



HAL
open science

Multimodal Analysis of Parentheticals in Conversational Speech

Manon Lelandais, Gaëlle Ferré

► **To cite this version:**

Manon Lelandais, Gaëlle Ferré. Multimodal Analysis of Parentheticals in Conversational Speech. Multimodal Communication, 2014, 3 (2), pp.197-217. 10.1515/mc-2014-0008 . hal-01422573

HAL Id: hal-01422573

<https://hal.science/hal-01422573>

Submitted on 26 Dec 2016

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

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

Multimodal Analysis of Parentheticals in Conversational Speech

Manon Lelandais*, Gaëlle Ferré*

ABSTRACT

Based on a video recording of conversational British English, this paper aims at describing the relation between verbal and non-verbal signals in the production process of parentheticals within the framework of Multimodal Discourse Analysis. Parentheticals are described in linguistics as side sequences interrupting linear development. Although their syntactic, prosodic and discursive characteristics have been deeply analysed, few studies have focused on the articulation of the different communicative modes in their production process. Beyond showing that gesture brings complementary information in regard to prosody, contributing to a composite collateral message, the results allow better delineation and understanding of skip-connecting phenomena as constructing coherence. Changes in the modal configuration throughout the parenthetical sequence suggest modes are dynamic and flexible resources for indexing parentheticals and their particular framing function.

KEYWORDS

Parenthesis, discourse structure, prosody, gesture

INTRODUCTION

In linguistics, parentheticals have often been relegated to editing or processing errors, indexing the unplanned character of speech and contravening certain co-operating norms. From the point of view of Discourse Analysis, parentheticals are short segments inserted into the current stream of discourse, momentarily suspending the progression of a sequence to construct another unit, bringing information at a second discursive level (Payà 2003) such as in example (1) below¹. When analysing how semiotic units form larger sequences of action in discourse and conversation, spontaneous speech presents disruptions where the parts no longer follow one another, some interrupting or abandoning an ongoing utterance, others

*Corresponding authors: Manon Lelandais, Gaëlle Ferré, LLING, Université de Nantes, Chemin de la Censive du Tertre BP 81227, 44312 Nantes cedex 3, France. E-mails : manon.lelandais@etu.univ-nantes.fr; Gaelle.Ferre@univ-nantes.fr

¹ Transcription conventions are provided in the Appendix at the end of the paper.

returning to previous statements. Conversely, some can project an upcoming utterance, such as this reflection announcing an upcoming French term:

- (1) Rhianna # (h) # I don't wanna be a judge # I don't wanna be a prosecutor
Alex yeah
Rhianna so I'd kind of be more interested in uh
well there's no English equivalent
but uh commissaire de police (h)

These breaks in contiguity are extremely relevant to examine language production and comprehension in real time: a particular negotiation of meaning takes place between participants, presenting implications for discourse modelling. Often taken as momentary 'deviations', parentheticals show an interesting degree of flexibility in speech planning and production, in which segments and other objects of discourse are organised into another hierarchy than that of temporal succession. The study of parenthetical sequences also questions the construction of coherence, i.e. the logical relations created by the linkage between constitutive elements of a unit.

Parentheticals display particular interactional strategies, yet their analysis has often concentrated on speech, taking precedence over other communicative modes. In the past, substantial work on such sequences has focused on the relation of parenthetical segments to their host from the perspectives of syntax, relevance theory, or prosody alone, describing them as forms of side sequences which can often be deleted without affecting the syntax or semantics of their host. The development of analytical tools and schemes now facilitates an account of parenthetical sequences as multimodal phenomena.

In face-to-face conversation, participants negotiate meaning through multimodal contributions, in which the linguistic resources of speech interface with gesture. These modes do not work independently from one another, although a particular mode may weigh more than the others at some points (Norris 2004). Within speech itself, the vocal or the verbal mode may stand out as more prominent at particular points in time. The present qualitative analysis aims at establishing links between prosody and gesture in discourse organisation, shedding light on the production process of parentheticals as particular activities structuring information, constructed as resources for conversation.

Viewing language as integrating speech and gesture in an organised system makes it possible to investigate why parentheticals are traditionally seen as disruptive (Quirk et al. 1985: 911; Mazeland 2007: 1856) on the one hand, but somehow related to mainline speech on the other (Mondada & Zay 1999: 401). If parentheticals create a break (Payà 2003: 223) and complicate the flow of discourse, we may wonder how they are recognised as ‘side sequences’ by participants. Regarding them as incremental audio-visual signals oversteps the opposition between competence and performance, which had constrained parentheticals to special formal categories.

From the hypothesis that parentheticals are signalled in an integrative way according to their specific function with a set of features that co-participants attend to, this paper investigates the participants’ use of the array of communicative modes in their construction, with a grounding in Discourse Analysis. Through an empirical annotation scheme, we first enlarge the set of their formal and functional verbal characteristics, allowing the formulation of specific hypotheses for their vocal and visual cues. We then test whether both sets of cues index a break, bring background information, and demarcate the specific types of parentheticals. The results confirm that parentheticals are composite framing devices (i.e. projecting a space in which a semiotic entity makes meaning, providing unity, relation, and coherence) showing a multimodal distribution of labour between discourse organisation and information status. Modal density (Norris 2004) is suggested as a dynamic resource for constructing coherence.

I- THEORETICAL BACKGROUND

1. Parentheticals and discourse

The vast terminology concerning side sequences suggests little consensus over the description of insertion phenomena (Payà, op. cit.), given the various perspectives along which they are observed. The expression derives from ‘parenthesis’, expressing ‘the external and marginal source of the inserted element’ (p.207). Recent studies indeed refer to ‘parenthetical’ sequences/clauses/structures or ‘parentheses’ (Payà 2003; Mazeland 2007; Banik 2008; Blakemore 2009), not indicating any syntactic form or discourse function of the segment in question. ‘Parenthetical’ is used here as a generic term to cover the phenomenon of interpolating an element in discourse, with expressions of varying length and complexity (Dehé 2009) halting an ongoing utterance or turn. These terms sometimes apply to fragments introducing direct speech (*she said*) or to expressions containing verbs of opinion (*I think*) or

declaratives (*I must say*), which are excluded from the definition in this study. The term does not cover comment clauses (*as you know*) either, described in Quirk et al. (op. cit.) and in Brinton (2008). The present paper rather links up to the ‘incidental sequential practice’ referred to in Mazeland (op. cit.: 1817).

Parentheticals are said to organise information in turns and shape the participation framework (Duvallon & Routarinne 2005; Bolden 2009), which reflects the way the speakers and hearers can relate to one another as well as their orientation towards utterances. They are subsidiary activities referred to as skip-connecting, i.e. relating an ongoing utterance to some early talk (Sacks 1995). Likewise, they are considered a useful resource for the handling of topics (Mondada & Zay, op. cit.).

They also allow interactional reconstruction (Fox 1987), as a specific means of framing in discourse with cues directing interpretation (Tannen 1993). Through the production of a parenthetical, the speaker indicates the co-speaker has to re-interpret a part of her utterance. ‘Speaking while monitoring addressees’ (Clark & Krych 2004: 62) shows the speaker’s ability to reflect and modify her utterances in real time for the benefit of the co-speaker, giving background information (Duvallon & Routarinne, op. cit.).

In addition, Clark & Fox Tree (2002:78) regard parentheticals as collateral signals in a discourse model where speakers refer to the ‘official business, or topics, of discourse with primary signals’, and to the ‘performance itself with collateral signals’ (Clark 1996). Such devices manage the on-going performance (Goffman 1981), commenting upon the problems encountered. Collateral signals can either be integrated in a same mode or produced in another form of modality, although only referred to as ‘concomitant’ when produced otherwise than verbally. Recent studies extend this view in suggesting collateral messages can be integrated verbally, prosodically and/or gesturally (Rodrigues 2005).

2. Prosody of parentheticals

Parentheticals are marked off from their hosts by special intonation contours, as well as a by a variation of tempo and of loudness (Crystal 1969; Bolinger 1985). They are uttered in a separate Intonational Phrase (Selkirk 1984). A falling-rising tone on the preceding domain and potential pauses signal their boundaries (Local 1992; Payà, op.cit.).

