

# A quantitative view of feedback lexical markers in conversational French

Laurent Prévot

Brigitte Bigi

Roxane Bertrand

Aix Marseille Université & CNRS

Laboratoire Parole et Langage

Aix-en-Provence (France)

firstname.lastname@lpl-aix.fr

## Abstract

This paper presents a quantitative description of the lexical items used for linguistic feedback in the Corpus of Interactional Data (CID). The paper includes the raw figures for feedback lexical item as well as more detailed figures concerning inter-individual variability. This effort is a first step before a broader analysis including more discourse situations and featuring communicative function annotation.

**Index Terms:** Feedback, Backchannel, Corpus, French Language

## 1 Objectives

Conversational feedback is mostly performed through short utterances such as *yeah*, *mh*, *okay* not produced by the main speaker but by one of the other participants of a conversation. Such utterances are among the most frequent in conversational data (Stolcke et al., 2000). They also have been described in psycho-linguistic models of communication as a crucial communicative tool for achieving coordination or alignment in dialogue (Clark, 1996).

The general objective of the project (ANR CoFee: Conversational Feedback)<sup>1</sup>(Prévot and Bertrand, 2012) in which this work takes place is to propose a fine grained model of the form/function relationship concerning feedback behaviors in conversation. The present study is first exploration aiming at knowing better the distribution of these items in one of our corpus. More precisely, we would to verify how much inter-individual variability we will face in further study and whether we can identify a structure in this variability (e.g speaker profiles). Second, we tried

to check there some strong trends in terms of evolution of use of these items in the course of the conversation. This later point was not conclusive and is not developed in this paper.

Some data-intensive works exist for English (Gravano et al., 2012), Japanese (Kamiya et al., 2010; Misu et al., 2011) or Swedish (Allwood et al., 1992; Cerrato, 2007; Neiberg et al., 2013) but not on many other languages such as French for example. On French, the work of (Muller and Prévot, 2003; Muller and Prévot, 2009) concerned a smaller scale (A hour corpus) and very specific task. (Bertrand et al., 2007) was focussed on the feedback inviting cues and also on a smaller scale (2 × 15 minutes). They showed that particular pitch contours and discursive markers play a systematic role as inviting-cues both for vocal and gestural back-channels.

The paper is structured as follow. Section 2 presents the conversational corpus used for this study, then section 3 presents how this corpus has been processed. Section 4 is related to general figures for the feedback lexical items, followed by more detailed information about inter-individual variability (section 5).

## 2 The corpus

The Corpus of Interactional Data (CID) (Bertrand et al., 2008; Blache et al., 2009)<sup>2</sup> is an audio-video recording of 8 hours of spontaneous French dialogues, 1 hour of recording per session. Each dialogue involved two participants of the same gender. One of the following two topics of conversation was suggested to participants: conflicts in their professional environment or unusual situations in which participants may have found themselves. It features a nearly free conversational style with only a single theme proposed to the participants at the beginning of the experiment. This

<sup>1</sup>See the project website: <http://cofee.hypotheses.org>

<sup>2</sup><http://www.sldr.org/sldr000027/en>

corpus is fully transcribed and forced-aligned at phone level. Moreover, it has been annotated with various linguistic information (Prosodic Phrasing, Discourse units, Syntactic tags, ...) (Blache et al., 2010) which will allow us later to take advantage of these levels of analysis.

Numerous studies have been carried out in prepared speech. However, conversational speech refers to a more informal activity, in which participants have constantly to manage and negotiate turn-taking, topic changes (among other things) without any preparation. As a consequence, numerous phenomena appear such as hesitations, repeats, backchannels, etc. Phonetic phenomena such as non-standard elision, reduction phenomena, truncated words, and more generally, non-standard pronunciations are also very frequent. All these phenomena can impact on the phonetization, then on alignment.

### 3 Processing the corpus

The transcription process is done following specific conventions derived from that of the GARS (Blanche-Benveniste and Jeanjean, 1987). The result is what we call an enriched orthographic transcription (EOT), from which two derived transcriptions are generated automatically : the standard orthographic transcription (the list of orthographic tokens) and a specific transcription from which the phonetic tokens are obtained to be used by the grapheme-phoneme converter. From the phoneme sequence and the audio signal, the aligner outputs for each phoneme its time localization. This corpus has been processed with several aligners. The first and main one (Brun et al., 2004) is HMM-based, it uses a set of 10 macro-classes of vowel (7 oral and 3 nasal), 2 semi-vowels and 15 consonants. Finally, from the time aligned phoneme sequence plus the EOT, the orthographic tokens is time-aligned.

The alignment for this paper is another version that has been carried out using SPPAS<sup>3</sup> (Bigi, 2012). SPPAS is a tool to produce automatic annotations which include utterance, word, syllabic and phonemic segmentations from a recorded speech sound and its transcription.

