Crossing Empirical and Formal approaches for studying French feedback items

Laurent Prévot and Jan Gorisch
Laboratoire Parole et Langage, AMU, Aix-en-Provence, France

Abstract. Feedback utterances are ubiquitous in dialogue and are identified as a crucial aspect of language interaction. Existing studies about these utterances are either broad communication models that are not systematically related to linguistic data or rich conversational analyses that do not aim at proposing a formal model. The present paper proposes a two-step formal model using fairly standard semantics for lexical item entries and attempting to derive the pragmatic communicative functions from the semantics thanks to rich context modeling. Our formal modeling work is grounded into empirical evidence based on two annotated corpora for a total of more than 10 hours of spontaneous dialogues.

1 Introduction

Following [Bunt, 1994], we take *feedback utterance* as an utterance through which a dialogue participant provides information about his processing of the partner's previous utterances. *Feedback* is among the most frequent dialogue acts in dialogue [Stolcke et al., 2000,Misu et al., 2009]. It is also a crucial aspect for theory of communication and interaction Clark1996. Most of feedback utterances are made of specific lexical markers and of repetitions. Despite this apparent simplicity semantic and pragmatic theories have not provided a comprehensive account for these utterances. Interestingly, only few studies [Allwood et al., 1992,Muller and Prévot, 2003,Bunt, 2012,Ginzburg, 2012] have addressed the problem from a semantic/pragmatic perspective. The present paper, situated within the broader CoFee project[1] [Prévot and Bertrand, 2012], is an attempt at capturing the semantic and pragmatic constraints holding for the choice of lexical makers for producing these feedback utterances. Based on an unpublished corpus study providing our empirical grounding, we focus on couple items in this abstract, propose a semantics for these items and suggest how it can be used in a discourse theory.

[1] `http:\hypotheses.cofee.org`

2 Empirical evidence

2.1 The data

Our empirical study is currently grounded on four different corpora: an 8 hour conversational data corpus [Bertrand et al., 2008] and 2,5 hours MapTask Corpus [Bard et al., 2013], a 2,5 hour face-to-face MapTask Corpus [Gorisch et al., 2014] and a DVD negotiation corpus [2]. We first identified the small set of most frequent items composing feedback utterances by building the token distribution for Inter Pausal Units[3] (IPUs) of length 3 or less. The 10 most frequent forms are : *ouais / yeah (2781), mh (2321), d'accord / agree-right (1082), laughter (920), oui / yes (888), euh / uh (669), ok (632), ah (433), voilà / that's it-right (360)*. The next ones are *et / and (360), non / no (319), tu / you (287), alors / then (151), bon / well (150)* and then follow a series of other pronouns and determiners with frequency dropping quickly. We excluded *tu, et, alors* as we considered their presence in these short isolated IPUs were related to feedback. We then selected all isolated utterances in which the remaining items where represented and treating now each IPUs as an instance. For instance, the utterance *ouais ouais ouais* contributed 3 *ouais* to our initial count and one *ouais ouais ouais* utterance to our dataset. As a result, in the current work *'oui'* is a different form than *'oui oui'* and so far we made no attempt to relate these forms. As a reviewer pointed out these utterances differ radically. Whether a semantics of *'oui'* can be determined to account for both is more a research question of this work than an assumption.

This yielded us a dataset of a bit less than 12000 utterances. We then ran an annotation campaign on this dataset to analyze the form-function relationship for feedback utterances. So far, 4390 utterances from CID and MTR corpora have been cross-annotated. Once reduced to this annotated data and filtered by a frequancy above 50 in one of the two corpora, the dataset with regard to frequently occurring utterances is presented in Table 1.

[2] Accessibles through the Ortolang platform `http://sldr.org/sldr000720/en`, `http://sldr.org/sldr000732/en`,`http://sldr.org/sldr000875/en`, `http://sldr.org/sldr000891/en`

[3] A approximation of utterance when no analysis of the corpus is available yet.