Bolinger (1985) and Delomier & Morel (1986) agree on a true set of common features for parentheticals which are not representative of the phenomenon if singled out, although

some of them might not be present in a same insertion sequence. Wichmann (2001: 181) calls these common features the ‘canonical’ or typical prosodic pattern for parentheticals.

Intonation has more specifically been confirmed as the main cue of such sequences, through a disjunction of levels (Debaisieux & Martin 2010). The typical ‘parenthetical intonation’ is lower F0 (i.e. *fundamental frequency* = pitch) and no modulation (Hirschberg & Grosz 1992). Local (2007) pinpoints that the pre- and post-insert talk match each other in pitch height and that the post-insert talk is higher. F0 generally rises on the final syllable of the inserted segment, indexing it as prefacing further speech, continuing the paragraph and the point being treated (Morel & Danon-Boileau 1998). Editing out a parenthetical then leaves the remaining utterance sounding prosodically well-formed and coherent, speakers resuming the pitch contour at the point at which it was interrupted (Wichmann 2000).

However, if some canonical prosodic features have been observed, parentheticals often show more variety than described. Their prosody depends on a wide range of criteria such as length, syntactic weight and position (Bolinger, *op. cit.*). Certain short sequences are prosodically integrated into an adjacent segment (Wells 2006; Lee 2007).

3. Gesture and side sequences

A considerable body of research regards gesture as part of a speaker’s message together with language, sharing the same integration system (McNeill 1992; 2005; Bernardis & Gentilucci 2006; Ferré 2014). If a communicative intention fuels the production of a signal, speech is regarded as gesture at a multimodal communication level, as a composite auditory-visual message. Since parentheticals essentially index changes in these communicative intentions, this theoretical stance presents interesting implications.

Additionally, recent research on the construction of coherence in discourse (Laursen 2005; Lascarides & Stone 2009) focused on repeated gestures throughout an interaction by same or other participants in terms of connecting, relating to McNeill’s (*op. cit.* 1992) theory of gesture catchments, considering repetition as correlated to underlying cognitive units in language production.

3.1. Hand gestures

Calbris (2011: 47) studies how the hand gestures related to verbal units render a visual hierarchic organisation of discourse through patterns. If maintaining a gestural shape corresponds to the ‘maintaining of the idea it represents’, concurrent change in gestures

segments discourse. ‘Change creates a gap in the continuum and performs a demarcative function, whereas the type of change provides information and performs a referential function’ (p.46). An earlier study (Calbris 2008: 40) refers to ‘side issues’ as presented by switching hands, the process of utterance production rendered as a ‘path towards a goal to which the central line is the topic’. The initial hand movement is suspended for the duration of the parenthesis, and reused after its completion, the speaker continuing along the same path.

Furthermore, Rodrigues (op.cit.: 7) notes that ‘discontinuity happens in movements with opposing direction’, i.e. to the right and then to the left. Continuity is then established with movements with the same direction, or the progression of circular ones.

Besides, Deppermann (2013) observes pointing gestures as devices for establishing reference as early as possible. Preceding verbal reference, they can orientate recipients to relevant spatial domains and possible referential targets of the turn (Schegloff 1984, Mondada 2004). Lascarides & Stone (op.cit.) approach deictic meaning in gesture as reusing salient information: ‘when [a pointing gesture] is interpreted with respect to a previously established virtual space, they offer similar forms in similar spatial configurations to represent the same objects over time’ (p.19).

In this sense, gestures for the same referent have been observed as more schematic, while gestures for new information are sharper and clearer (Kita, Van Gijn, & Van der Hulst 1998).

3.2. *Head gestures*

Correspondingly, the significant distinction for head gestures can be that of ‘dynamic units’, i.e. of change of direction as opposed to ‘static units’ (Rodrigues, op. cit.: 8).

Among the specific formal types, head nods can provide segmental information, singling out particular entities and participating to their salience, while side-to-side shakes correlate with expressions of intensification and inclusivity. Some other gestures such as beats are connected with discourse structure in their function (Kendon 1972: 7), having a ‘visual rhythmic character that seems to mark out the rhythmic organization of the utterance’.

3.3. *Gaze direction and eyebrow movement*

A shared visual perspective, as part of common ground, is extremely useful in that speakers can choose to convey information visually so that the collaborative effort to establish mutual understanding is reduced to a minimum (Goodwin 1981).

Gaze and head traditionally move away from the co-speaker for discourse elaboration as soon as the speaking turn is taken and secured. The head comes back towards the co-speaker followed by gaze at the end of the turn. A gaze towards the co-speaker functions either as a call for attention or as potential turn-yielding.

Often used for punctuation functions and word search (Ekman & Friesen 1969), eyebrow movement may provide suprasegmental information in participating to focalisation and increasing the detection of prominence.

II- CORPUS AND METHODOLOGY

1. Corpus recording

A spontaneous conversation between two native British young women of 20 and 23 was recorded and filmed for 30 minutes. The few criteria along which the two speakers were selected were the absence of any heavy regional accent and English as a mother tongue. Alexandra was 23 by then and had just settled in France after finishing her degree at UCL in London. Alexandra is from Cambridge. Rhianna was 20 when she was recorded and has mostly lived in France, although she comes from Wimbledon. They had only met once before the recording.

Little information concerning the object of study was given to the speakers before the recording. Alexandra and Rhianna signed an information and consent form, which guaranteed the confidentiality of the data in any other context than academic studies. Disclosing the real purpose of the study would certainly have influenced their conversation as well as their attitude, especially their gestures. The object of study was only referred to as an analysis of spoken English. The consent form also stipulated the duration of the recording.

The speakers were asked to talk just as they do in everyday life, and some specific themes had been suggested, should their conversation be at a standstill (e.g. plans for the forthcoming summer holidays, unusual situations they had lived). Some were used during the recording, others were not.

Sitting in armchairs in a recording studio offering good lighting and soundproofing, they were filmed in a static, wide-angled shot. The two participants were visible from head to ankle, sitting three-quarters turned towards each other: if on the one hand they had been sitting next to each other, they would have instinctively turned their heads towards each other while talking, making it impossible to know whether looking away was deliberate or a means

to rest their heads. If on the other hand they had been completely face-to-face, the camera would not have caught subtle face movements or facial expressions such as raising one's eyebrows. They were left alone during the session and knew they were not overheard by the authors.

Both had a lavalier microphone, providing two separate audio tracks. The native video recording was transformed into an MPEG-4 stereo file, with a rate of 25 frames per second. Two audio files corresponding to each microphone were created in a WAV format, facilitating the analysis of overlapping speech. This material was decided upon as a compromise between the necessity of a quality recording, enabling detailed prosodic and gestural analyses, and that of naturally occurring data.

2. Corpus transcription

2.1. Prosodic coding

The corpus was first edited in Praat (Boersma & Weenink 2011) for a standard orthographic transcription segmented into tone-units, according to the British school of intonation (Halliday 1967; Crystal 1969; Wells 2006) based on dynamic pitch contours.

The Momel-Intsint algorithm (Hirst 2007; Bigi 2012) was used for the automatic annotation of the F0 target points in the signal. Annotations are made in two respects: the algorithm notes pitch height (in Hz) on target syllables, which then allowed us to calculate mean F0 values for specific segments. The algorithm also codes symbolic (relative) values of intonation, in which each measured F0 value is compared to preceding ones, i.e. significant changes in the F0 curve either regarding the speaker's pitch range (Top, Bottom) or regarding the neighbouring tones or sequences of tones (Upstep, Downstep, Same, Low, High). In this work, we are particularly interested in values which indicate a significant pitch reset (Top, Bottom), or a significant change in pitch key (Upstep – change towards higher pitch range, Downstep – towards lower pitch range). We are also interested in the value 'Same' which, if found in greater number in our sequences, would indicate that there is no break in between the different elements of the sequence.

