CHAPTER 6

FRENCH FACE-TO-FACE INTERACTION: REPETITION AS A MULTIMODAL RESOURCE

Roxane Bertrand[1], *Gaëlle Ferré*[2], *Mathilde Guardiola*[1]
[1]Aix-Marseille Université, CNRS, Laboratoire Parole et Langage,
Aix-en-Provence, France,
roxane.bertrand@lpl-aix.fr; mathilde.guardiola@lpl-aix.fr
[2]Laboratoire de Linguistique et Université de Nantes,
Nantes, France
Gaelle.Ferre@univ-nantes.fr

1. Multimodal analysis of human communication and interaction

Human-human interaction implies numerous studies to significantly improve the efficiency, naturalness and persuasiveness in human-computer interaction (HCI) systems. But there is still inadequate knowledge on what and how cues interact in face-to-face interaction. The complexity of human-human interaction involving the description of verbal and non-verbal modalities still needs theoretical and empirical foundations. To achieve this goal, researchers need to develop resources and tools that enable them to take into account the different modalities. Verbal, vocal and gestural cues have been studied separately for a long time. This favored the precise description of the mechanisms and rules governing each domain. But today the question of how these various cues in the different modalities are connected, has become important for linguists.

In this study, we present the perspective adopted in the national OTIM project (Blache et al. 2009) which aimed to precisely answer some of the issues raised in multimodality in French face-to-face interaction. To achieve the global aim of the project, i.e. to better understand how the different linguistic levels interact, several steps were necessary, among which the specification of a standardized way of representing multimodal information, the development of generic and reusable annotated resources based on the elaboration of a multimodal annotation schema, the development and/or the adaptation of different annotation tools (see http://www.lpl-aix.fr/~otim/ for details of conventions, tools and annotations).

Drawing on the Corpus of Interactional Data (CID, Bertrand et al. 2008), the project involved different steps, from the development of the various coding schemes in the different modalities to the annotation and analysis.

The corpus itself is an audiovisual recording of 8 hours of French conversational dialogs. The recording of the corpus was born out of an interest in human interaction based on a very fine-grained analysis of each linguistic domain and their relationships. Such an analysis in the phonetic domain requires a semi-experimental setting with a high quality of recordings enabling the acoustic analysis of speech. In the same way, the gestural level requires a particular setting, both in terms of lighting, framing and placement of the speakers in respect to each other and to the camera. The various recordings should be comparable and the frame chosen for the recording should allow good visibility of fine movements as well as larger ones made by the speakers. At the same time, conversational analysis requires yet consider other criteria such as the level of (in)formality, the symmetric or complementary status between participants, the absence of pre-determined discursive role of participants, the presence/absence of a third party to regulate turn-taking, etc. This corpus affords a good balance between the elicited and very controlled corpora usually used by phoneticians or prosodists until recently and 'natural' conversational data analyzed in the field of Conversational Analysis (Couper-Kuhlen & Selting 1996) on which the present study on repetition is drawing.

In this latter framework, the authors claim that every aspect of talk-in-interaction is collaboratively accomplished through participants' ongoing negotiations in situ (Szczepek Reed 2011: 8). In the same way, the collaborative model of Clark (1996) defines conversation as a *joint-action* implying a coordination of actions by participants at the level of content and at the level of process. Joint-action is achieved through different phenomena in interaction, among which backchannel signals, but also collaborative or competitive turn completion. Repetition naturally contributes to the co-construction of interaction as it supposes that one of the participants is taking into account what has been produced by the other at a certain time. Repetition then supposes some kind of adaptation in between participants to an interaction.

The use of such terms as adaptation as well as alignment (Garrod & Pickering 2004, Pickering & Garrod 2006), accomodation (Giles et al. 1987) or mimicry (Kimbara 2006 among others) to quote but a few studies, refers to convergence phenomena. In the Interactive-alignment Model of Dialogue (Pickering and Garrod 2006), the alignment observed at one level is automatically extended to other levels, resulting in a similarity at the level of discourse or gesture and at the level of representations. For the authors, this alignment is the basis of successful communication in dialog. Adaptation refers to the fact that participants adapt their responses to the other interactant(s)' productions. In the Communication Accomodation Theory (CAT) (Giles et al. 1987), adaptation and accomodation can be used indifferently. Speakers are tailored to their partners (*adaptive behavior*) to affiliate with their social status for example. Mimicry is a direct imitation of

what the other participant produces (exact match at prosodic level, Couper-Kuhlen and Selting 1996; exact or very close match at gesture level, Jones 2006). A discussion about the relevance of one term or another is out of the scope here (see Guardiola, in progress)[1]. It can nevertheless be noted that the choice of a term varies according to the linguistic field (psycholinguistics, sociolinguistics, phonetics,…) but also the modality considered in the type of study (see Section 3).
In this chapter, after presenting the corpus as well as some of the annotations developed in the OTIM project, we then focus on the specific phenomenon of repetition. After briefly discussing this notion, we show that different degrees of convergence can be achieved by speakers depending on the multimodal complexity of the repetition and on the timing in between the repeated element and the model. Although we focus more specifically on the gestural level, we present a multimodal analysis of gestural repetitions in which we met several issues linked to multimodal annotations of any type. This gives an overview of crucial issues in cross-level linguistic annotation, such as the definition of a phenomenon including formal and/or functional categorization.

2. Corpus & annotations

A multimodal analysis of interaction requires the encoding of many different pieces of information, from different domains, with different levels of granularity. All the information has to be connected and synchronized (with the signal for example). Different steps in the annotation were adopted in the OTIM project to achieve this goal. Before presenting the annotation process and some of the annotations used in this study on repetition, it is important to consider that the project aimed not only to provide and develop conventions and tools for multimodal annotation but also to define the organization of annotations in an abstract description from which a formal XML scheme could be generated (Blache & Prévot 2010).

2.1 Corpus

For a few years, numerous programs have been conducted in different countries to provide large-scale spontaneous speech interactions involving the creation and development of resources (in terms of both corpora and annotations). Among others, one can mention the Map-Task corpus (Anderson et al. 1991) that is one of the first semi-elicited corpus and which has been reduplicated in many languages, the Columbia Game Corpus

[1] For further details: see SPIM ANR-08-BLAN-0276-01: http://spim.risc.cnrs.fr/

(http://www.cs.columbia.edu/speech/games-corpus/), the Buckeye Corpus (Pitt et al 2005), the Corpus of Spontaneous Japanese (Furui et al. 2005), DanPASS (the Danish phonetically annotated spontaneous speech corpus, Grønnum 2006), as well as corpora annotated at the gestural level such as the Göteborg Spoken Language Corpus (Allwood et al 2000), the MIBL Corpus (Multimodal Instruction Based Learning, Wolf and Bugmann 2006) or the D64 Corpus (Campbell 2009), among others (for a more exhaustive list, see Knight 2011). The *Corpus of Interactional Data* (CID) (Bertrand et al. 2008) described here is an audiovideo recording of conversational French (eight dialogs of 1 hour each, 110.000 words). Participants were filmed by a single camera and recorded with a head-set microphone (one track by speaker, in order to enable the acoustic analysis of speech and segments produced in overlap by both speakers). Participants were asked to tell about conflicts or unusual events in their personal lives.

2.2 Corpus transcription

The first and most important step in the annotation process is transcription because most of the annotations in the different domains are based on this particular level.