Table 1. Feedback utterances type with frequency > 50 in one corpus

item	translation	#CID	#MT
ouais	yeah	912	588
mh	mh	551	82
mh mh	mh mh	320	245
oui	yes	119	180
d'accord	agreed	35	235
voilà	that's right / that's it	54	170
ah ouais	oh yeah	194	15
ok	ok	5	146
ouais ouais	yeah yeah	106	21
bon	well	65	79
ah oui	oh yes	52	8
laughter	*laughter*	674	54

2.2 Annotation campaign

We asked our annotators to characterize a range of elements that required interpretation (*basic feedback function, scope of the feedback, previous dialogue act type, potential attitude revealed*). We had between 2 and 7 concurrent annotations per item. Intercoder agreement (multi-κ computed with NLTK evaluation metrics API ranged between 0.2 and 0.7 for base function but exhibit large variations for the other dimensions as discussed below. The annotation schema is summed-up in the type feature structure of Figure 1.

$$
\textit{feedback}\left[\begin{array}{l}
\textsc{base}\left\{\textsc{contact, acknowledgment, evaluation, answer, elicit, other}\right\} \\
\textsc{evaluation}\left[\begin{array}{l}
\textsc{approval}\left\{\textsc{approve, disapprove}\right\} \\
\textsc{expectation}\left\{\textsc{expected, unexpected}\right\} \\
\textsc{amusement}\left\{\textsc{amused, annoyed}\right\} \\
\textsc{epistemic}\left\{\textsc{confirmation, doubt}\right\}
\end{array}\right] \\
\textsc{scope}\left\{\textsc{local, global, None}\right\} \\
\textsc{structure}\left\{\textsc{continuation, transition, None}\right\} \\
\textsc{previous}\left\{\textsc{assertion, question, feedback, try , request, incomplete, uninterpretable}\right\}
\end{array}\right]
$$

Fig. 1. Sum-up of the annotation schema for Conversational Feedback

The labels are interpreted in the following way. As for the BASE FUNCTION: CONTACT is a signal of presence and attention, ACK: acknowledgment of receipt, EVALUATION is used for expressing

some kind of attitude or opinion on the content of the target, ANSWER to a question, ELICIT is for utterances that are actually requiring feedback from the other speaker, and OTHER correspond to utterances not clearly associated with feedback[4]. Concerning EVALUATION category, each subcategory correspond to a pair of polarities. (which includes attitudes and emotions values), APPROVE / DISAPPROVE is used for any form of content (dis-)approval / (dis-)agreement, UNEXPECTED / EXPECTED is associated to surprice or absence of surprise, AMUSEMENT / ANNOYED were used for producers amused / bothered by the target, CONFIRM meant that the feedback producer already posses the information targeted. STRUCTURE only splits into CONTINUATION (on the same topic) and *transition* to a different topic[5]. Finally we asked coders to give a dialogue act type to the previous utterance: REQUEST include instructions and other directive moves ; *Try* include confirmation requests ; we will consider other labels of this category to be self-explanatory.

Coding reliability strongly varies from one category to another. The way to evaluate also differs from categories to categories. BASE function is straightforward since there must be a value proposed by each coder and there is no default value. Therefore looking at the κ results should be done while keeping in mind the number of instances actually annotated for a given value. Finally, SCOPE, STRUCTURE and PREVIOUS had default values that have been used massively by naive coders.

Default values The chosen default values are dominant in the data for reducing the annotation burden, but changes to the default are important. We were actually expecting to perform the analysis with the values different than the defaults. What happened is that experts did make the effort to track non-default values while some naive coders facing a relatively complex and repetitive task seemed to forget about it (and focus on categories on which there were clearer decisions to take for each instance). An adjudication procedure is unlikely to change this picture.

Rare categories EVALUATION categories did not have default values but they were derived from on polarity and extended to the other one. In practice, for all these dimensions, among the instances for which a value was proposed in this dimension, one polarity dominated (namely APPROVAL, UNEXPECTED, AMUSED, CONFIRMATION). Moreover, there are a lot less annotations that are distributed (unevenly) in these 4 dimensions and APPROVAL is overall dominant. The others, even the dominant polarity of each pair ended up being rare with regard to amount of data scanned. Agreement on rare categories is specially difficult to reach (because any error on such category strongly damage the intercoder agreement) and therefore scores are low except for APPROVAL (which is not rare), AMUSED and UNEXPECTED that despite begin relatively rare seems to be very reliably identified by coders probably thanks to less unambiguous surface forms. Adjudication on functions that were less well represented but present and with low agreement could make sense because errors are due to ambiguities and to the fact that there are not that many instances to check.