2.2. Gestural coding

The gestures were annotated in Elan (Sloetjes & Wittenburg 2008), based on Kendon's (2004) definition of a gesture as a visible action of any body part that is used as an utterance or part

of an utterance. Hand gestures, head and eyebrow movement as well as gaze direction were manually given labels.

Descriptive annotation

The descriptive annotation was based on gesture units. Each gesture unit considered here starts at the onset of the gesture and ends at the return to rest position if there is one. In the case of two consecutive gestures, the first gesture phrase (adopting Kendon's terminology) ends at a significant change in shape and/or trajectory. A main dimension or function regarding the co-occurring speech corresponds to each gesture unit.

In a series of different tiers, head movement was labelled into nods, shakes, tilts, beats, or jerks. Gaze direction was annotated as either towards the co-participant or away, eyebrow movement distinguished between rise and frown, and hand gestures, which values are described below, were categorised into iconics, metaphorics, pointings, beats, emblems, butterworths, and adaptators.

Functional categories

The hand gesture typology draws mainly from McNeill's work. If iconics are 'images of concrete entities and/or action', metaphorics are 'images of the abstract' involving a metaphoric use of form and/or space (McNeill 2005: 39). Pointing gestures are deictics whereas beats are linked with speech rhythm (McNeill 1992: 80), emblems are conventionalised signs and butterworths are disorganised gestures made in lexical retrieval. Adaptators, i.e. self-contact gestures used for comfort, were included given their important number and the fact that they give the experimenters information on the organisation of turns, being more frequent when the participant is listening. As hand gestures may have several dimensions (McNeill 2005), two values were noted in the same label and were both counted in such a case.

3. Treatment of parentheticals

3.1. Localisation

The parentheticals were localised and transcribed as Ps over the course of the whole interaction. The preceding clause was called A (antecedent), the subsequent one called R (resumption). The object of study was then defined as an APR sequence.

A total of 28 APR sequences were annotated in the thirty-minute conversation. The data were analysed separately for the two speakers because of an asymmetrical distribution of Ps: one speaker produces 22 parenthetical segments, making up a total of 1 min 15 s out of 13 min 31 s of speaking time (i.e. 1.65 P/min) while the other produces only 6, totalling 14 s out of 15 min 10 s of speaking time (i.e. 0.40 P/min)². The similarities and differences between the speakers were nonetheless studied.

3.2. *Verbal annotation*

Syntactic classification

Mainly forming independent clauses, Ps were classified as non-introduced or introduced by conjunctive morphemes. From a macro-syntactic point of view, they are close to parataxis (McCawley 1982; Debaisieux & Martin 2010), i.e. a system of interdependency between two juxtaposed elements of equal status.

Most of the Ps are in a suffix position to the main (or nuclear) construction, respectively identified as [SUF] and [NUCL] in example (2), featuring marked syntactic structures and verbs with low transitivity, prone to identification and subject focus.

- (2) Rhianna (A) no [NUCL] it's a special program # uh for (h) people who speak good English
 (P) [SUF] so hopefully # hem # that wouldn't be a problem (h)

From the point of view of parataxis as a form of structuring beyond other realisation modes of complex constructions, they are propositions juxtaposed in a junction mode between subordination and syntactic independence. This 'third dimension' of phrase structuring, closer to discourse than to sentence, has further semantic and illocutionary implications. If Ps are in majority grammatically autonomous, independent enunciations (introduced with a non-governing conjunctive morpheme such as 'so', or non-introduced), their semantic, modal, and illocutionary autonomy is relative. Their level of autonomy was ranked as shown in Figure 1 below: some Ps are speech acts and syntactically independent, whereas others have a syntactic fixed form and/or are semantically bound to their hosts. Their global autonomy can then be seen on a scale rather than with absolute values. If most of the syntactically independent Ps are speech acts as well, the Ps with no verbal complex are likely to be illocutionarily dependent on their hosts.

² The discrepancy between the two speakers may be explained by the fact that one of them is more familiar with the principal investigator of this study and therefore feels more in charge of the interaction.

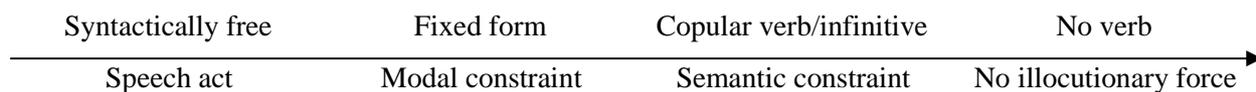


Figure 1. Gradual autonomy of parentheticals

Discursive annotation

Some cohesion (Halliday & Hasan 1976) and reference (Fox & Thompson 2009) marks were surveyed; Ps are prone to lexical repetition and deictics, and their discourse markers work as sequential frames, suggesting Ps are closer to text monitoring than to sentential organisation. Example (3) is a reference-tracking P where the pronoun ‘that’ specifies the identity of a head referent located in A.

- (3) Alex (A) cause it was during freshers’ week
 (P) **oh you don’t have that in France**
 (R) anyway so it was during freshers’ week

Speech acts

Each P was then indexed as either anaphoric or cataphoric, i.e. either commenting or orientating information. Example (4) is a cataphoric APR in a narrative, where the pronoun ‘it’ points forward to a concept that has not yet been introduced.

- (4) Alex (A) and it’s a four-year course cause when you do #
 (P1) **I don’t think it’s not- right it’s not the same thing in France**
 (P2) you know when you do languages
 (R) you have to do well we had to do a third year abroad

Ps were also organised as either essential or non-essential to the sequential development. The APR shown in example (5) below is essential to the narrative, adding some necessary information for a correct reception and interpretation by the co-speaker.

- (5) Alex (A) (h) but hem # and then this other time quite recently i came over (h) and #
 (P) **cause we were meant to London-Tours**
 (R) and the weather was really really bad it was really foggy and overcast

Conversely, example (6) shows a non-essential APR, indexing the speaker’s processing, without which the sequence develops in the same way.

- (6) Alex (A) yeah you should try on like hem le bon coin # there’s hem that

(P) **and what else did I use I used hem #**

(R) (h) # hem # like vivastreet or something like that (h)

Three kinds of break are observed in Ps, in regard to temporal rupture, change in subject, or overstep of sequential expectations (i.e. contradicting the traditional schema for a given sequential type). Example (7), uttered in a narrative, features a temporal switch from past tense to present, the utterance going out of the narrative frame.

(7) Alex (A) cause I was working last summer in the train station

(P) **(h) you know in the kind of like small café things like these ones south and north (h)**

(R) like cause I did that last year

Table 1 below shows that Ps offer hindsight for (re)contextualisation at three different levels. Some Ps act on single words and have a metadiscursive function. Some others have a wider scope in commenting upon a whole stretch of discourse as explanations, concessions or evaluations. A last group of Ps act at the interactional level and are concerned with the state of shared knowledge between participants. The most widespread type is that of discourse Ps. The following APR within a narrative works for instance as an evaluation of the event described in the antecedent.

(8) Alex (A) so i did the second year of # hem # like English # here

(P) **(h) (laughs) which was interesting**

(R) (h) hem yeah and then i just graduated now #

Metadiscourse	Discourse	Representations
lexical/word search	explanation	check common ground
code-switching	concession	add common ground
disambiguation	evaluation	recall common ground

Table 1. Action level of parentheticals depending on their pragmatic function

4 Working hypotheses

Based on the theoretical background defined by the literature and on our observations from the verbal annotation, three main hypotheses are addressed.

If parentheticals signal a break regarding A and R, a difference should be expected in terms of pitch, speech rate, and pauses. A new gesture unit, different in form, should begin,

gaze should be turned towards the co-participant, and eyebrow rising movements should be increased.

If parentheticals bring ‘background’ information regarding A and R, a non-modulated pitch and a faster speech rate are expected. They should feature more abstract metaphoric and pointing gestures.

If the different types of parentheticals are vocally and visually distinguished, metadiscursive Ps should feature more pauses, less modulation, low pitch, fast speech rate, as well as self-centered and suspended, held up gestures.