In a preliminary stage, the speech signal was automatically segmented in inter-pausal units (IPUs), speech blocks surrounded by 200 ms silent pauses. The transcription process takes this series of IPUs as input. The transcription conventions adopted in the project derive from the ones defined by the GARS (Blanche-Benveniste & Jeanjean 1987). They take into account some remarkable and frequent phonetic phenomena: non-standard elisions, phoneme substitution or additions, assimilation phenomena, word truncation, silent pauses, filled pauses as well as some specific phenomena such as the pronunciation of schwas in Southern French and laughters. From this initial transcription, two versions were generated: i/ a *standard orthographic transcription* from which the orthographic tokens are extracted to be used for semantics, syntax and discourse analysis and their related tools (POS tagger, parser, etc) and ii/ a *phonetic transcription* from which the phonetic tokens are used in the next steps of *grapheme-phoneme conversion* and alignment presented below.

The enriched orthographic transcription is time consuming: three passes have been made for each speech file. In the first one, the entire transcription was made by one transcriber. The second and third passes involved a correction of this first transcription. However, it guarantees a faithful transcription and improves the phoneme/signal alignment.

The grapheme-phoneme converter is a dictionary and rule-based system (Di Cristo & Di Cristo 2001); it takes a phonetic token sequence extracted from the transcription as input and provides a sequence of phonemes as output. From this, the aligner assigns each phoneme its time localization.

This aligner is HMM-based (Brun *et al*. 2004), and relies on acoustic models based on standard French. The alignment is done for each IPU separately. In a first pass, labeling was automated. A second pass involving hand-correction of vowel boundaries was conducted on two speakers. From the time-aligned phoneme sequence and the enriched orthographic transcription, the orthographic tokens are also time-aligned.

From this tokenization and its alignment on the signal, a wide range of annotations have been conducted in each of the different domains: prosody (phrasing, pitch contours), morphosyntax and syntax, discourse and interaction (discursive units, reported speech, disfluencies, backchannel signals, etc.[2]). Not all the annotations will be fully described here, since there have been many in several linguistic fields and not all of them are relevant to the present study.

2.3 Morphosyntactic annotations

Morphosyntactic annotations were done in two steps. In a first stage, the enriched orthographic transcription was filtered of information to which no morphosyntactic category could be assigned, such as laughter or disfluencies, in order to form the input for a modified version of the syntactic parser for written French text (Blache & Rauzy 2008). This was then modified in order to account for the characteristics of spoken French. In a second stage, the output of the parser was manually corrected for the totality of the CID. The annotation process is time-consuming whether it is manual or automatic. The manual annotation requires several annotators (either expert or not, sometimes both) and tests of labeling consistency to measure inter-annotator agreement. The automatic annotation is less time-consuming but also requires evaluation between the different tools or involves manual corrections, which enable to evaluate the performance of the parser (only 5% of error rates).

2.4 Prosodic annotations

The prosodic level can be annotated in a manual or an automatic way depending on whether we observe rather phonological (more abstract) phenomena or phonetic parameters. In OTIM, we focused on the prosodic phrasing which corresponds to the structuring of speech material in terms of boundaries and groupings. The manual annotation is very time-consuming but was necessary to improve the knowledge of prosodic domains in French. In a first stage, such a manual annotation made by experts enabled us to test the robustness of annotation criteria. A previous study involved two experts; results have shown a very good inter-coder

[2] More details are provided on http://aune.lpl.univ-aix.fr/~otim/

agreement and kappa score for the higher level of constituency (IP) (Nesterenko et al 2010). In a second stage, the elaboration of a guideline for transcribing prosodic units in French by naïve annotators enabled us to test the reduplicability of these annotation criteria. Naïve transcribers have to annotate 4 levels of prosodic break defined in terms of a ToBI-style annotation[3] (0 = no break; 1 = AP break; 2 = ip break; 3 = IP break) in Praat (Boersma & Weeninck 2009). The global aim is to develop this phonologically-based transcription system for French that would be consistent enough to be amenable to automatic labeling. One of the step is to compare the manual annotations. Another step consists in improving existing automatic tools (such as *Intsint* for example, Hirst et al. 2000) by comparing the output of different annotation tools and manual expert/naïve annotation (Peskhov et al. 2012).

At last, another aspect of prosodic annotation concerns the intonation contours associated to intonational or intermediate phrases (levels 3 and 2 above). Pitch contours are formally and functionnaly defined (Portes et al. 2007 for details). Intonation contours were coded for 6 speakers.

2.5 Gesture annotation

90 minutes of the CID involving 6 speakers were coded for gestures. We manually annotated hand gestures, head and eyebrow movements as well as gaze direction with Anvil (Kipp 2001).

Different typologies have been adopted for the classification of hand gestures, based on the work by Kendon (1980) and McNeill (1992, 2005). The formal model we use for the annotation of hand gestures is adapted from the specification files created by Kipp (2004) and from the MUMIN coding scheme (Allwood et al. 2005). Both models consider McNeill's research on gestures (1992, 2005).

The changes made to existing specification files only concerned the organization of the different information types and the addition of a few values for a description adapted to the CID. For instance, we added a separate track 'Symmetry'. In case of a single-handed gesture, we coded it in its 'Hand_Type': left or right hand. In case of a two-handed gesture, we coded it in the left Hand_Type if both hands moved in a symmetric way or in both Hand_Types if the two hands moved in an asymmetric way. For each hand, the scheme and annotation file in Anvil has 10 tracks.

2.5.1 Functional categories

The gesture types we annotated are mostly taken from McNeill's work. *Iconics* present "images of concrete entities and/or actions", whereas

[3] http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html

Metaphorics present "images of the abstract", they "involve a metaphoric use of form" and/or "of space". (McNeill 2005: 39). *Deictics* are pointing gestures and *Beats* bear no "discernible meaning" and are rather connected with speech rhythm (McNeill 1992: 80). *Emblems* are conventionalized signs and *Butterworths* are gestures made in lexical retrieval. *Adaptors* are non-verbal gestures that do not participate directly in the meaning of speech since they are used for comfort. Although they are not linked to speech content, we decided to annotate these auto-contact gestures since they give relevant information on the organization of speech turns. For gesture phrases, we allowed the possibility of a gesture pertaining to several semiotic types using a boolean notation.

$$
\begin{bmatrix}
GestureUnit & & \\
Symmetry & & \\
Hands\_Type & \begin{bmatrix}
\textsc{Phase} & \\
\textsc{Phrase} & \\
\textsc{Lexicon} & \\
\textsc{HandShape} & \\
\textsc{HandOrientation} & \\
\textsc{GestureSpace} & \\
\textsc{Contact} & \\
\textsc{Movement} & \begin{bmatrix} MovementTrajectory \\ MovementVelocity \\ MovementAmplitude \end{bmatrix}
\end{bmatrix} &
\end{bmatrix}
$$

Figure 1: Formal model for the annotation of hand gesture

2.5.2 Descriptive annotations

A gesture phrase (i.e. the whole gesture) can be decomposed into several gesture phases i.e. the different parts of a gesture such as preparation, stroke (the climax of the gesture), hold and retraction (when both hands return to rest) (McNeill 1992). The scheme presented in figure 1 also enables us to annotate gesture lemmas (Kipp 2004:237), the shape and orientation of the hand during the stroke, suppress gesture space (where the gesture is produced in space in front of the speaker's body, McNeill 1992:89), and contact (hand in contact with the body of the speaker, of the addressee, or with an object). We added three tracks to code the hand trajectory (adding the possibility of a left-right trajectory to encode two-handed gestures in a single Hand_Type, and thus save time in the annotation process), gesture velocity (fast, normal or slow) and gesture amplitude (small, medium and large). A gesture may be produced away from the speaker in the extreme periphery, while having a very small amplitude if the hand was already in this part of the gesture space.

Head and eyebrow movements, as well as gaze direction and global facial expressions (laughters and smiles) were annotated as well, although not all the items projected in the coding scheme provided in Figure 2 were

noted.