2.3 Final dataset and analysis

To overcome this problem, we calculated a score (# coder for a label / total # coder for this item) for each label and each instance in order to build a reference dataset. We then arbitrarily

[4] Some false starts appeared in the data set due to their inclusion of items such *euh, ouais*.

[5] Originally TRANSITION was split into CLOSURE and PIVOT but low frequency of these categories made us merge them into one.

fixed a threshold of 0.4[6] on this score to consider that a given item was likely to hold a label. Finally, we compared the estimated distribution of the lexical items (the distribution of items for the whole dataset) vs. the distribution of the items holding a given function and checked χ^2 value for these distributions. When the χ^2 yielded significant differences between distributions ($p < 0.05; df = 10$), we compared *estimated* and *observed* percentages for each lexical form and retain for the present abstract deviations over 10%. We ended-up with the correlations presented in table 2. This quantitative filtering does not pretend to exhaust the semantic / pragmatic values of these items. Many other aspects (in particular prosody and multi-modal aspects) are expected to affect these values. Moreover, some uses of these items may occur in different discourse contexts. This analysis is only a way to focus our attention on robust data filtered by naive coder (but native speakers) intuitions on these items.

Table 2. Correlations lexical item vs. function

	BASE FUNCTION	EVALUATION	STRUCTURE	PREVIOUS UTTERANCE
mh	+ack, +contact	-agree		
ouais	+answer, -elicit, -other		+continuation	+try, +question, +request
oui	+answer			+question
voilà		+confirm	+transition	+try
d'accord	+elicit			+request, +feedback
ah ouais		+unexpected		
laughter		+amused		

All these correlations are based on normalised relative distributions. That means that it is not because a function does not appear to be correlated to an item that this item cannot be used this function. It only means that this item does not hold significantly more this function in comparison to the normalised average of all the items. So for example, *ouais /yeah* which intuitively could correspond to AGREEMENT does not have this function in Table 2. This is because the distribution of *ouais* in the function categories does not differ significantly from overall normalized distribution. In this case, AGREEMENT is a function that many of the items concerned have and at a similar percentage. So this table only shows contrasts and is a way to secure that the association studied are robust.

To provide an example, let us focus on the CONFIRMATION category with a threshold of reliability employed later of 0.66[7] The overall percentage of utterances holding CONFIRM label is a bit less than 5%. This is the same percentage for utterances made only of a *'ouais'* or *'oui'*. Despite the fact that *'oui'* and *'ouais'* are used for CONFIRM they do not appear in this table. On the other hand *'voilà'*-utterances hold this label 35% of the time. We give +CONFIRM to *voilà* in the table. Many utterances types (*mh, mh mh, d'accord, okay, laughter*) never received this label but given the overall low frequency of CONFIRM we prefer to stay on the safe side for this first approach.

A more detailed explanation for each lexical item of the table is provided in the following:

[6] This score is just a starting point, it can be adjusted in case results show that higher reliability is required (very likely) or in case that more data with lower reliability is preferred.

[7] We changed the threshold here just to strengthen our point. The new threshold guarantees that at least 2 out 3 coders have use it or for example 2 out 2 did, etc...

- *ouais*: Not strongly associated particularly with any label. This is inline with the intuition of a very underspecified *ouais* (which is also the most frequent of the items).
- *oui*: Used more intensively to answer questions.
- *mh*: Strongly associated with simple acknowledgement. It seems blocked for answering questions, show confirmations or approvals.
- *d'accord, ok*: Strongly associated with simple acknowledgement and seems specific to feedback on requests and other feedback.
- *voilà*: correlates strongly with confirmation and less clearly with *closing*
- *ah ouais, ah*: unexpected
- *laughter*: amusement

3 Lexical items semantics

We will focus on the items *mhm, oui-ouais, okay, d'accord* and *voilà*.

mhm It is quite difficult to provide a semantic content to *mhm*. It seems literally void and, at least for the time being we will consider it to be. This does not means that *mhm* has no impact in the dialogue but its contribution only comes from the corresponding utterance event, the locutionary act and not from its semantics.