Discourse Ps should show less pauses and modulation, but more head and eyebrow movements as well as metaphors.

Representational Ps should be characterised by high pitch, modulation, rising contours and slow speech rate. Gaze should be towards the co-participant, and the gestures should be addressee-centered.

III- RESULTS

1. Prosody

1.1. *Speech rate and length*

The results show that P is indeed faster than A and R; P is also the shortest segment whereas R is the longest. Table 2 below shows the mean speech rate and length of each APR segment.

	A	P	R
Speech rate	4.27	6.60	4.44
Length	3.41	3.15	4.50

Table 2. Mean speech rate (in syll/sec.) and length (in seconds) for each part of the parenthetical sequence

Example (9) is a metadiscursive P that is shorter and faster than A and R, as a word search in an argumentation:

(9) Alex	(A) yeah you should try on like hem le bon coin # there’s hem that	3.00 syll/sec
	(P) and what else did I use I used hem #	6.69 syll/sec
	(R) (h) # hem # like vivastreet or something like that (h)	2.75 syll/sec

If metadiscursive Ps are the fastest and shortest, discourse Ps are the slowest and longest as seen in Table 3 below.

	Metadiscursive	Discourse	Representation
Speech rate	7.73	5.81	6.25
Length	1.64	4.39	2.19

Table 3. Mean speech rate (in syll/sec) and length (in sec) for each type of parenthetical

The following discourse P working as an evaluation in an argumentation is indeed longer and slower:

- (10) Alex (A) and then they've got their scratch card it's like 4.55 syll/sec
(P) **(h) we're on a plane for an hour you know wh- # I don't want a scratch card (laughs)**
like (laughs) seriously (laughs) 3.92 syll/sec
(R) so always pointless things 6.29 syll/sec

1.2. *Pauses*

Filled and silent inter-unit pauses occur more between P and R than between A and P, marking Ps at their ends. As far as intra-unit gaps are concerned, R features the greatest number of pauses. Ps are more characterised with 'hem', signalling a local interruption, than with 'uh', signalling a strong boundary. Discourse Ps feature more pauses than the other types.

This metadiscursive P, produced in an answer to Rhianna's enquiry about Alex's master's degree, features a pause between P and R. Silent and filled intra-unit pauses create a prosodic break in R whereas there are no pauses in P:

- (11) Alex (A) (h) mh to use the French word médiation culturelle
(P) **cause I don't know what it is in English**
(R) (h) hem # and then hopefully do something like kind of # work in culture

1.3. *Intonation*

Pitch is lower in P than in A and R, especially in the discourse type, as shown in Table 4 gathering a mean F0 height for each segment.

F0/ APR segment	A	P	R
F0	192.6	182	185.7
F0/ P type	Metadiscursive	Discourse	Representation
F0	189.3	177.1	188.1

Table 4. Mean pitch height (in Hz) in function of APR segment and P type

This metadiscursive P is produced by Alex in a descriptive sequence about the student loans she was allocated in the UK:

VIDEO 1: YEARLY INSTALMENTS

- (12) Alex (A) and then you get # hem like three yearly ins- like # no #
- (P) **that doesn't make sense three yearly instalments**
- (R) (h) yeah # three instalments per year of like living money

Figure 2 below shows the observed pitch height difference. Pitch is lower in P than in A and R (mean pitch in A: 161.1 Hz; mean pitch in P: 84.2 Hz; mean pitch in R: 176.4 Hz). The final contours in A and P can however be rising and indicate further speech to come (Figure 2, Figure 3).

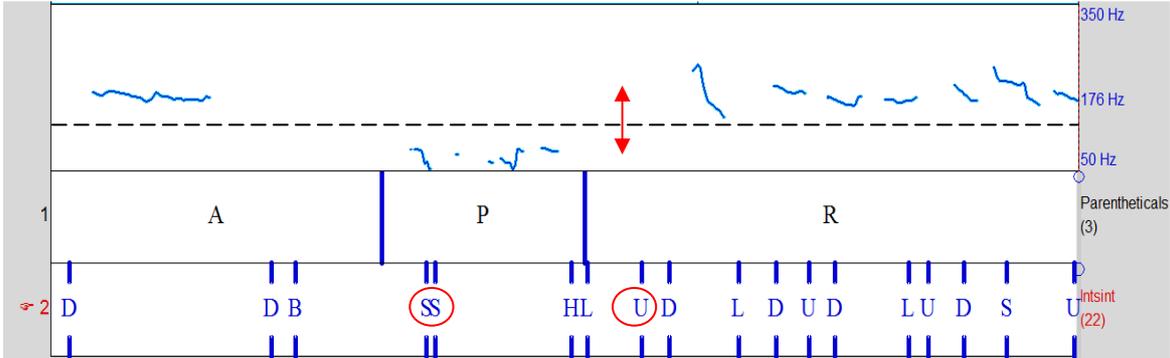


Figure 2. Lower pitch in a metadiscursive P in Praat (first transcription track shows segments – A, P, or R – in the APR sequence, second transcription track shows Intsint’s coded values)

For both speakers, R shows a greater prosodic break than A and P in featuring more demarcative tones values (hence pitch movement) than the other two segments.

Tones/APR	A	P	R
Bottom	8	15	11
Top	0	2	1
Downstep	49	45	76
Upstep	42	41	53
Total	99	103	141
Same	37	33	69

Table 5. Number of occurrences of symbolic relative values of intonation for each segment of the parenthetical sequences extracted from the INTSINT-MOMEL algorithm

Table 5 above also shows flat contours detected by the value ‘Same’ (S): resumptions are prototypically downstepped plateaus for both speakers.

In the context of a dialogic sequence about University, Alex produces a representation P, checking whether UCL is part of the common ground.

VIDEO 2: UNIVERSITY COLLEGE LONDON

- (13) Alex (A) (h) hem I just finished my deg # I did a four-year degree in London
 (P) uh University College London I don't know if you've heard of it UCL #
 (R) and hem
-

Figure 3 below shows Alex's resumption as a downstepped plateau, pitch decreasing from 202 Hz at the end of P to 190 Hz during R. R shows a flat pitch curve as the S-value signals same pitch height after a downstep (D).

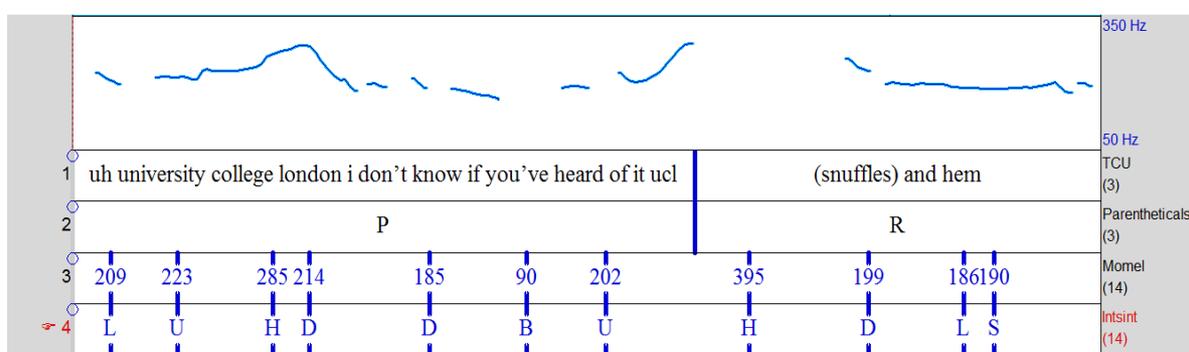


Figure 3. Resumption as a downstepped plateau in a discourse APR in Praat (first transcription track shows segments – P, or R – in the APR sequence, second transcription track gives Momel's corrected F0 values in Hz, and third transcription track shows Intsint's coded values)

Not only do discourse parentheticals show a greater prosodic break in gathering more tone values, but they also gather more S points than the other types, i.e. they are flatter, as shown in Table 6 below. The parentheticals acting on discourse are downstepped and upstepped plateaus in majority. If discourse Ps are flat, representation Ps show modulation (few S points).