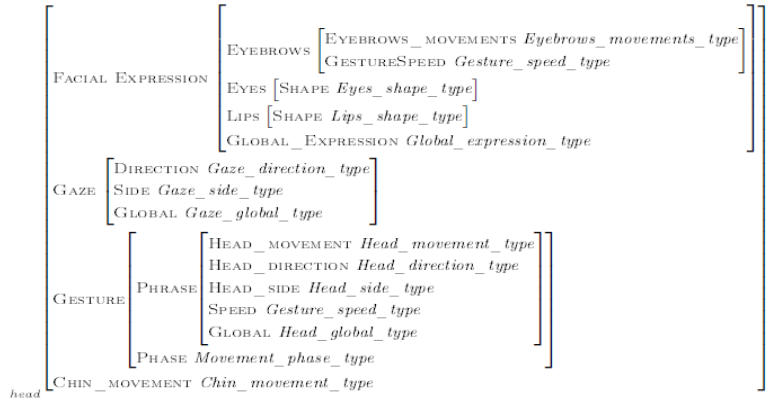

Figure 2: Formal model for the annotation of head and eyebrow movements, gaze and facial expressions

3. Repetition

3.1 Theoretical background: a general definition

To take a rather objective term, "repetition" in interaction has been observed by many researchers working in different fields. Chartrand & Bargh (1999), but also Garrod & Pickering (2004) argue that repetition is needed to make conversations easier or more fluent and that speakers align "their representations at different linguistic levels at the same time" (2004:9), thus reducing the processing load for each participant in a conversation. Several terms have been coined to refer to repetition that do not, however, necessarily refer to the same process. Working on gesture and sound repetition of adults by infants, Jones (2006:3) distinguishes between emulation, a repetition of an "outcome produced by a model, with no requirement that the actual motor behavior should match that of the model", and mimicry (the most widely used term in multimodal studies), "a behavior that matches or closely approximates the movements of another". She also goes against the general view that mimicry is innate and contends instead that it is a learned behavior that arises in infants around 18 months and that is the result of infants being mimicked by their caregivers. An infant's response to an adult protruding their tongue with the same movement would not be imitation according to her but rather a general response to any interesting stimulus. The debate is out of the scope of this paper, but Garrod & Pickering (2004) adopt a similar view when they describe conversations amongst adults, as they establish a sequential link between primed representations – what is called by Chartrand & Bargh (1999) the chameleon effect, i.e. a perception-behavior link – which lead to

imitation, which in turn leads to alignment of representations (op. cit., 9), or what is termed elsewhere convergence. They illustrate this sequential process with 'yawning': if one sees someone yawn, one yawns in return (primed representation), but one also tends to feel more tired or bored (imitation and alignment of representations with the initial yawner). They also mention that the whole process is unconscious and largely automatic but "is also conditional to the extent that it can be inhibited when it conflicts with current goals and purposes, or promoted when it supports those goals" (op.cit, 10). At last, both Garrod & Pickering (op. cit.) and Shockley et al. (2009) mention that alignment does not mean that speakers have to be in agreement in a conversation and that it is rather a process speakers use to simply understand each other. Tabensky (2001:217), however, states that what she calls echoing "can be merely a sign of co-presence, and not necessarily an indication of understanding or alignment with the speaker's proposition".
The consensus is larger on the fact that a conversation can be considered as a joint action (Garrod and Pickering 2004; Holler and Wilkin 2011; Kimbara 2006; Shockley et al. 2009; Tabensky, 2001, to cite but a few studies) which entails the co-construction of meaning by all the participants to the interaction as suggested in the introduction.

3.2 Behavior and gesture repetition

There has already been quite a large body of work on the role of behavior and gesture-pattern repetition and on the conditions for their emergence. Lakin *et al.* (2003) observe that some situations activate a desire to affiliate in the participants to an interaction and thus encourage mimicry. This work is derived from Chartrand & Bargh (1999) who noted the social role of mimicry. On experimental data, they observed that postures and adaptors (in their study, the shaking of one's foot) were regularly mimicked by the participants, and that when the confederate mimicked the participants, the latter felt greater empathy with the model. Working on posture and gaze, Shockley *et al.* (2009) find that similar gaze patterns emerge in participants together with the increase of joint understanding. Also working on experimental data, they observe that participants adopt more postural coordination when they see the same words on a screen than when the words are different. Mol *et al.* (2009) go further on experimental data as well. They find that reproduced iconic gestures are not just imitation: only gestures that are consistent with verbal content are copied. Much in the same vein, Holler & Wilkin (2011) find that mimicked hand gestures in experimental conditions play an active role in the grounding process and help create mutually shared understanding. Their classification of mimicked gestures is both semantic and formal. To count as repeated, a gesture has to represent the same meaning or have the same referent, use

the same mode of representation and have the same overall form. Drawing on data from a joint narration task, Kimbara (2006:45) adds that temporal proximity together with co-referentiality between a gesture and its repetition show "realizations of a shared image construal". Besides, she observes that not all the features of gestures are repeated, but a subset has to be present in the repetition for the gesture to be considered as mimicked. From her study, she concludes that hand-gesture mimicry creates *gesture catchments* (McNeill, 2001) across speakers.

In a later study, Kimbara (2008) notes that gesture repetition is not a chance phenomenon. In experimental conditions, she notices that participants produce gestures which are more similar in terms of handshape when they can see each other than when they cannot. Parrill & Kimbara (2006), also working on experimental data, note that observing mimicked gestures induces more mimicry in the participants. They consider a gesture is repeated when two of the following features are reproduced: motion, handshape or location. Hand-gesture features are also central in Mol *et al.* (In press) who state that imitators in laboratory speech are influenced in their mimicry by features of the original gesture, for instance handedness. They show as well that participants are influenced in their repetition of hand gestures by the cognitive perspective adopted by the confederate (like the description of items on a map from a vertical or a horizontal viewpoint).

Instead of focusing on exact matches between the verbal and gestural productions of participants to conversations in three languages, Tabensky (2001) describes what she calls *rephrasings*, namely how speakers mimic some semantic features while adding new features at the same time. She is also concerned with temporal alignment of the productions and what forms a language unit. From her corpus, she observes that in some instances, the semantic features which formed a package in speech and/or gesture in the original production are separated into different units in the rephrasing (a process she terms *separation*), whereas in other instances, semantic features expressed in several units in the original production are merged into a single unit in the rephrasing (a process she terms *fusion*).

At last, von Raffler-Engel (1986) observes full or partial gesture imitation in *transfers*, namely the gestures made by an interpreter into another language in consecutive translation. She describes gesture repetition in terms of "repetition of parts to the whole" and determines a series of *components* which must be proportional to the model for a gesture to be considered as repeated but need not be identical: muscular tension, gesture duration and movement extension, a series of components that we also consider in the present study. Other gesture characteristics have to be identical to the model to give an impression of sameness. With this in mind, she notices that in many instances, interpreters retake the gestures they observe in the speaker they translate, instead of changing the original gestures in the re-packaging of information involved in a translation. Yet,

she mentions that the gestures produced in the translation were judged natural enough by native speakers of this language, so that no culturally inappropriate gesture was copied into the translation.