$$[\![mhm]\!] = - \tag{1}$$

oui (yes), ouais (yeah) These two items are sometimes very close in pronunciation and transcribers do not agree systematically about their transcription. However, aside marking a difference of register (*ouais* being more colloquial) it has been claimed that do exhibit different functions distribution [Péroz, 2009]. From the quantitative study above, we can only state that *ouais* is well correlated to a broader range of functions and contexts. Their semantics is inspired from the semantics of answers tokens [Krifka, 2001] ($[\![yes]\!] = \lambda P.P$), here expressed a bit differently as carrying an anaphoric propositional (type $\langle t \rangle$) meaning expressed in Discourse Representation Theory in its concise form. *oui, ouais*, once resolved, carry a full proposition meaning and therefore can be associated to a discourse unit.

$$[\![oui]\!] = [\![ouais]\!] = [P_{\langle t \rangle} | P_{\langle t \rangle} =?] \tag{2}$$

d'accord, okay In French *accord* literally means *agreement*. However, *d'accord* can only be a contracted form of *être d'accord (be in agreement with)* and therefore intruduces an eventuality ($\langle evt \rangle$). We therefore propose a semantics with two arguments, the target of the agreement (a proposition) and the syntactic subject (an agent, $\langle agt \rangle$). Following our empirical analysis, *ok*, at least for a first approach we will take it as a synonym of *d'accord*.

$$[\![d'accord]\!] = [\![okay]\!] = [e_{\langle evt \rangle}, x_{\langle agt \rangle}, P_{\langle t \rangle} | P_{\langle t \rangle} =?, x_{\langle agt \rangle} =?, agree(e_{\langle evt \rangle}, x_{\langle agt \rangle}, P_{\langle t \rangle})] \tag{3}$$

voilà *voilà* is the form with the most tricky etymology in our set of discourse markers. *Voilà* is either a defective from of a verb *(see there)* or a preposition. Both forms are present in spoken French but it is the defective form of the verb that is used as a DM. According to our analysis, the key is that *voilà* comes from a contraction of *voir+là (to see there)* which gave it a demonstrative twist.Moreover, *voilà* took an anaphoric resumptive meaning.[8] The speaker using such a demonstrative literally means '*see there that*'. We take this demonstrative nature as way to express that there is direct accessible evidence for backing what is concerned by *voilà*. We will capture this evidence by the predicate *manifest* relative to the speaker.[9]. Similarly to the *oui-ouais* case, *voilà* carries a full proposition and cannot take an extra syntactic subject. We treat it as a discourse unit, leaving presentative uses like *voilà mon chien (here is my dog)* aside.

$$[\![voilà]\!] = [x_{\langle agt\rangle}, P_{\langle t\rangle} | P_{\langle t\rangle} =?, Spk(x_{\langle agt\rangle}), manifest(x_{\langle agt\rangle}, P_{\langle t\rangle})] \quad (4)$$

4 Discourse usage

In this section, we show how to use such semantics to derive the correct discourse use of these items. We adopt a SDRT (Segmented Discourse Representation Theory) [Asher and Lascarides, 2003] approach because it allows for a clean division between literal and communicative meaning. This is achieved thanks to integration of participant inferences (that can involve non-linguistic ingredients) and discourse representation coming from different modules. It also provides rich mechanisms for modeling scope of backward-looking acts such our feedback.

4.1 Continuers

In example (1) feedback items do not even ground the semantic content of what the main speaker says. Although in general such lack of grounding is difficult to assess we made it explicit here by a statement showing the previous lack of attention of the main speaker.

(1) A: You take ///the/// second right ///Then/// first left (t_3)
B$_1$: ///mh///(t_1) ///mh///(t_2)
B$_2$: Wait a second, I was not listening (t_4)

We present in Table 3 the evolution of each participant linguistic representation (as SDRSs) and of the observable common ground that we consider here as directly accessible to the participants.

In this game board, K_1 and K_4 are derived through standard compositional semantics ; $K_2 = K_3 = -$; and we do not analyze precisely K_5.

Uttering *mhm* only results with the locutionary act without any meaning associated. However, an effect is to signal that its producer is here at the utterance time. Based on this fact, given a minimum of cooperativity, one can infer that the interlocutor is ready to hear what is coming next.

Formally, the following rules are assumed to be part of speakers's communicative skills, modeled here as a module expressed in SDRT's Glue Logic. A crucial propery of Glue Logic is to have two logical implications : $\rightarrow$ standard material implication, and $>$ a default implication ($A > B$ reads as *Normally, If A then B*) [Asher and Morreau, 1991].