Tones/P type	Metadiscourse	Discourse	Representation
Bottom	4	10	1
Top	1	1	0
Downstep	12	30	3
Upstep	8	28	5
Total	25	69	9
S	9	21	3

Table 6. Number of occurrences of symbolic relative values of intonation for each type of parenthetical extracted from the INTSINT-MOMEL algorithm

1.4. Restarts and resumptions

Frequent interruptions mark the end of A, either creating orphans, i.e. suspension of speech between words, or word fragments, i.e. suspension of speech within words. The majority of orphans over word fragments suggests a preferential preliminary syntactic commitment, in that syntax weighs more than phonology in the process planning of speech during APRs.

Orphans and near repeats can be seen in this discourse P working as a justification in a narrative about luggage restriction:

- (14) Alex (A) because # they didn't #
 (P) **cause you know they have the like cabin baggage restriction that says that all the stuff fits in the overhead # lockers**
 (R) (h) and they obviously didn't have enough space

2. Gesture

2.1. *Hand gestures*

Crucially, speakers gesture more during P and R. If no hand gesture is produced in A, metaphoric gestures preferentially appear in P or R. As for adaptators, the same gesture unit is produced throughout the whole sequence or appears in R, indexing processing or retrieval.

The following P, working on discourse as an explanation in the context of a narrative about changing subjects at University, features emerging gestures co-occurring with different segments. In (b), she produces the emblematic gesture for TWO, in (c), she presents the emblematic gesture for MAYBE with an oscillating movement of her right hand (her wrist shifting from the vertical plane to the horizontal plane), and at last produces an adaptator in (c), touching her lips during the silent pause that follows the filler 'hem':

VIDEO 3: INTERPRETER

- (15) Alex and (h) I was like oh god why am I # doing German cause I actually don't really like German very much as like I'm not very good at it
 Rhianna (giggles)
 Alex (A) [(a) (h) but hem #]
 (P) [(b) **cause I wanted two languages**] [(c) **to maybe be an interpreter or translator**]
 Rhianna mhmh
 Alex (R) (h) but hem [(d) # anyway] I decided that # no it wasn't for me cause I really can't stand German
-

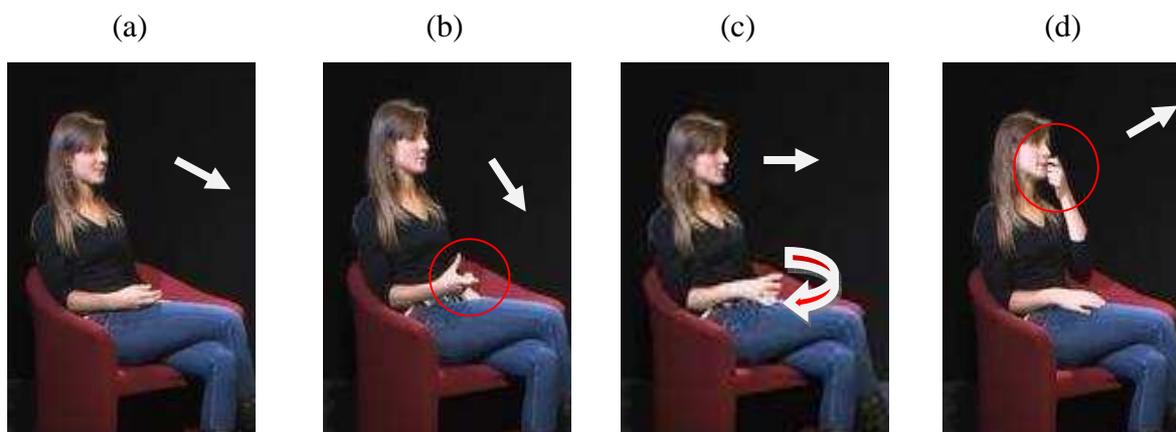


Figure 4. Gestures emerging in an explanatory APR sequence

Figure 4 above shows different moments from Alex's APR: (a) corresponds to a moment in A, (b) and (c) show moments in P, (d) is in R. Each gestural activity illustrated in a picture is represented in between brackets.

If the higher proportion of metaphors over iconics only marks a general trend, we also note that gestural tendencies are increased in P rather than contradicted: if a specific type of hand gesture is produced in A, more of this specific type is produced in P, such as in the following APR, increasing the common ground in Rhianna's description of her childhood in Montmartre. Figure 5 below shows a succession of abstract pointing gestures sharing similar forms and values regarding the co-occurring speech. What is interesting though is that each abstract pointing does not correspond to a precise referent (the town hall as opposed to the next district), but rather serves as a demarcation in between speech acts. Yet, the same hand shape used by the speaker maintains the gestural coherence of the passage:

VIDEO 4: TOWN HALL

- (16) Rhianna (A) I [(a) was closer to the hem] # to the town hall #
 (P) cause [(b) now they have a town hall] [(c) and it's just right] the next district
 (R) [(d) (h) and hem # i was closer to there #]
-

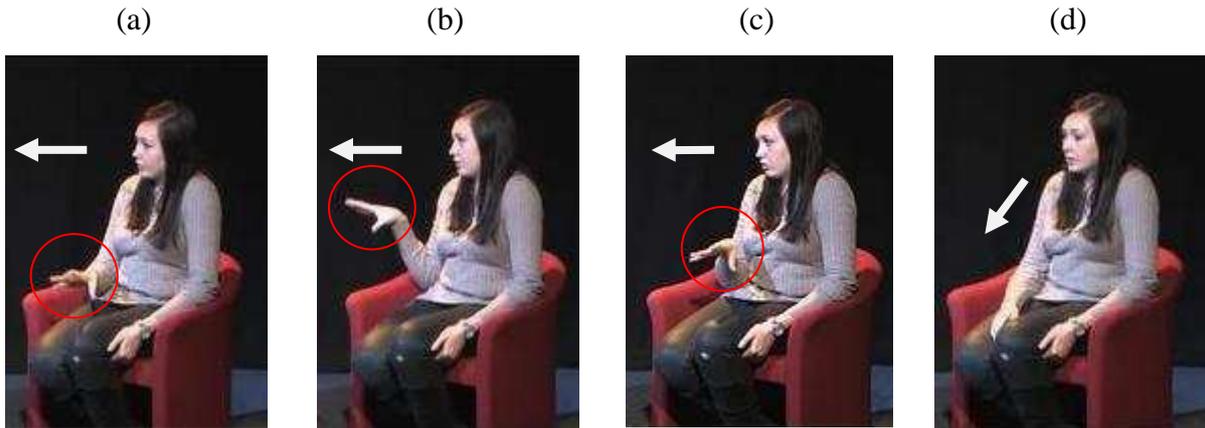


Figure 5. Series of abstract pointing gestures sharing the same configuration and value throughout an APR

Ps are also loci for emerging small pointing gestures, all correlated with specific lexical affiliates. This metadiscursive P in a narrative about giving English lessons features for instance two pointing gestures correlated with ‘advertisements’, the key lexical element of the sequence: the appropriation of the term is commented upon as an insecure repair to the disfluencies in A. Alex freezes an adaptator and produces a first pointing gesture with her right hand while proposing a potential repair, still showing hesitation (b). She then ends the adaptator to reinforce the pointing gesture she is holding with a similar sign of her left hand (c), appealing to the co-speaker. She returns to a rest position as soon as the lexical retrieval is achieved (d).