3.3 Verbal other-repetition

Verbal other-repetitions (henceforth OR) consist in repeating a word or a sequence of words that have been previously uttered by another interactant. This process leads to a lexical similarity of the participants' speech that can be analyzed as a means to align with the interlocutor. OR have been identified as forming an important mechanism in face-to-face conversation through their discursive or communicative functions (Norrick 1987, Tannen 1989, 2007, Perrin *et al.* 2003). According to Tannen (2007), participants notably use lexical repetition to show their involvement in the interaction. She argues that repetition is useful at several levels of verbal communication: production (easier encoding), understanding (easier decoding), connection (better cohesion in discourse), and interaction (repetition maintains the link between participants). Repetition can also be considered as a *specific* form of feedback (in the sense of Bavelas et al. 2000). Perrin (2003) proposes a four-function typology for other-repetitions, nearly corresponding to backchannel functions: taking into account, confirmation request, positive reply and negative reply. More largely, repetition functions as a device for getting or keeping the floor (Norrick 1987).

3.4 Prosodic repetition

In a similar way to the verbal or gestural level, the main issue raised by prosodic repetition is to know when one speaker's repetition of a prosodic pattern can be considered as mimicry (Couper-Kuhlen & Selting 1996: 366). More recently, this question is addressed by Szczepek Reed (2006) through the notion of prosodic orientation that refers to the "interactional orientation whereby (…) speakers display in their sequentially "next" turns an understanding of what the "prior" turn was about" (Hutchby & Wooffitt 1998: 15). Szczepek Reed defines several types of prosodic orientation such as the prosodic matching (copy) of the previous speaker's prosodic design, the complementation of a prior turn with a second structurally related prosodic design or a continuation of the previously unfinished prosodic pattern. Gorisch et al. (2012) report some works on pitch matching and interactional purpose. To the authors' credit, they provide precise définitions of different terms often used in the same way. They propose to consider that prosodic matching is used for continuing the project in hand, aligning or affiliating with the previous speaker. In line with Stivers (2008) and Barth-Weingarten (2011), they distinguish between the terms 'alignment' and 'affiliation' that have been used until recently in an indefinite way. Alignment refers then to the endorsement of the

sequence/activity in progress and contrasts with the notion of affiliation which refers to the endorsement of the previous speaker's evaluative positioning, or stance (cited by Gorish et al 2012: 7). More precisely, Stivers (2008) uses the term of alignement to describe actions by a second speaker which support the activity being undertaken by the first speaker. In this way, the production of backchannel signals in conversation can be considered as adapted and expected responses from the listener during conversation (Bertrand et al. 2007; Heldner et al. 2010 among others) and more particularly in a storytelling activity in which the main speaker (narrator) is indeed ratified as main speaker by the listener (Stivers 2008). At last, Gorish et al.'s study also constitutes a first attempt to develop a method enables the measurement of the acoustic similarity (in terms of f0 and intensity) of pitch contours in naturally occurring data. Similar parameters were considered by De Looze et al. (2011) in a study on prosodic convergence in spontaneous conversations. In Gorish et al.'s study, the prosodic matching observed is then considered as a resource used to demonstrate alignment with the prior action. In a similar way, Bertrand & Priego-Valverde (2011) have shown that a copy of some prosodic cues by both participants could be a resource to demonstrate orientation to a humorous utterance expressed by the speaker. A series of prosodic matching repetitions by both participants is leading to the creation of a short sequence called *joint fantasy* (Kotthoff 2006).

4. Identification of multimodal repetition

4.1 Gesture repetition

The criteria we adopted for the repetition of hand gestures were very much inspired from von Raffler-Engel (1986). For a hand gesture to be considered as repeated, gesture phrase, lexicon and movement trajectory have to be identical, that is the functional category of the gesture, whereas other descriptive features like gesture space, tension, amplitude and velocity do not have to be strictly identical to the model for the gesture to be considered as a repetition of the model. The criteria for the repetition of head movements and gaze orientation are stricter than for hand gestures since we considered a gesture was repeated if the movement in the repetition was strictly identical or mirrored between the repetition and the model.

4.2 Lexical repetition

In the same way as for gesture repetition, for a lexical repetition to be considered as repeated, different formal and functional criteria are involved.

First of all, we proposed to formally define verbal repetition as the production of a word or phrase that has already been uttered by another

speaker. We specify that too frequent words cannot be considered as repeated, in order to avoid 'accidental' similarities in discourse.

Annotations concerning lexical other-repetition were made in two steps: a first automatic output, followed by a manual correction by two experts. The first automatic stage, based on the transcription of tokens, allows the detection of potential other-repetitions. It is based on a set of rules and on relevance criteria themselves based on word frequency (for each speaker). The rules were elaborated during a previous study (Bigi et al 2010).

A preliminary processing transforms the words into lemmas, containing no morphological mark of conjugation or plural. A set of two rules is then applied on the data: Rule 1- An occurrence is accepted if it contains one or more 'rare' words, (the rarity is measured on the vocabulary of the speaker who makes the repetition). Rule 2 – An occurrence which contains at least 5 words is accepted if the order of words is strictly identical in both speakers' discourse.

The tool locates co-occurrences of relevant lemmas: the words which were uttered by a speaker in an IPU, and by the other speaker in a simultaneous or in a previous IPU.

Obviously, these formal criteria are not sufficient to only select occurrences of other-repetition. Following Perrin et al. (2003), a repetition has to have an *ostensive* character to be considered as a real repetition (intention of quotation). Then only an expert analysis enables to eliminate co-occurrences that are not other-repetitions. The tool, however, greatly reduces the amount of time necessary for the detection of repetitions. In a last step, two experts checked the speech segments identified by the tool as possible repetitions on the basis of formal criteria. 350 consensual cases were then retained in the CID.

4.3 Prosodic repetition

In the same way as for gesture or lexical repetition, for a prosodic repetition to be considered as repeated, the repetition and the model have to be identical, either at a phonetic level (duration, fundamental frequency) or at a phonological level (phrasing, pitch contours, or rhythm pattern). Following Szczepek Reed (2006), we use the term *prosodic matching* to refer to the repetition of a prosodic pattern.

5. Analysis

First, it must be noted that *other-repetition* is not quite frequent in our corpus when considering gesture. Although we have seen in the introduction and section 3 that conversations constitute joint activities in which meaning is co-constructed by participants, this co-construction does not necessarily entail gesture repetition. Gesture repetition is therefore probably dependent of many other factors like topics, collaborative tasks,

but also the conversational history of the participants. Indeed, Tabensky (2001) mentions that gesture repetition occurs pretty much at the beginning of the conversations she works on, in which participants are not acquainted with each other. In our corpus, participants know each other quite well and therefore have a long conversational history which could explain why less gestural adjustment is needed between them. Tabensky (op. cit.:232-233) also states that "textual repetition of words and retakes with small adjustments are generally not accompanied by gesture", which is also what we find since there is a much vaster quantity of verbal repetition and hardly any at all involve gesture repetition. However, we will see in this section that when gesture repetition is present, it can be quite complex especially when related to verbal and prosodic repetition in a multimodal perspective.

5.1 Repetition as a confirmation request

In example 1 below, speaker A describes the hard work he had to put in to redecorate a room in his house since the walls were coated with several layers of wallpaper and paint. An approximate translation of the example is given in italics and the transcription conventions are provided at the end of this paper.

Example 1

1 A: c'était tapissé peint [alors c'était l'enfer quoi] *the wall was papered, painted, so it was a nightmare*

2 B: [ah ouais avant que t'enlèves les couches] *oh yes, before you can take off all the layers*

3 A: oh putain j'ai mis j'ai mis un mois quoi /mh/ *oh fuck it took it took me a month /mh/*

4 A: enfin bon c'est un mois en faisant que le week-end si tu veux /ah ouais/ mais *well a month but I was only working on week-ends you know /oh yeah/*

5 A: tu vois là tout à la ra[clette] *you see, all of it with a scraper*

{Gesture 1 ----------}

6 B: [c'était peint sur la ta]pisserie [truc comme ça] *it was painted on the paper and stuff*

{Gesture 2 ----------}

7 A: [et ouais] *yes it was*

Figure 3a. Iconic gesture produced by speaker A (Gesture 1) in example 1.