[8] We will not discuss whether this meaning could actually be derived from the first one.
[9] $manifest(x, P)$: Participant x has positive evidence (for example, he saw it) for P

t	$SDRS_A$	$\mathcal{OC}$	$SDRS_B$
t_1	$\emptyset$	*utter*(e_{1a}, A, u_1)	$\emptyset$
		utter(e_{1b}, B, u_2)	$\pi_2 : K_2$
t_2	$\pi_1 : K_1$	*utter*(e_{2a}, A, u_3)	
		utter(e_{2b}, B, u_4)	$\pi_3 : K_3$
t_3	$\pi_4 : K_4$	*utter*(e_3, A, u_5)	
t_4		*utter*(e_4, B, u_6)	$\pi_5 : K_5$

Table 3. Gameboard evolution for example (1)

Glue Logic 1 (Presence) $\forall e, x, y$ *utter*$(e, x, y) \rightarrow$ *present*$(x, STE(y))$[10]

Glue Logic 2 (Attention) $\forall x, y$ *present*$(e, x, y) >$ *ready*(x, y)

In our example, we infer therefore that the simple observation of the utterance acts lead to $present(B, STE(u_2)$, $ready(B, STE(u_2))$ and $present(B, STE(u_4)$, $ready(B, STE(u_4))$ among many other *ready* information.[11]

Following [Lascarides and Asher, 2009], this case can be handled at discourse relations level through the *Acknowledgment** meta-talk relation. This relation only acknowledge the target speech act without grounding the content. Although being an interesting proposal, we consider that *continuers* can concern any dialogue activity or unit (e.g word or phrase), not necessarily a full speech act.

4.2 Acknowledgment

In example (2), the feedback utterance simply grounds the meaning of the previous turn. Here the turn is composed of two discourse units related by a *contrast* relation.

(2) A: le code tu le reçois mais c'est en deux temps *(the code, you receive it but it is in two steps*
B: ouais *(yeah)*
Corpus CID

The game board evolution is presented in the table. The first steps are standard monological interpretation. The constituents π'_x are the version of the interlocutor of the constituent π_x. There are several way of dealing with these constituents. [Lascarides and Asher, 2009] assume they are equally accessible to both speaker and therefore has a single label. Here we prefer to keep them distinct. They can be seen has a way to access directly the other participant discourse structure or as a partial copy for it. In case of long monologues constituting a turn, this copy could be reduced to a discourse topic summarizing the content of the monologue.

[10] $STE(x)$ represent the spatio-temporal extension of x.

[11] In practice, this rule generates too many $ready$ signals. We should limit these inferences to time point where complete processing chunks (elementary discourse units) are produced and also integrates prosody that could act as more specific (and therefore overruling our default inference) signal to take turn.

$SDRS_A$	$SDRS_B$
$\pi_1 : K_1$ $\pi_2 : K_2$ π_{12} : *contrast*(π_1, π_2)	$\emptyset$
	$\pi_3 : K_3$ *ack*(π'_{12}, π_3)

Fig. 2. Gameboard evolution for example (2)

In this game board, K_1, K_2, K_{12} are derived through standard compositional semantics ; K_3 = $[P|P\ =?]$. Here P is resolved as $K_1 \wedge K_2 \wedge \phi_{contrast(\pi_1,\pi_2)}$. Therefore K_1, K_2 and $\phi_{contrast(\pi_1,\pi_2)}$[12] are becoming part of the common ground.

To infer acknowledgment discourse relation, we check that the utterance attached is redundant with its target, following [Walker, 1992]. The triggering of this rule is straightforward in our example. In this exemple $models$ sands for the standard entailment relation. In this case it means that the content of α is logically can be entailed from the content of β.

Glue Logic 3 (Inferring Ack) *speaker*$(\alpha) \neq$ *speaker*$(\beta) \wedge K_\beta \models K_\alpha >$ *Ack*(α, β)

Once the acknowledging role the discourse unit has been established, its contribution has been modeled. There is therefore no need for semantic constraints on *Acknowledgment*. It is however useful to specify that it is a subordinating relation since it should not affect which discourse accessibility.