VIDEO 5: ADVERTISEMENTS

(17) Alex (A) [(a) (h) so I put # an an- like hem #
 (P) uh not announces hem #] [(b) like advertise- #] [(c) ments is it] *is it what it is* [(d)
 yeah
 (R) (h) on hem (swallows) online about a week ago]

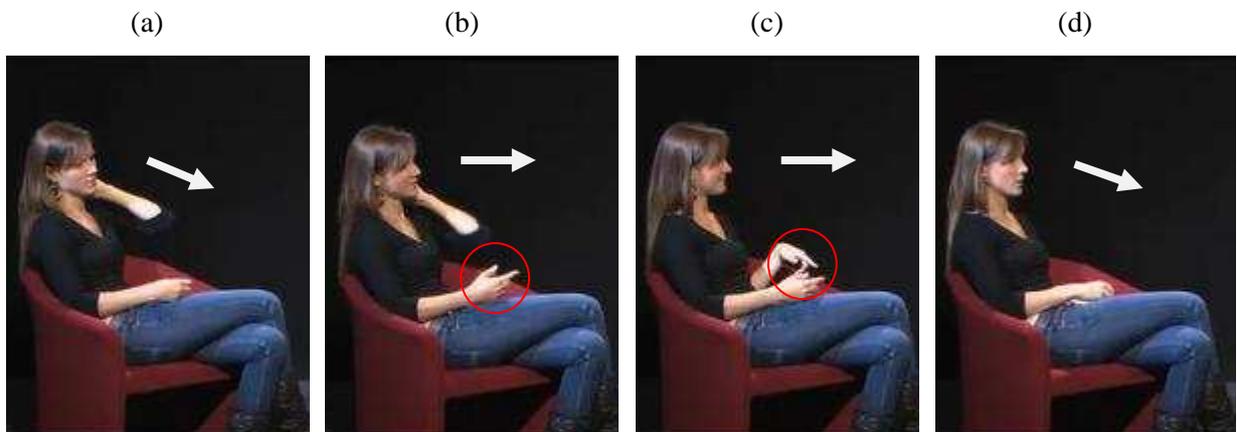


Figure 6. Pointing gestures during a metadiscursive APR sequence

2.2. Head gestures

Beats and tilts occur in greater number during P and R. Just as they are more populated with hand gestures, discourse Ps also gather more head movement, such as beats, connected to parsing and rhythmic structure. Evaluations, concessions and explanations also link with stance-introductory tilts and jerks, such as the tilt in Figure 7 below during an evaluative P in a dialogic sequence:

VIDEO 6: ERASMUS PROGRAM

(18) Alex oh what an Erasmus year
Rhianna (A) no it's a special program # uh for (h) [(a) people who speak good English]
 (P) so [(b) hopefully #] [(c) hem # that wouldn't be a problem (h)]
 (R) and # then I have another year but I don't know if I'm doing it in Nantes or in Paris so

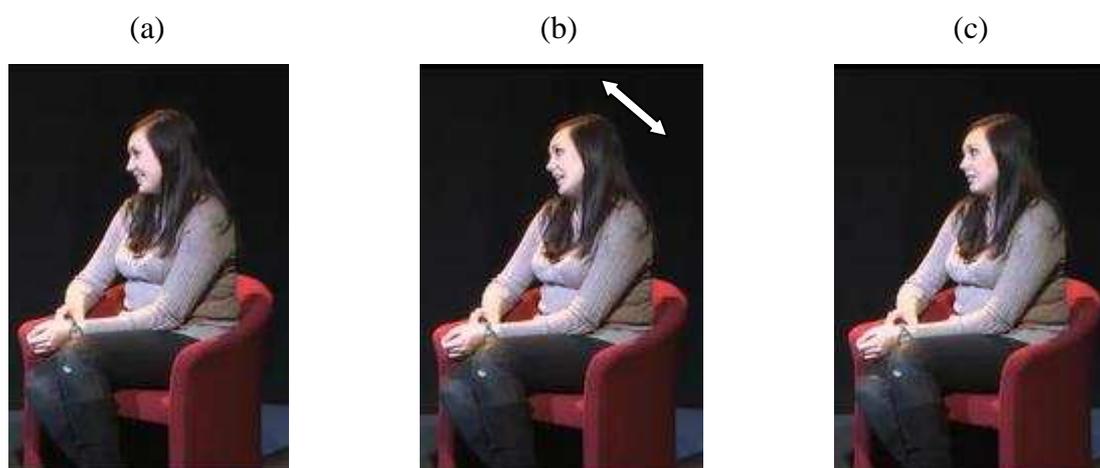


Figure 7. Head tilt produced in an evaluative parenthetical

2.3. Gaze direction

Gaze direction reveals two opposite tendencies: if one speaker looks away throughout the APR, using gaze in an ‘unmarked’ way to hold her turn, the other looks up towards the co-participant in P, which is in contrast with the other parts of the APR.

The two examples below, along with Figure 8, show how gaze is used differently as a resource. Alex’s APR acting on representation, as an addition to the common ground in a

narrative about a flight being rescheduled (19), contrasts with Rhianna's metadiscursive APR in a dialogic sequence about her degree (20).

VIDEO 7: LONDON-TOURS

(19) Alex (h) but hem # and then this other time quite recently I came over (h)
 (A) [(a) and #]
 (P) [(b) cause we were meant to London-Tours
 (R) and] [(c) the weather was really really bad it was really foggy and overcast]

VIDEO 8: PROSECUTOR

(20) Rhianna # (h) # I don't wanna be a judge # I don't wanna be a prosecutor
 (A) [(d) so I'd kind of be more interested in uh]
 (P) [(e) well there's no English equivalent
 (R) but uh] [(f) commissaire de police (h)]

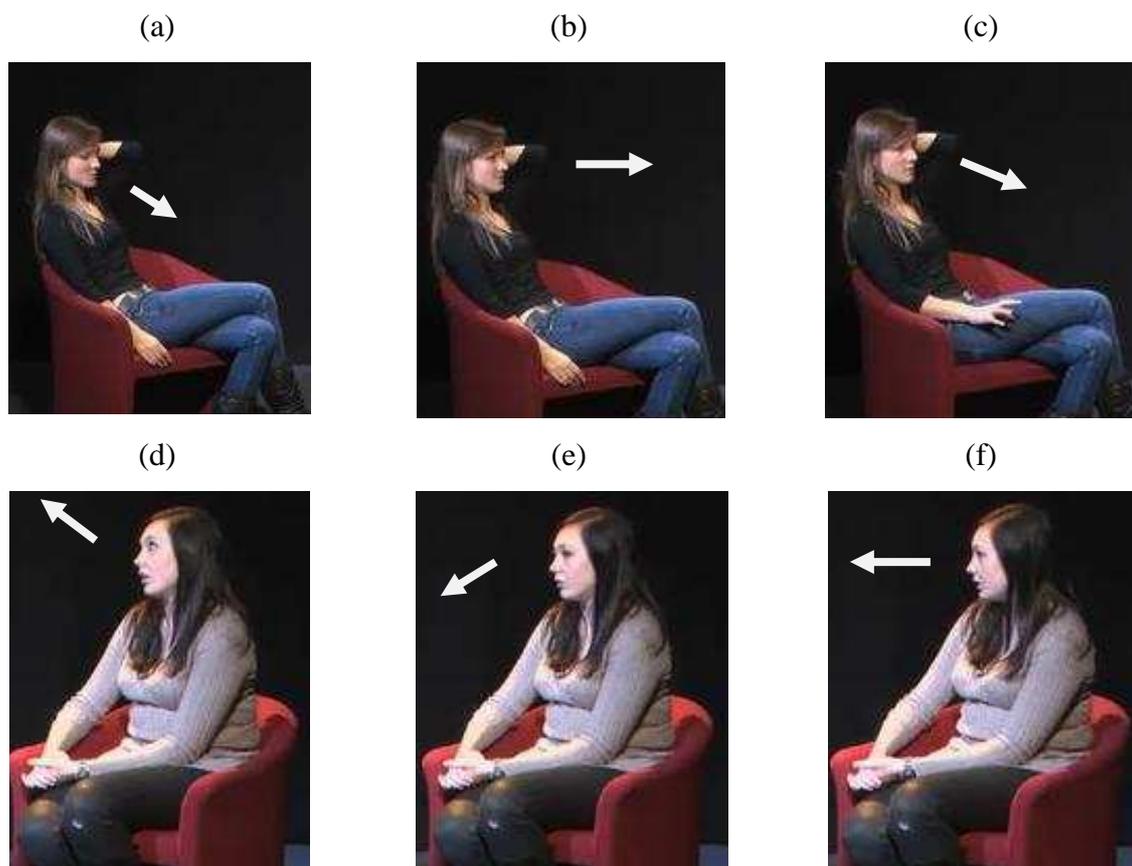


Figure 8. Different use of gaze direction during two APRs

2.4. Eyebrow movement

Conversely, eyebrow movement does not frame the different APR segments but relates to their function: metadiscursive APRs are the preferred loci for frowning (stance-introductory) while discourse APRs are prone to eyebrow rise (focalisation on a lexical item).