Figure 3b. Iconic gesture reproduced by speaker B (Gesture 2) in example 1.

Figure 3a illustrates the gesture made by speaker A while saying *all of it with a scraper*. He produces an iconic gesture as if he was holding a scraper, a gesture in which his right hand goes up and down twice. Speaker B repeats the iconic gesture adopting a mirror perspective in a double up and down movement of his right hand with the same hand shape, although the gesture is not as high in the gesture space as Gesture 1 and a bit more towards his side instead of being in front of his chest. Whereas Gesture 1 lasts 1.60 seconds, Gesture 2 in figure 3b is slightly shorter with a duration of 1.52 seconds, yet the difference in length between the two gestures is not dramatic. What is interesting though is their temporal alignment. Speaker B prepares for the gesture just as speaker A is preparing for his second "scraping" gesture and the stroke of Speaker B's gesture is beginning 3 frames before the end of Speaker A's stroke. This means that at the same time as he is producing his gesture, speaker B is attending to speaker A's own gesture. He cannot know at the beginning of his gesture that the model will stop after the second "scraping movement".

The example illustrates a certain multimodal parallelism between gesture and speech: the case does not correspond to the fusion nor to the separation described by Tabensky (2001) since the repetition contains two pieces of information pertaining to two different modalities. The iconic gesture repeats the information of the scraper, whereas the utterance produced by B is not an exact repetition of what was said by speaker A, although some information is similar. At the prosodic level, however, there

is a similarity. The two utterances considered here form one intonational phrase (IP) each (the IP of Speaker B starting before the end of Speaker A's turn, with an overlap of 0.655 s). Both IPs present similar configurations. We can note three things that are particularly striking though: first of all, Speaker A produces a slightly emphatic accent on "tout" (*all*) which is realized as a slight reinforcement of the initial plosive /t/. The same reinforcement is met in the initial plosive of the emphatic word "peint" (*painted*) for Speaker B. Towards the end of the IP we can see a similar list pitch contour even if the second utterance (B) could be considered as a confirmation request. Speaker B seems indeed to ask confirmation that he understood well when saying "it was painted on the wallpaper" as the first verbal mention of the utterance by speaker A (line 1) did not make it explicit that the coat of paint had been applied onto the wallpaper (the utterance "the wall was papered, painted" could be understood as a chronological description of two actions with no link between them, not necessarily as the wallpaper being painted).

Nevertheless, the prosodic matching is also expressed by the strong lengthening associated with the last syllables of words "raclette" (*scraper*) and "tapisserie" (*wallpaper*), which was described by Portes et al (2007) as the main cue of the list contour. And at last, Speaker A adopts a flat trailing contour around 135 Hz on the whole phrase which is also copied by Speaker B with the same F0 height, although Speaker B generally has a much lower voice than Speaker A.

The match which occurs both at the gestural and at the prosodic level is interesting in two respects: first, considering the fact that the two utterances do not constitute the same kinds of speech acts – Speaker A's utterance is a statement, whereas Speaker B's could be a confirmation request – the two utterances would probably have had completely different prosodic contours in another context. Then, because of the overlapping speech, it means that Speaker B is copying prosodic information while Speaker A is still speaking and this exactly matches the pattern we have for gesture since there was also a gesture overlap in between the model and the copy.

5.2 Repetition as a hedge

Just before the extract below, two male participants were discussing a school experience one of them had. His teacher was very strict and forbade the children to leave class. Once, he needed to go to the bathroom, didn't dare to ask the teacher and messed himself. As his mother actually worked in the school as a teacher, he went to see her. In the example, speaker A, after acknowledging the narrative with a backchannel, asks if the mess showed in a verbally elliptical utterance ("parce que t'étais tout", *because you were all*). The question is, however, not exactly elliptical as it is completed by a gesture, which is repeated by speaker B in his answer.

Example 2

1 A: parce ce que t'étais tout (0.075) *because you were all*
{gesture 1 ----------}

2 B: non ça se voyait peut être /non/ je me rappelle plus trop /ouais/ mais je crois pas que ça se voyait mais bon euh @ ça ça devait sentir tu vois *@ and then, no perhaps it didn't show /no/, I don't quite remember /yeah/, but I think it didn't show, but uh it must have smelt you see*
{gesture 2 --}{gesture 3 ----------}{gesture 4 ----------}

Figure 4a illustrates the metaphoric gesture produced by speaker A who starts with both hands slightly rising from his lap and places them in the lower periphery, palms oriented towards his body. He then extends his hands away from his body thus representing the extent of the mess. The whole gesture from the beginning of the preparation phase to the end of the retraction lasts 0.96 second. 1 frame before the end of the retraction of Gesture 1, speaker B initiates a repetition of the gesture (Figure 4b), yet the two gesture strokes are not in overlap. The difference between the two gestures as illustrated in the figures is that Gesture 2 is much shorter (0.60 second) than gesture 1, the movement is not as ample and the fingers are much more relaxed than those of Gesture 1.

Figure 4a. Metaphoric gesture produced by speaker A (Gesture 1) in example 2.

Figure 4b. Metaphoric gesture reproduced by speaker B (Gesture 2) in example 2.

Figure 5. Second reproduction of metaphoric gesture by speaker B (Gesture 4) in example 2.

What is interesting, however, is that immediately after Gesture 2, B produces an emblem (Gesture 3) – which is of no particular interest here – without any retraction of Gesture 2, and then produces Gesture 4 without retracting his hands. Gesture 4 happens to be a second repetition of Gesture 2. This time, it is slightly longer than the first repeat (0.92 second) and the gesture features are more similar to the original production of the gesture as illustrated in Figure 5. In this second repetition, the amplitude of the movement is slightly larger than in the first repetition and finger tension is also greater. What is different from the original gesture is the position of the hands: the palm of the left hand is facing up instead of the hand being on its side.

What can be added in this example is that both Gesture 1 and Gesture 2 are not redundant with the message content. When speaker A produces Gesture 1, he is anticipating some assumption on the part of speaker B who was narrating what happened at school. The gesture in this context completes what is left unsaid in the elliptical utterance, probably out of decency. Although speaker B repeats speaker A's gesture (with Gesture 2), he contradicts the verbal assumption, so that the gesture which was consistent with the initial verbal message is repeated (as such gestures tend to be repeated as pointed out by Mol *et al.* 2009), but then becomes quite inconsistent with the answer. A gesture linked to the syntactic negation would rather have been expected here. In this example, the prosodic level is in accordance with the content of the utterances. In the first turn, speaker A formulates the elliptical question with a low and trailing pitch and a very strong lengthening on the last word that is typically used in unfinished turns. By contrast, the next turn produced by B, starting with the answer "no", exhibits a rising-falling contour while at the same time speaker B is repeating the gesture previously produced by A in the first turn. Gesture 4 is also a repetition that comes together with the repetition of the contradiction and this looks as what has sometimes been called a *hedge*, i.e. a way of softening a contradiction, contradictions being generally not preferred by interactants. It is interesting to note that the same rising-falling

contour is again produced by Speaker B on "voyait" (*showed*) (which is also the second repetition of this word) as Speaker B is once again repeating Speaker A's gesture. Therefore, we can say that there is a complete dissociation between the double repetition of the other participant's gesture by Speaker B, and the contrast expressed both in verbal content with two negations and prosody with the self-repetition of the contrastive prosodic contour. Speaker B then develops with "mais ça devait sentir" (*but it must have smelt*), an utterance which is later repeated as well as a self-confirmation.