Alignment of items of the list given in (1) were then manually verified. Largest errors were corrected to obtain reliable alignments.

DM pronunciations are the standard ones except

<sup>3</sup><http://www.lpl-aix.fr/~bigi/sppas/>

for a few cases. There are only two items with non standard cases that are over 2 occurrences: sampa: m.w.e.) that is an hybrid between *mh* and *ouais*, and sampa w.a.l.a, a reduction of v.w.a.l.a *voilà*.

The extraction themselves have been realized by the authors with a Python script and all the statistical analyses and plots have been produced with *R* statistical analysis tool.

### 4 Descriptive statistics for the lexical markers used in feedback

All the lexical items of the list given in (1) were automatically extracted and categorized into two categories: (i) *Isolated* items are items or sequence of items surrounded by pauses of at least 200 ms and not including any extra material than the items of this list ; (ii) *Initial* items (or sequence items) are located in front of some other items (but there is no other material within the sequence). Most of these items also occur in final or even *surrounded* positions but we did not consider these cases since they do are not clearly related to feedback. More precisely *surrounded* items are mostly consisting in breaks of disfluencies or genuinely integrated construction (e.g *j'étais d'accord avec lui / I agreed with him*). Final ones can play a role in eliciting feedback or sometimes bring some kind of closure at the end of the utterance (what has been described as Pivot Ending in (Gravano et al., 2012)).

- (1) ah (ah), bon (well), ben (well), euh (err, uh), mh (mh), ouais (yeah), oui (yes), non (no), d'accord (agreed), OK (okay), voilà (that's it, right)

Strictly speaking, the list (1) is not exhaustive. However, other items are already in the thin part of the distribution's tail. Moreover, some of the items such as *euh / err* are not necessarily related to feedback. However, by crossing lexical values with position we expect to get close enough the full set of tokens involved in feedback. For example, initial *euh* not followed by a feedback related item will not be included in the final dataset. This is also an objective of the present work to identify these situations.

The different markers exhibit very different figures with regard to their location as it can be seen in 1. While some are specialized in isolated feedback such as the continuer *mh* which is most of the

time backchanneled, others are found at the beginning of utterances such as *euh*, *ah*. The later makes sense since *euh* is also a filled pause.

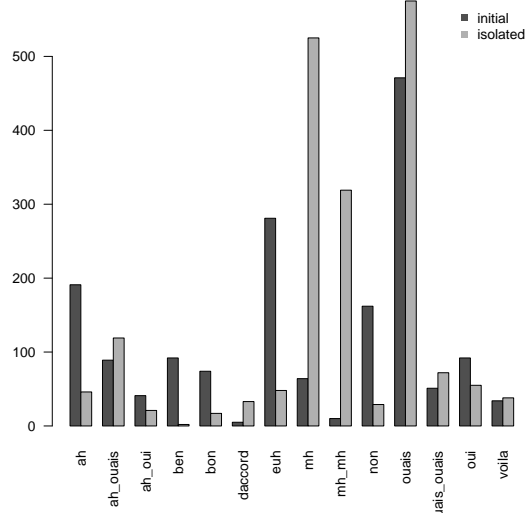


Figure 1: Distribution of isolated vs. initial position for the most frequent lexical items

In total 197 different combinations of the basic markers were identified. The most frequent are the simple repetitions of items such as *ouais* (up to nine times) or *mh*. There are also more complex structures as exhibited in (2) that seem to mix two kinds of items: base ones and *modifiers* (*ah*, *euh*). The base ones seem by default to carry general purpose communicative functions as described in (Bunt, 2009; Bunt, 2012) while the others can also be produced alone but are generally dealing specific dimension such as turn-taking, attitude expression or time management.

- (2) a. ah ouais d'accord ok (*ah yeah right okay*)  
 b. voilà oui non (*that's it yes no*)

With regard to duration, the data is rather messy concerning the very long items. There are extreme lengthening on these units. Aside that and the filler *uh* that exhibit a wide spread, the other items are not produced with huge variations. Monosyllabic remain well centered around 150-250 ms while disyllabic and repeated items are distributed in the 250-500 ms range. This is important for our next step in which automatic acoustic analysis of these items will be performed.

