quantitative systematic analysis of responses suggests that the morphosyntactic richness of the listeners' responses increase during the narrative. Results only show a slight tendency but we could improve this richness measurement by accounting for only some narrative formal phases. We have shown indeed in the example 3 that confirmation requests can appear in the first phases of narratives, in order to improve the shared knowledge between participants. But this kind of production (rich in terms of morpho-syntactic categories) in the beginning of narratives, could explain why the tendency is so tenuous. Next step will be to improve the measurement by taking into account this point. For instance, we will suppress some phases as the "parenthesis" phases of narrative that lead to digressions in the narrative structure (as defined by [15]) while they are very frequent in conversational data in which participants can more easily take the floor (even in narratives). Then we will use the available multi-level annotations in order to better take into account the evolution of listeners' responses within formal phases (orientation, complication, evaluation/resolution) of the narrative, and also in investigating prosodic cues (intensity crescendo). At last, we plan to increase the number of speakers and narratives analyzed, while making more accurate the syntactic criteria.

6. References

- [1] Schegloff, E. 1982. Discourse as an interactional achievement: some use of "uh uh" and other things that come between sequences, *Text and Talk*, 71-93.
- [2] Clark, H.H. 1996. *Using Language*, Cambridge: Cambridge University Press.
- [3] Guardiola, M. in progress. "Contribution multimodale à l'étude de phénomènes de convergence dans l'interaction", PhD dissertation, Aix-Marseille Université, France.
- [4] Ward, N. & Tsukahara, W. 2000. Prosodic features which cue backchannel responses in English and Japanese, *Journal of Pragmatics*, 23: 1177-1207.
- [5] Gravano, A., Hirschberg, J., Benus S. 2012, Affirmative cue words in task-oriented dialogue, *Computational Linguistics*, 38: 1, 1-39.
- [6] Þórunn, B. 2005. Feedback in Conversational Storytelling, in *Feedback in Spoken Interaction*, Nordtalk, Gothenburg Papers in Theoretical Linguistics 1-17.
- [7] Bavelas, J. B., Coates, L., & Johnson, T. 2000. "Listeners as co-narrators", *Journal of Personality and Social Psychology*, 79, 941-952.
- [8] Couper-Kuhlen, E. , Selting, M. 1996. *Prosody in conversation*, Cambridge University Press.
- [9] Bertrand, R. , Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., Rauzy, S. 2008. "Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle", *Traitement Automatique des Langues*, 49, 3. 105-134.
- [10] Blache, P., Bertrand, R., Ferré, G. 2009. Creating and Exploiting Multimodal Annotated Corpora: The ToMA Project. In: M. Kipp, et al. (Eds.), *Multimodal Corpora, From Models of Natural Interaction to Systems and Applications*, Springer-Verlag, Berlin, Heidelberg, 38-53.
- [11] Blache, P., Rauzy, S. 2008. Influence de la qualité de l'étiquetage sur le chunking : une corrélation dépendant de la taille des chunks, *Proceedings of the TALN conference*, 290-299, Avignon, France.
- [12] Benus S., Gravano, A., Hirschberg, J. 2011. Pragmatic aspects of temporal accommodation in turn-taking", *Journal of Pragmatics*, 43, 3001-3027.
- [13] Bertrand, R., Priego-Valverde, B, Guardiola, M. 2010. The prosodic cues of humorous reported speech in conversation, AAAL Conference, Atlanta, USA
- [14] Guardiola, M., Bertrand, R. 2011. Mise en évidence de discours rapportés « en écho » dans la conversation, Actes Rencontres Jeunes Chercheurs en Parole, Grenoble, France.
- [15] Labov, W., Waletzky, J. 1966. Narrative analysis : oral versions of personal experience, In J. Helm (ed), Essays on the verbal and visual arts: Proceedings of the Annual Spring Meeting of the American Ethnological Society. Seattle, University of Washington Press, 12-44.
- [16] Tannen, D. 1989-2007. *Talking Voices, Repetition, Dialogue, and Imagery in Conversational Discourse*, Cambridge University Press, Cambridge.
- [17] Laforest, M. 1992. *Le back-channel en situation d'entrevue*. Québec : CIRAL/Recherches sociolinguistiques, 2.
- [18] Bates, D., Maechler, M. , Bolker, B. 2011. *LME4: Linear mixed-effects models using S4 classes*. R package version 0.999375-42. <http://CRAN.R-project.org/package=lme4>.
- [19] R Development Core Team. 2011. *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, <http://www.R-project.org/>.