A concessive P in an argumentation (21) contrasts with a metadiscursive P in a narrative (22) in the type of eyebrow movement shown below. If raising one's eyebrows highlights both an element of contest in P and a means of transport standing in opposition in R to the general topic (i.e. flying), frowning indexes processing difficulties in A and uncertainty in P towards a potential ambiguous element (freshers' week):

VIDEO 9: TRIP TO ENGLAND

(21) Rhianna (A) yeah well I mean I'm going to England on the twenty eighth
(P) **although [RISE I'm not staying] very long**
(R) hem but we're [RISE going by # car] and then taking the boat

VIDEO 10: FRESHERS' WEEK

(22) Alex (A) it was [FROWN cause it was] during freshers' week
(P) **oh [FROWN you don't have that in France there's no]**
Rhianna no we don't really have
Alex (R) anyway so it was during freshers' week and I went round

IV- DISCUSSION

1. Gesture vs. Prosody?

The prosodic cues fully confirm our first hypothesis according to which APRs create a break, in that they feature a difference in terms of pitch, speech rate and pauses. However, this break is not indexed where it was expected: it lies in the resumption part rather than in P itself, meaning that a break is not signalled between the antecedent and the parenthetical, but rather between the parenthetical and the resumption. The gestural features, conversely, fully confirm our second premise about APRs bringing background information in that the proportion of metaphoric and abstract pointing gestures are increased in P and R, subscribing to the previous typological tendencies of a sequence in producing more gestures of a same type. Prosody and gesture, then, do not bring the same message regarding parentheticals.

If the different types can be distinguished, only discourse Ps feature both distinct prosodic and visual characteristics: they are slower and longer than the other P types,

featuring more pauses. They are also accompanied with more head movement, eyebrow rises, and abstract gestures. Metadiscursive Ps are signalled gesturally with eyebrow movement (frowns) and adaptators, whereas representation Ps are modulated, not signalled with gesture at all. Some Ps show self-attentiveness (lower F0, faster speech rate, gaze away, adaptators) whereas others show other-attentiveness (change of F0, shorter, implicational tones, gaze moved or maintained towards the co-speaker). Other-attentiveness is however dominant in parentheticals, whatever their pragmatic type.

Gesture brings extra information in regard to prosody, contributing to a complex and composite message in which modes simultaneously bring information at local and global levels: APR sequences feature prosodic action at a linear level with segmental cues indexing an unexpected disruption (demarcative function), while gesture is used at a sequential level, defining the status of the information presented in the APR as background (referential function). This distribution of labour between modes can also be observed through several temporalities: gesture, relying on space as well as on time, ‘ties’ A, P, and R together, escaping the linearity of sound. Prosody is preferred for its local segmental functions, framing P or R alone. This does not mean, however, that both segmental and suprasegmental cues always index discontinuity: key and final contours can mark further incoming speech, organising information (Coulthard & Brazil 1979; Hirschberg & Pierrehumbert 1986), but there are more prosodic cues expressing demarcation than prosodic signs of continuum in APRs.

‘Complicating locally to simplify globally’ allows Ps to take local configuration and form, without becoming the central activity of the interaction; information is marked as different just as the sequence is indexed as not finished, the conversational agenda still pending and attended to. The speaker is identified as ‘doing’ the clarification: the opening of a specific sequence by the co-speaker is not wanted. Turn-holding, if not already clearly indicated, is secured.

2. Change in modal configuration

The highest modal density, i.e. the moment where all modes are the most active, drawn on and interwoven at the same time (Norris 2004; 2011) lies in R, where most of the ‘tying up’ to the main sequence is achieved and where P is also indexed as ‘particular’ speech retrospectively. The negotiation of the communicative importance between the different modes fluctuates not only as different action sequences take place, but also within a particular sequence.

Although one mode is generally dominant in a discourse segment, both gesture and prosody play a part in language integration: neither of these two modes can be taken into account independently from one another. Recent work goes towards a shared integration system between action observation and language processing: language and action-related information (i.e. gesture) contribute to a single yet composite, incremental speaker’s message.

The multimodal configuration can be seen in Figure 9 below, in which the squares stand for each particular action performed through discourse. As APRs are conversational phenomena, the verbal discourse is each time at the core of this action. P stands for prosody while G stands for gesture, weighing differently in the distinct segments. When the visual, vocal and verbal modes happen to be equally drawn on, each one brings information at different levels, global and local. Depending on from which ‘window’ they are looked upon, Ps can be indexed (prospectively, simultaneously, retrospectively) as contributing unusually to the embedding sequence or showing a different degree of modality, while remaining cohesive in terms of concepts through reference.

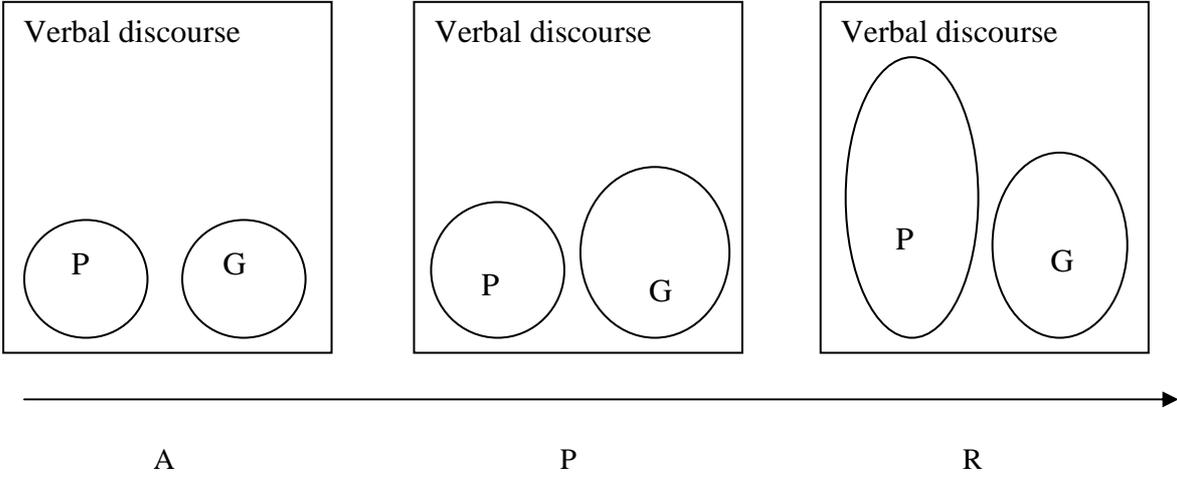


Figure 9. Change in modal configuration during APRs

3. Global coherence

This dialectic between speech and gesture can also be seen as semantic interaction. In terms of coherence relations, which ‘structure communicative actions by showing how the speaker is grouping ideas together to highlight the meaningful relationships among them’ (Lascarides & Stone 2009: 3), information is not presented in isolation but in a contiguity relationship. Laursen (2005) observes that when gesture projects an increment, indicating which element is followed by an increment and the element to which it is tied, the verbal mode specifies the

type of relation that is to be expected in the increment (i.e. that makes the incremented action relevant and coherent), while gesture makes the action visibly cohesive. Likewise, the function of some gestures constructing coherence and cohesion throughout the APR in establishing reference (e.g. small pointing gestures, repetition of metaphors, continuation of a frozen gesture) can be compared to that of repeats and pronouns in the verbal mode. APRs are action sequences in which the referent is likely but not certain to be recognisable for interlocutors, in the way Enfield, Kita & de Ruiter (2007: 1733) describe ‘insecure reference environments’, also prone to try-markers.