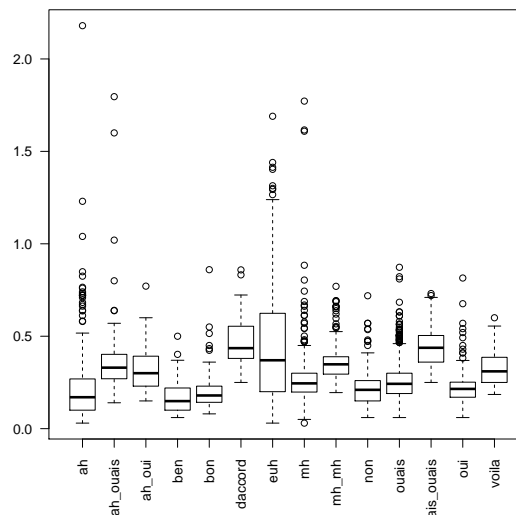


Figure 2: Duration (in seconds) of each lexical type

## 5 Inter-individual variability

Inter-individual variation is a big issue on the way to the generalizability. We would like to understand some of the feedback producing profiles. Our intuitions coming from familiarity of the data is that there are strong variation but they correspond to a few different speaking styles. In future work, we would like to see in a second step whether we can identify and characterize these styles.

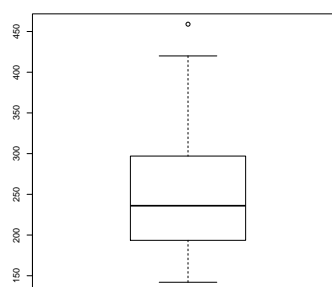


Figure 3: Number of feedback items per speaker

Figure 3 illustrates the total figures of feedback per speaker. As expected variation is huge, from 132 to 425 but with in fact with few outliers with a nice batch of speaker in the 200 – 300 range. The wider spread of the distribution in the high range comes from two factors. First of all, there are participants producing a high quantity of feed-

back items. They produce a massive amount of light backchannels (*mh*, *ouais*) compared to low-quantity feedback producers. The later also produce feedback during the long pauses of the main speaker but they produce much less overlapping backchannels. This should be double checked with a specific measure (adding overlapping as a factor). However, a second effect seems important for at least one speaker (the outlier): the amount to time holding the floor. In fact the speaker producing the most feedback did so because she was rarely the main speaker.

In order to get a global idea of the different uses of these items, Figure 5 represents the proportion of each item per speaker. As expected, the variation is important but one can spot some tendencies. For examples for the most frequent items, the rank seems to be preserved across speakers.

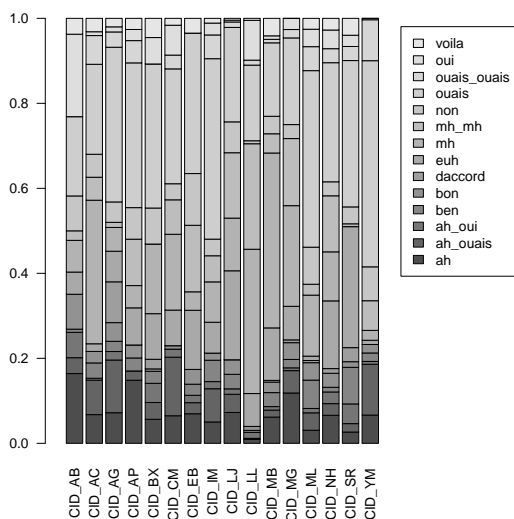


Figure 4: Distribution of the lexical items

Based on their feedback profile (proportion of use of each items as illustrated in Figure 5), we attempted to cluster the participants as showed in 5. While the lower parts of the dendrogram are hard to interpret the higher part matches well with the impression acquired by listening to the corpus (no backchannels and rather formal feedback vs. lots of backchannels and very colloquial style).

## 6 Current and Future Work

About this first batch of analyses, we will complete the analysis of the evolution during the conversation. More precisely, we will go at the individual level looking for time-based changes

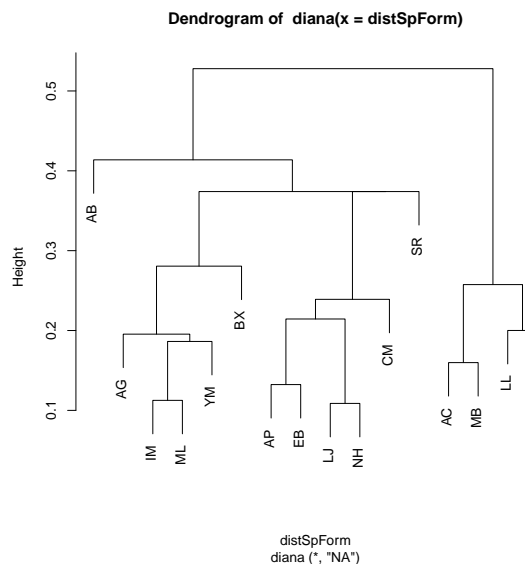


Figure 5: Dendrogram of the participants cluster based on their feedback profile

in their profiles as well as looking at the pairs for tracking potential convergence effect either in terms of distribution of lexical marker types or in their duration.