5.3 Cross-repetition

In example 3 below, the two speakers are discussing the arrangements speaker A will make to look after the baby his wife is expecting.

Example 3

1 A: bien par exemple t'façon Laure elle est prof *well, for example, anyway, Laure is a teacher*
2 A: donc elle travaille pas tu vois /ouais/ tout le tem[ps] *so she doesn't work, you see /yeah/ all the time*
3 B: [to]ut le temps ouais *all the time yeah*
4 A: puis à ce moment là les matinées où où elle est au au co- si elle doit aller au collège (0.62) *so then on the mornings when when she is at at scho- if she must go to school*
5 A: [bien moi moi je reste ici je prends euh enfin je m'en occupe] *so I I stay here, I take uh, well I look after it*
{Gesture 1}{Gesture 2}{Gesture3}{Gesture 4 ----------
6 B: [ouais toi tu restes ouais ouais vous euh] *yeah yeah you stay yeah yeah you uh*
{Gesture 5}

Example 3 is slightly different from what we have seen above. As he utters *so I*, speaker A produces a metaphoric gesture (Gesture 1) which is an asymmetrical double-handed gesture (the left hand moves a bit more than the right one), and which consists of his hands being oriented palm up at the beginning of the stroke. Then he rotates his wrists, so his hands are on the side and opens them a little again. Speaker B has a similar movement of his right hand only although the configuration of his fingers is different (Gesture 5). The twisting movement is actually what makes the gesture look as a repetition of A's metaphoric. This is shown in Figures 6a and 6b.

Figure 6a. Two-hand metaphoric gesture produced by speaker A (Gesture 1) in example 3.

Figure 6b. Single-hand metaphoric gesture produced by speaker B (Gesture 5) in example 3.

In terms of temporal alignment, Gesture 5 starts 0.88 second after Gesture 1, yet, contrary to what we saw in example 1, the repeat is much longer than the model (0.88 sec vs. 0.36 sec) and is also more complex as well. Whereas Gesture 1 is only composed of a stroke because it is part of a series of gestures which we will not describe here since they are not relevant to the present study, Gesture 5 contains a preparation and a retraction. If we consider the stroke only, then the repeat is shorter than the model as it lasts 0.32 second.

One may consider that there is a redundancy between gesture and verbal repetition in this example with no new information added. However, the repetition plays a role in the message structure as it is the global pattern which is repeated including words and gesture and which has a function of backchannel. The whole extract is very collaborative: speaker A produces some argumentation as to who will take care of the baby and speaker B collaborates to the argumentation first producing the backchannel signal "ouais" (*yeah*), then repeating "tout le temps" (*all the time*) and repeating his own *yeah* again. His whole utterance forms a complex backchannel with a function of acknowledgement. At the prosodic level, the model and the repeat are clearly distinct at least because of the location in the IP and the discursive function of each one. A produces "tout le temps" in the end of the IP with a major terminal rising contour (about 100 Hz) followed by a high plateau while B produces "ouais tout le temps ouais" as a single IP

with a minor rise on "temps" (around 30 Hz). At this point, the two speakers seem to be reaching the end of a conversational sequence which could be the reason why their overall pitch is so low. The configuration of this repetition exhibits a compressed span as it is often the case in backchannels. Concerning the gesture repetition (Gesture 5), although it is quite clear in the video that Speaker B is repeating Speaker A's gesture, the timing in the verbal modality is rather a repetition of B's speech by A. In fact, B's "restes" (*[you] stay*) begins 0.197 s before A's "reste" (*[I] stay*), so that when Speaker A is beginning to utter "reste", he has enough acoustic material to know what is being said by B. B's "restes" is much longer so that speech rate is not similar for the two speakers. The lengthening on "restes" by Speaker B directly corresponds to the lengthening of his copied gesture. However, according to the location of the word "reste" in the two repetitions, their contour is not quite the same (see Figure 7). Whereas Speaker B's contour is a low plateau at the end of an IP, followed by another IP ("ouais") A's "reste" is in the middle of the IP that ends on "ici" expressed with a rise. The configuration of the repetition that also functions as an acknowledgement is in accordance with the previous verbal repetitions of this sequence.

This analysis provides a good example of cross-repetition: whereas one of the participants is repeating the other's gesture, the other participant repeats the first one's speech in terms of verbal content. Prosody is matching between the two speakers only in the fact that they both use compressed span as projecting the end of a conversational sequence.

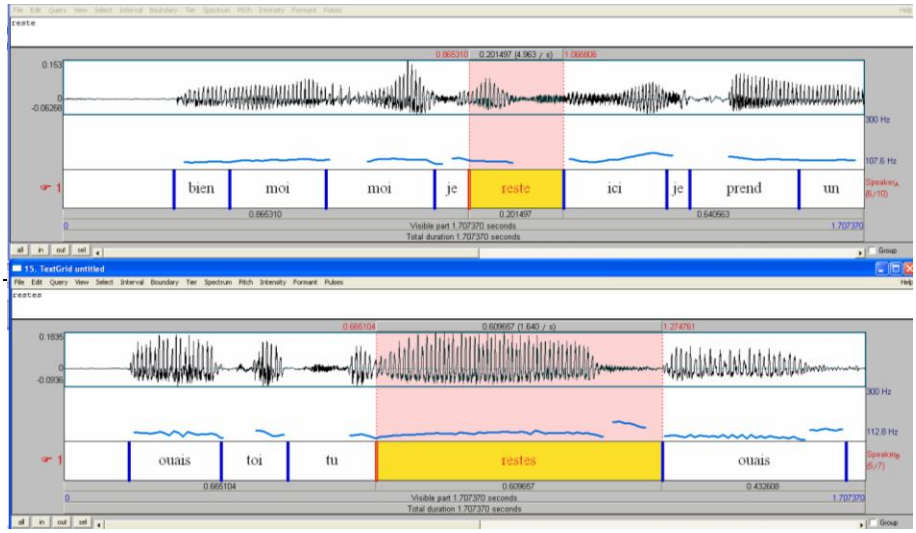

Figure 7. Pitch curves (Hz) of the utterances produced by Speaker A (top) and Speaker B (bottom) in example 3.

5.4 Posture-match: a case of extreme convergence?

The example below illustrates the social role of posture coordination between participants in a way quite similar to the observations made by Chartrand & Bargh (1999) and Shockley *et al.* (2009). It occurs just after the beginning of an interaction between two female participants. They have been urged to speak about unusual things that might have happened to them and at the very beginning of the recording they were thinking about what to say and each of them was turned away from the other, looking up while discussing the meaning of the word "unusual" as illustrated in Figure 8a. At the beginning of example 4, they both turn their heads towards each other with their chin slightly raised without changing the orientation of their body (Figure 8b) and both encourage the other to come up with a narrative.