Definition 1 (Acknowledgment). *Ack*$(\alpha, \beta) \rightarrow \emptyset$
Ack is a subordinating relation (no structural constraint)

4.3 Approval

In example (3), the target of the feedback is an imperative. The agreement leads to some kind of commitment in action from the feedback producer. Agreement targets can also be assertive leading to propositional commitments [Hamblin, 1970].

(3) A: Tu descends tout droit. (you go straight down)$(t1)$
B: ok $(t2)$
Corpus MTR

The game board evolution is similar to the previous one except for the discourse unit of the feedback (π_2). However, the discourse relation used is the same.

K_2 = $\lambda x, P\, \exists e$ *agree*(e, x, P) resolved into *agree*(e, B, K_1). Here the commitment is an action commitment.

4.4 Confirmation

In example (4), the follower of a route explanation dialogue is either taking the lead or recapitulating some previous explanation. The giver only confirms what is asserted by the follower.

[12] The semantic constraints of the $contrast(\pi_1, \pi_2)$ relation.

t	$SDRS_A$	$\mathcal{OC}$	$SDRS_B$
t_1	$\pi_1 : K_1$	*utter*$(e_1, A, u1)$	$\emptyset$
t_2		*utter*$(e_2, B, u2)$	$\pi_2 : K_2$ *Ack*(π'_1, π_2)

Table 4. Game board evolution for example (3)

(4) F: ouais je contourne la la mine par le haut à gauche *(yeah I turn around the the mine through the left upper part)*
G: voilà *(that's right)*
Corpus MapTask

$SDRS_A$	$SDRS_B$
$\pi_1 : K_1$	$\emptyset$
	$\pi_2 : K_2$ *confirm*(π'_1, π_2)

Fig. 3. Game board evolution for example (4)

Following our lexical semantics for *voilà*, $K_2 = [I, P | P =?, manifest(I, P)]$ resolved into $manifest(B, K_1)$. Infer *Confirmation* is similar to *Acknowledgment* inference. However, there is an extra need of the content targeted to be manifest from the feedback producer perspective. As a result *Confirmation* entails *Acknowledgement* but force a coordinating nature for it.

Glue Logic 4 (Inferring Confirmation) *speaker*$(\alpha) \neq$ *speaker*$(\beta) \wedge K_\beta \models K_\alpha \wedge$ *manifest*(*speaker*$(\beta), K_\alpha) >$ *Confirm*(α, β)

Confirmation is a way to validate some information and once confirmed they should be how removed from the focus of attention. This is captured by considering *confirmation* as a coordinating relation. Moreover *Confirmation* has for semantic effects to require that the producer of the confirmation knows about the element confirmed before its interlocutor said it as detailed in the formula below[13].

$$\textit{confirm}(\alpha, \beta) \rightarrow \mathcal{K}^t_{\textit{speaker}(\beta)} K_\alpha \wedge t <_t \textit{time}(\alpha) \tag{5}$$

Transition uses of *voilà* are proposing to close the current discourse topic. However, we do not need a specific treatment for handling this case. The coordinating nature of the *confirm* relation provides the closing effect desired. Therefore all confirmations are removing their target from the *Right Frontier*. When the target unit is a sequence, the whole sequence is closed (at least in the representation of the closure producer) thanks to the coordinating nature of the $confirm$ relation.

[13] In which $\mathcal{K}^t_x \alpha$ stands for *'x knows* α *at time* t*'* and $time(\alpha)$ is the time of the uttering event corresponding α discourse unit.

(5) A : Ensuite tu arrive place des Carmes et c'est la rue au fond à gauche *(Then you arrive on the Carmes square and it is the street across the square on the left)*
B : et c'est quel numéro *(and what is the street number?)*
A: c'est au 27 rue des Polinaires. *(it is at the 27 Polinaires street.)*
B: ouais je vois en fait c'est rue des Polinaires. *(yeah I see in fact it is Polinaires street)* (t_n)
A: voilà *(that's it)*(t_m)
Corpus Apéro-Toulouse

t	$SDRS_A$	$\mathcal{OC}$	$SDRS_B$
...	...	...	...
t_n	π : ...	...	
t_m		*utter*(u_m, A, s_m)	$\pi_m : K_m$ *confirm*(π, π_m)

Fig. 4. Gameboard evolution for example (5)