In parallel to this work, we are launching independent prosodic and kinesic analyses of the forms, as well as a discourse analysis of the functions. Moreover the work is being extended by adding two corpora in the study in order to allow for a better situation generalisability: A French MapTask; and a third corpus consisting in a less cooperative situation. The idea is later to bring together the observations from the different levels in order to propose a multidimensional model for feedback in French dialogues.

Those are steps toward more extensive studies in the spirit of (Gravano et al., 2012) or (Neiberg et al., 2013) on French language and in which we hope to address more directly the issue of discourse situation generalisability.

## Acknowledgment

This work has been realized with the support of the ANR (Grant Number: ANR-12-JCJC-JSH2-006-01) and exploited aligned data produced in the framework of the ANR project (Grant Number ANR-08-BLAN-0239). We would like to thank all the members of these two projects.

## References

- J. Allwood, J. Nivre, and E. Ahlсен. 1992. On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.
- R. Bertrand, G. Ferré, P. Blache, R. Espesser, and S. Rauzy. 2007. Backchannels revisited from a multimodal perspective. In *Proceedings of Auditory-visual Speech Processing*. Citeseer.
- R. Bertrand, P. Blache, R. Espesser, G. Ferré, C. Meunier, B. Priego-Valverde, and S. Rauzy. 2008. Le cid-corpus of interactional data-annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues*, 49(3):1–30.
- B. Bigi. 2012. SPPAS: a tool for the phonetic segmentation of speech. In *Language Resource and Evaluation Conference*, pages 1748–1755, ISBN 978–2–9517408–7–7, Istanbul (Turkey).
- P. Blache, R. Bertrand, and G. Ferré. 2009. Creating and exploiting multimodal annotated corpora: the toma project. *Multimodal corpora*, pages 38–53.
- P. Blache, R. Bertrand, B. Bigi, E. Bruno, E. Cela, R. Espesser, G. Ferré, M. Guardiola, D. Hirst, E. Muriasco, J.-C. Martin, C. Meunier, M.-A. Morel, I. Nesterenko, P. Nocera, B. Palaud, L. Prévot, B. Priego-Valverde, J. Seinturier, N. Tan, M. Teller, and S. Rauzy. 2010. Multimodal annotation of conversational data. In *Proceedings of Linguistic Annotation Workshop*.
- C. Blanche-Benveniste and C. Jeanjean. 1987. *Le français parlé. Edition et transcription*. Paris, Didier Erudition.
- A. Brun, C. Cerisara, D. Fohr, I. Illina, D. Langlois, O. Mella, and K. Smaili. 2004. Ants: le système de transcription automatique du loria. In *Actes des XXV Journées d’Etudes sur la Parole*, Fès, Morocco.
- H. Bunt. 2009. Multifunctionality and multidimensional dialogue act annotation. In *Proceedings of DiaHolmia, SEMDIAL*.
- H. Bunt. 2012. The semantics of feedback. In *16th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL 2012)*, pages 118–127, Paris (France).
- L. Cerrato. 2007. *Investigating Communicative Feedback Phenomena across Languages and Modalities*. Ph.D. thesis.
- H.H. Clark. 1996. *Using language*. Cambridge: Cambridge University Press.
- A. Gravano, J. Hirschberg, and Š. Beňuš. 2012. Affirmative cue words in task-oriented dialogue. *Computational Linguistics*, 38(1):1–39.
- Y. Kamiya, T. Ohno, and S. Matsubara. 2010. Coherent back-channel feedback tagging of in-car spoken dialogue corpus. In *Proceedings of the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 205–208. Association for Computational Linguistics.
- T. Misu, E. Mizukami, Y. Shiga, S. Kawamoto, H. Kawai, and S. Nakamura. 2011. Toward construction of spoken dialogue system that evokes users’ spontaneous backchannels. In *Proceedings of the SIGDIAL 2011 Conference*, pages 259–265. Association for Computational Linguistics.
- P. Muller and L. Prévot. 2003. An empirical study of acknowledgement structures. In *Proceedings of Diabruck, 7th workshop on semantics and pragmatics of dialogue, Saarbrücken*.
- P. Muller and L. Prévot. 2009. Grounding information in route explanation dialogues. In *Spatial Language and Dialogue*. Oxford University Press.
- D. Neiberg, G. Salvi, and J. Gustafson. 2013. Semi-supervised methods for exploring the acoustics of simple productive feedback. *Speech Communication*.
- L. Prévot and R. Bertrand. 2012. Cofee-toward a multidimensional analysis of conversational feedback, the case of french language. In *Proceedings of the Workshop on Feedback Behaviors*. (poster).
- A. Stolcke, K. Ries, N. Coccaro, E. Shriberg, R. Bates, D. Jurafsky, P. Taylor, R. Martin, C.V. Ess-Dykema, and M. Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.