Example 4

1 A: insolite (0.674) euh si le p- *unusual um yes the p-*

2 B: [bon je vois que tu es tellement à court d'idées allez vas-y tu démarres vas-y vas-y @] *so I see you're really lacking an idea, here you go, you begin, here you go, here you go*

{Both A and B turn their head towards each other in exact synchrony ----------

3 A: [si si vas-y vas-y j'ai j'ai un truc qu'est qui était] extrêmement marrant *yes yes, here you go, here you go, there is something that is that was extremely funny*

Figure 8a. Posture of the 2 participants in line 1 of example 4

Figure 8b. Posture of the 2 participants in line 2 & 3 of example 4

When considering the head orientation of the two participants in example 4, one is compelled by the exact match both before and after they turned their heads towards each other. In determining the presence of repetition, simultaneity plays as important a role in gesture as in the other modalities. One cannot talk of gesture repetition in this example because the change in head direction starts at exactly the same time for each participant. However, there is a strong convergence in both the visual and verbal

modality as not only do both participants turn their heads towards each other at the same time but they also speak in overlap repeating the phrase "vas-y" (*here you go*) several times both in self- and other-repetitions. Prosodically, this overlapping sequence presents a high pitch and intensity for both participants. This characteristic is known to indicate a competitive sequence to gain the floor (French and Local 1986). To describe what exactly happens in this sequence, we can say that there is a real adjustment between speech turns. For each speaker, we observe a similar phrasing in three units. For A: "allez vas y vas y // tu démarres// vas y vas y" and for B: "si si// vas y vas y// j'ai un truc //", the second IP for B being a repetition of A's first IP and A's last IP being the repetition of B's second one. This precise timing inside the overlapping sequence provides evidence that both speakers are in a legitimous position to take the floor according to the rules governing the organization of turn-taking (Sacks et al 1974). After a certain time lag, both participants to the interaction are entitled to take the turn at speech and are therefore potential next speaker. They then start speaking at the same time because the time lag is shared by both participants, a process which is described as case of blind-spot overlap (Jefferson 1987). The effect is to achieve some sort of 'social convergence' insofar as conversation can be seen as a social activity governed by a certain number of rules of politeness. Politeness does not only involve what is said and in what manner but also involves behavior patterns like gaze alternation in between speakers and listeners as well as body orientation towards the co-participant.

Example 4 can be contrasted to later moments in the same interaction where the two speakers are not involved in the interaction to the same degree as illustrated in figure 9. Their body is not oriented towards the co-participant and they do not gaze at each other. At these moments, the previous topic was finished and they had not started a new topic yet. They nevertheless repeat phrases such as "à part ça" (*apart from this*) and "et sinon" (*and otherwise*) which carry little semantic content. The repeated phrases are similar from a lexical and prosodic viewpoint (echo utterances) and they seem to be the only link left between the two speakers, playing a role in the regulation (in terms of cohesion, Tannen 2007) of the interaction. The repetitions show an interactional alignment at the level of forms, but also at a meta-interactional level (both speakers express convergence in their search for a new topic).

Figure 9. Postural misalignment.

4. Conclusion

One of the issues raised in the field of multimodality is precisely how the verbal, the vocal and the visual modalities articulate with one another in the construction of interaction. We know that information is conveyed not only through words and sentence types at the semantic and syntactic levels, but also through prosodic phrasing and contours used by the speaker. It has been shown more recently that co-speech gestures also participate in the conveying of semantic information, and that they play a role in the organization of discourse by speakers. At last, much like what happens in the verbal and vocal modalities, they reveal something of the interpersonal relationship between participants to an interaction. It would, however, be simplistic to suggest that in any utterance, exactly the same information is conveyed in the three modalities at the same time. It cannot be expected therefore that when information is repeated by a participant to an interaction, all of the information will be copied. Rather, the participant is more liable to copy different pieces of the message: part of what was said (semantic information) and/or part of its format (prosodic and gestural information). And since the main role of repetition, as seen in previous studies described in section 3 of this chapter, is to help participants to an interaction achieve some sort of convergence, it is to be expected that depending on the amount of information repeated by a participant, the degree of convergence will be lower or higher.

In order to test this, we analyzed some examples with a focus on gesture repetition. The examples were drawn from the Corpus of Interactional Data (CID) recorded at Aix en Provence. It comprises a series of video recordings of unprepared dialogs in French which were transcribed and annotated in several linguistic domains, including gesture for part of the corpus.

The examples confirmed results from previous studies showing that gesture repetition does not have to be strictly identical to be considered as repetition and that it is rather what makes the semantics of the gesture (namely the type and direction of movement, general hand shape) which has to be copied, whereas other features are not strictly necessary in the repeat (gesture speed or gesture space for instance). These may be

considered as variable features of the gesture. It became apparent as well that although the copy goes towards a reduction of the model in most cases, it sometimes happens that the copy is an improved version of the model, both in terms of length and structure. Gesture repetition may be used to accompany a confirmation request on the part of one of the participants, and therefore as a means to achieving a convergence which is not yet there. In some cases, when the speaker repeats a gesture whereas the prosodic pattern and the verbal message are in contradiction with what was said by the other participant, the gesture repetition may be seen as a means to fake convergence. This reveals how important convergence is to participants in an interaction. It also reveals that, although co-speech gesture is sometimes considered as forming a single idea unit with speech (McNeill 1992), there must be some kind in independence between the different modalities, for them to be repeated or not independently from each other.

We saw as well that timing between model and repetition is of extreme relevance in terms of convergence. When two gestures are produced in complete overlap (and therefore cannot be termed model and repetition) convergence between interactants is at its highest. These particular occurrences of gesture match between participants are also generally accompanied by verbal and prosodic matches.

Beyond the study of repetition, we presented here the more global perspective of the OTIM project which aims to create resources in terms of multimodal corpus and annotations. Thanks to the annotations now available we can investigate numerous phenomena in conversation, that we can compare and that we hope to be able to analyze shortly in a more systematic way, thanks to the adaptation of tools, and automatization in gesture annotation.

Acknowledgement

This research is supported by the French National Research Agency (Project number: ANR BLAN0239). The OTIM project is referenced on the following webpage: http://aune.lpl.univ-aix.fr/~otim/.

BIBLIOGRAPHY

Allwood, J., *et al.*, (2005). The MUMIN Multimodal Coding Scheme, NorFA yearbook 2005. http://www.ling.gu.se/~jens/publications/B%20files/B70.pdf

Allwood, J.; *et al.*, (2000). The Spoken Language Corpus at the Department of Linguistics, Göteborg University. *Forum: Qualitative Social Research* 1(3), 1-20.

Anderson, A., *et al.*, (1991). The HCRC Map Task Corpus. *Language and Speech* 34, 351-366.

Barth-Weingarten, D., (2011). Double Sayings of German JA—More Observations on Their Phonetic Form and Alignment Function. *Research on Language & Social Interaction* 44(2), 157-185.

Bavelas, J. B., Coates, L., & Johnson, T., (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology, 79*, 941-952.

Bertrand, R., *et al.*, (2008). Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues* 49, 105-133.

Bertrand, R.; Priego-Valverde, B., (2011). Does prosody play a specific role in conversational humor? *Pragmatics and Cognition* 19(2), 333-356.

Bigi, B., Bertrand, R., Guardiola, M., (2010). Recherche automatique d'hétéro-répétitions dans un dialogue oral spontané. *In Proceedings of XVIIIèmes Journées d'Étude sur la Parole*, Mons (Belgium), Cederom, 4 pages.

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.

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.

Blache, P. & Prévot, L. (2010). A Formal Scheme for Multimodal Grammars, Proceedings of COLING-2010.

Blanche-Benveniste, C., Jeanjean, C., (1987). *Le français parlé : transcription et édition*, Publication du *Trésor de la langue française*, INALF, Didier Érudition.

Boersma, P., Weenink, D., (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Available: Retrieved May 1, 2009, from http://www.praat.org/

Brun, A., Cerisara, C., Fohr, D., Illina, I., Langlois, D., Mella, O. & Smaïli, K. (2004). Ants : le système de transcription automatique du Loria . *Actes des XXV^e Journées d'Etudes sur la Parole*, Fès, 101-104.

Campbell, N., (2009). Tools and Resources for Visualising Conversational-Speech Interaction. In: Kipp, M.; Martin, J.-C.; Paggio, P.; Heylen, D. (Eds.). *Multimodal Corpora*: *From Models of Natural Interaction to Systems and Applications*. Springer: Heidelberg, 176-188.