5 Conclusion and current work

We have a proposed a quantitative analysis of French feedback items grounded on two different communicative situations. Although related to [Allwood et al., 1992], we provide more empirical footing (for French language) to values of feedback. On the empirical side, the work can be seen as a replication of some asepcts of [Muller and Prévot, 2003,Muller and Prévot, 2009] with three differences: addition of a discourse situation, 7 naive coders (instead of 2 over-trained experts), scaling up the experiment (current annotated dataset of 4390 items). On the formal side, we tried to give feedback items a precise underspecified meaning that can be used at the discourse level for deriving correct way of functioning in context. As the formal aspects, future consists in further developing the model to include on repeated items, their semantics and possibly their compositional building. A similar extension work is required about co-verbal gestures and non-lexical items. Overall, the development of the empirical side on this work is also going toward more data annotated and analyzed.

Acknowledgments

This research was supported by the ANR project "Conversational Feedback": ANR-12-JCJC-JSH2-006-01. Many thanks to all the participants and our coders of the annotation campaign (undergrad and master students of the Language Science departement).

References

Allwood et al., 1992. Allwood, J., Nivre, J., and Ahlsen, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.

Asher and Lascarides, 2003. Asher, N. and Lascarides, A. (2003). *Logics of conversation*. Cambridge University Press.

Asher and Morreau, 1991. Asher, N. and Morreau, M. (1991). Commonsense entailment: A modal theory of nonmonotonic reasoning. In *Logics in AI*, pages 1–30. Springer.

Bard et al., 2013. Bard, E. G., Astésano, C., D'Imperio, M., Turk, A., Nguyen, N., Prévot, L., and Bigi, B. (2013). Aix MapTask: A new french resource for prosodic and discourse studies. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody*.

Bertrand et al., 2008. Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., and Rauzy, S. (2008). Le cid-corpus of interactional data-annotation et exploitation multimodale de parole conversationnelle. *Traitement Automatique des Langues*, 49(3):1–30.

Bunt, 1994. Bunt, H. (1994). Context and dialogue control. *Think Quarterly*, 3(1):19–31.

Bunt, 2012. Bunt, H. (2012). The semantics of feedback. In *16th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL 2012)*, pages 118–127, Paris (France).

Ginzburg, 2012. Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford University Press.

Gorisch et al., 2014. Gorisch, J., Astésano, C., Bard, E., Bigi, B., and Prévot, L. (2014). Aix map task corpus, the french multimodal corpus of task-oriented dialogue. In *Proceedings of The 9th LREC*.

Hamblin, 1970. Hamblin, C. (1970). *Fallacies*. Methuen.

Krifka, 2001. Krifka, M. (2001). For a structured meaning account of questions and answers. *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow*, 52:287–319.

Lascarides and Asher, 2009. Lascarides, A. and Asher, N. (2009). Grounding and correcting commitments in dialogue. *Journal of Semantics*.

Misu et al., 2009. Misu, T., Ohtake, K., Hori, C., Kashioka, H., and Nakamura, S. (2009). Annotating communicative function and semantic content in dialogue act for construction of consulting dialogue systems. In *INTERSPEECH*, pages 1843–1846.

Muller and Prévot, 2003. Muller, P. and Prévot, L. (2003). An empirical study of acknowledgement structures. In *Proceedings od Diabruck, 7th workshop on semantics and pragmatics of dialogue, Saarbrucken*.

Muller and Prévot, 2009. Muller, P. and Prévot, L. (2009). Grounding information in route explanation dialogues. In *Spatial Language and Dialogue*. Oxford University Press.

Péroz, 2009. Péroz, P. (2009). On ne dit pas ouais! *Langue française*, (1):115–134.

Prévot and Bertrand, 2012. Prévot, L. and Bertrand, R. (2012). Cofee-toward a multidimensional analysis of conversational feedback, the case of french language. In *Proceedings of the Workshop on Feedback Behaviors*. (poster).

Stolcke et al., 2000. Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Ess-Dykema, C. V., and Meteer, M. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics*, 26(3):339–373.

Walker, 1992. Walker, M. A. (1992). Redundancy in collaborative dialogue. In *Proceedings of the 14th conference on Computational linguistics-Volume 1*, pages 345–351. Association for Computational Linguistics.