Chartrand, T.L., Bargh, J.A., (1999). The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76(6), 893-910.

Clark, H. H., (1996). *Using language*. Cambridge, UK: Cambridge.

Couper-Kuhlen, E., Selting, M., (1996). Towards an interactional perspective on prosody and a prosodic perspective on interaction. In: E.C.-

K.a.M. Selting (Ed.), *Prosody in Conversation*. Cambridge University Press, Cambridge, 11-56.

De Looze, C., *et al.*, (2011). Measuring dynamics of mimicry by means of prosodic cues in conversational speech. In *Proceedings of ICPhS XVII*, Hong Kong, 1294-1297.

Di Cristo, A., Di Cristo P. (2001). Syntaix, une approche métrique-autosegmentale de la prosodie. *Traitement Automatique des Langues*, 42, 1, 69-111.

French, P., Local J. (1986). Prosodic Features and the Management of Interruptions. In C. Johns-Lewis, (ed.), *Intonation in Discourse*. San Diego: College-Hill Press. 157-180.

Garrod, S., Pickering, M.J. (2004). Why is conversation so easy? *TRENDS in Cognitive Sciences* 8(1), 8-11.

Giles, H., Mulac, A., Bradac, J., & Johnson, P. (1987). Speech accomodation theory: The first decade and beyond, Communication Yearbook, 10, ed by M.L. McLaughlin, Sage Pub., Londno, UK, 13-48.

Gorisch, J., Wells, B., Brown, G.J., (2012). Pitch Contour Matching and Interactional Alignment across Turns: An Acoustic Investigation. *Language and Speech* 55, 57-76.

Grønnum, N., (2006). DanPASS - a Danish phonetically annotated spontaneous speech corpus. *In Proceedings of LREC 2006*. Genoa, Italy: 5th LREC conference.

Guardiola, M. (in progress). Contribution multimodale à l'étude de phénomènes de convergence en interaction, PhD Thesis, Aix-Marseille Université.

Heldner, M., Edlund, J., & Hirschberg, J., (2010). Pitch similarity in the vicinity of backchannels. In *Proceedings Interspeech 2010* (pp. 3054-3057), Makuhari, Japan.

Hirst, D., Di Cristo, A., Espesser, R. (2000). Levels of description and levels of representation in the analysis of intonation, in M. Horne (ed) *Prosody* : *Theory and Experiment*, Kluwer : Dordrecht, Pays-Bas, 51-87.

Holler, J., Wilkin, K., (2011). Co-Speech Gesture Mimicry in the Process of Collaborative Referring During Face-to-Face Dialogue. *Journal of Nonverbal Behavior* 35, 133-153.

Hutchby, I., Wooffitt, R., (1998), *Conversation Analysis*, Cambridge, UK: Polity Press.

Jefferson, G. (1987). Notes on "latency" in overlap onset, In G. Button, P. Drew & J. Heritage (Eds.), Interaction and language use. Special issue of *Human Studies*, **9**, 153-183.

Jones, S.S., (2006). Infants learn to imitate by being imitated. In *Proceedings of International Conference on Development and Learning (ICDL)*, Bloomington, IN: Indiana University, 1-6.

Kendon, A., (1980). Gesticulation and Speech : Two Aspects of the Porcess of Utterance. In M.R. Key (ed.), *The Relationship of Verbal and Nonverbal Communication*, The Hague: Mouton, 207-227.

Kimbara, I., (2006). On gestural mimicry. *Gesture* 6(1), 39-61.

Kimbara, I., (2008). Gesture Form Convergence in Joint Description. *Journal of Nonverbal Behavior* 32, 123–131.

Kipp, M., (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech)*, Aalborg, Denmark, 1367-1370.

Kipp, M., (2004). Gesture Generation by Imitation - From Human Behavior to Computer Character Animation. Boca Raton, Florida, Dissertation.com.

Kotthoff H. (2006) "Oral genres of humor: On the dialectic of genre knowledge and creative authoring", *Interaction and Linguistic Structures*, No. 44.

Laforest, M.,(1992). Le back-channel en situation d'entrevue, in *Recherches Sociolinguistiques* 2, Québec : Université Laval.

Lakin, J.L., *et al.*, (2003). The Chameleon Effect as Social Glue: Evidence for the Evolutionary Significance of Nonconscious Mimicry. *Journal of Nonverbal Behavior* 27(3), 145-162.

McNeill, D., (1992). *Hand and Mind. What Gestures Reveal about Thought*, Chicago: The University of Chicago Press.

McNeill, D., (2001). Growth points and catchments. In: C. Cavé, *et al.* (Eds.), *Oralité et Gestualité (ORAGE) : "Interactions et comportements multimodaux dans la communication*". L'Harmattan, Aix-en-Provence, pp. 25-33.

McNeill, D., (2005). *Gesture and Thought*, Chicago, London : The University of Chicago Press.

Mol, L., *et al.*, (In press). Adaptation in gesture: Converging hands or converging minds? . *Journal of Memory and Language*.

Mol, L., *et al.*, (2009). Alignment in Iconic Gestures: Does it make sense? In *Proceedings of AVSP 2009 -- International Conference on Audio-Visual Speech Processing*, University of East Anglia, Norwich, UK, 1-8.

Nesterenko I., Rauzy S. & Bertrand R., (2010). Prosody in a corpus of French spontaneous speech: perception, annotation and prosody ~ syntax interaction. *Proceedings of Speech Prosody 2010,* May 11-14 : Chicago, United States of America.

Norrick, N. (1987). Functions of repetition in conversation, *Text, Interdisciplinary Journal for the Study of Discourse*, 7, 3, 245-264.

Parrill, F., Kimbara, I., (2006). Seeing and Hearing Double: The Influence of Mimicry in Speech and Gesture on Observers. *Journal of Nonverbal Behavior* 30(4), 157-166.

Perrin, L., Deshaies, D., Paradis, C. (2003). Pragmatic functions of local diaphonic repetitions in conversation, Journal of Pragmatics, 35, 1843-1860.

Peshkov, K., Prévot, L., Bertrand, R., Rauzy, S., Blache, P. (2012). Quantitative experiments on prosodic and discourse units in the corpus of Interactional Data, Seinedial, *16th Workshop on the Semantics and Pragmatics of Dialogue*, Paris, Sept. 19-21.

Pickering, J., Garrod, S., (2006). Alignment as the basis for successful communication, *Research on Language and Computation*, 4, 203-228.

Pitt, M.A., *et al.*, (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication* 45, 89-95.

Portes, C., Bertrand, R. & Espesser, R., (2007). Contribution to a grammar of intonation in French. Form and function of three rising patterns. *Cahiers de linguistique française* 28, 155-162.

von Raffler-Engel, W., (1986). The transfer of gestures. *Semiotica* 62(1-2), 129-145.

Sacks, H., *et al.*, (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* 50/4, part 1, 696-735.

Szczepek Reed, B. (2006). Prosodic orientation in English Conversation. Basingstoke, UK: Palgrave MacMillan.

Shockley, K., *et al.*, 2009. Conversation and Coordinative Structures. *Topics in Cognitive Science* 1, 305–319.

Stivers, T., (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction* 41(1), 31-57.

Tabensky, A., (2001). Gesture and speech rephrasings in conversation. *Gesture* 1(2), 213-235.

Tannen, D., 1989, (2007). *Talking Voices: Repetition, dialogue, and imagery in conversational discourse*, Cambridge, CUP.

Wolf, J.C.; Bugmann, G., (2006). Linking Speech and Gesture in Multimodal Instruction Systems. *In Proceedings of IEEE RO-MAN 2006*. Plymouth, UK.