Verbal/visual correspondences favour a global model of coherence across modalities. Interactional reconstruction (Fox 1987) is achieved through an interplay of verbal, visual, and vocal cues. Devices such as conjunction signal the type of context space being developed and their relation to preceding context spaces. This can be seen as extending Reichman’s (1978) model: an ongoing conversation is parsed on-line, based on verbal but also co-verbal and nonverbal cues, the results of this parsing are passed to a database, which acts as a model of the ongoing interaction.

It is in this sense that the multimodal cues in A and in P are substrates in Goodwin’s terms (2012), forming a basis for transformative operations preserving the structure. ‘Action uses as its input the structure and resources of a current substrate and produces as its output a new, transformed substrate’ as a point of departure for a next action (p. 17). Re-using material throughout an APR triggers reanalysis, as a specific recontextualisation changing interpretation.

V- CONCLUSIONS

One of the issues raised by multimodality is how the verbal, the vocal, and the visual modalities articulate with one another in the construction of meaning and interaction. In the production process of parentheticals, speakers increase communicative efficiency by simultaneously conveying different information in various modalities, at several levels.

The empirical annotation shows that the identification and description of Ps in the verbal mode are based on a whole cluster of features, not only discontinuity. The hindsight brought by Ps on mainline speech can be layered in three different scopes.

Ps open a focus of attention on some object of discourse, which has to be reanalysed by the co-speaker from a new perspective because speakers think their previous or following

speech might be ambiguous. Speakers mark the unexpected shift of frames with specific prosodic patterns suggesting a break in linearity (demarcative function), while gestural patterns ‘tie’ the parenthetical segment to mainline discourse in establishing coherence for the whole sequence (referential function). It is in the resumption part that local, editing action on discourse is indexed: parenthetical segments are signaled as such retrospectively.

Both verbal and nonverbal signals are related to each other, and contribute differently to the elaboration of the message. The vocal, verbal and visual modes are dynamic, adjustable resources for indexing Ps and their particular action.

We have mainly analysed APRs from the point of view of production. However, much finer, detailed work remains to be done on the multimodal cues for the reception and the perception of APRs, for instance through an extensive study of backchannels as a multimodal addressee response embedded in the inserted action, that could influence the speaker’s communicative choices.

Furthermore, multimodal data are known to provide rich insight for discourse phenomena and language production mechanisms. However, a thirty-minute dyadic conversation provides a qualitative analysis for APRs, in which hypotheses could be generalised only through their testing on larger multimodal corpora.

ACKNOWLEDGEMENTS

We would like to thank our two participants Alexandra and Rhianna, as well as the staff at the Audio-Visual Centre at the University of Nantes. We are also indebted to two anonymous reviewers for their valuable comments and suggestions on a previous version of this paper.

APPENDIX: TRANSCRIPTION CONVENTIONS

- (h) audible inbreath
- # pause
- XXX inaudible word (X stands for 1 syllable)
- (...) vocal activity (laughs, giggles, swallowing, sighs)
- *...* low intensity
- [...] illustrated gestural activity
- interrupted construction
- (A) antecedent

- (P) parenthetical
(R) resumption

REFERENCES

- Banik, E. 2008. "An Integrated Architecture for Generating Parenthetical Constructions." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Student Research Workshop*, Columbus, Ohio: 7–12.
- Bernardis, P., and M. Gentilucci. 2006. "Speech and Gesture Share the Same Communication System." *Neuropsychologia* 44 (2): 178–90.
- Bigi, Brigitte. 2012. "SPPAS: A Tool for the Phonetic Segmentation of Speech." In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, Turkey: 1748–55.
- Blakemore, D. 2009. "On the Relevance of Parentheticals." In *IDP 09*, Paris: 9–17.
- Boersma, P., and Weenink, D. 2013. "Praat: Doing Phonetics by Computer." Accessed January 30. <http://www.fon.hum.uva.nl/praat/>.
- Bolden, G. 2009. "Implementing Delayed Actions." In J. Sidnell (ed.) *Conversation Analysis: Comparative Perspectives*, Cambridge University Press, Cambridge: 326–53.
- Bolinger, D. 1985. *Intonation and Its Part: Melody in Spoken English*. Stanford University Press. Stanford, California.
- Brinton, L.J. 2008. *The Comment Clause in English: Syntactic Origins and Pragmatic Development*. Cambridge University Press. Cambridge.
- Calbris, G. 2008. "From Left to Right...": Coverbal Gestures and Their Symbolic Use of Space." In A. Cienki and C. Müller (Eds.), *Metaphor and Gesture*. Gesture Studies Vol. 3. John Benjamins. Amsterdam; Philadelphia: 27–53.
- . 2011. *Elements of Meaning in Gesture*. Gesture Studies Vol. 5. John Benjamins. Amsterdam; Philadelphia.
- Clark, H. H. 1996. *Using Language*. Cambridge University Press. Cambridge.
- Clark, H. H., and J. E. Fox Tree. 2002. "Using 'Uh' and 'Um' in Spontaneous Speaking." *Cognition* 84 (1): 73–111.
- Clark, H. H., and M. A. Krych. 2004. "Speaking While Monitoring Addressees for Understanding." *Journal of Memory and Language* 50 (1): 62–81.

- Coulthard, M., and Brazil, D. 1979. "Exchange Structure. Discourse Analysis Monograph No. 5." *English Language Research*, University of Birmingham edition. Birmingham.
- Crystal, D. 1969. *Prosodic Systems and Intonation in English*. Cambridge University Press. Cambridge.
- Debaisieux, J-M., and P. Martin. 2010. "Les Parenthèses: Étude Macrosyntaxique et Prosodique Sur Corpus." In Béguelin, M. J., Avanzi, M., & Corminboeuf, G. (eds): *La Parataxe: Tome 2*, Peter Lang, Bern: 307–37.
- Dehé, N. 2009. "Clausal Parentheticals, Intonational Phrasing, and Prosodic Theory." *Journal of Linguistics* 45 (3): 569–615.
- Delomier, D., and M-A. Morel. 1986. "Caractéristiques Intonatives et Syntaxiques Des Incises." *DRLAV* 34-35: 141–60.
- Deppermann, A. 2013. "Turn-Design at Turn-Beginnings: Multimodal Resources to Deal with Tasks of Turn-Construction in German." *Journal of Pragmatics* 46 (1): 91–121.
- Duvallon, O., and S. Routarinne. 2005. "Parenthesis as a Resource in the Grammar of Conversation." In Hakulinen, A. & Selting, M. (eds): *Syntax and Lexis in Conversation: Studies on the Use of Linguistic Resources in Talk-in-Interaction*. John Benjamins. Amsterdam; Philadelphia: 45–74.
- Ekman, P., and W. V. Friesen. 1969. "The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding." *Semiotica* 1 (1): 49–98.
- Enfield, N.J., S. Kita, and J.P. de Ruiter. 2007. "Primary and Secondary Pragmatic Functions of Pointing Gestures." *Journal of Pragmatics* 39 (10): 1722–41.
- Ferré, G. 2014. "A Multimodal Approach to Markedness in Spoken French." *Speech Communication: Special Issue on Gesture and Speech in Interaction* 57: 268–82.
- Fox, B. A. 1987. "Interactional Reconstruction in Real-Time Language Processing." *Cognitive Science* 11 (3): 365–87.
- Fox, B. A., and S. A. Thompson. 2009. "On Formulating Reference: An Interactional Approach to Relative Clauses in English Conversation." *Papers in Pragmatics* 4 (1): 183–96.
- Goffman, E. 1981. *Forms of Talk*. University of Pennsylvania Press. Philadelphia.
- Goodwin, C. 1981. *Conversational Organization: Interaction between Speakers and Hearers*. Academic Press. New York.
- . 2012. "The Co-Operative, Transformative Organization of Human Action and Knowledge." *Journal of Pragmatics* 46 (1): 8–23.
- Halliday, M. A. K. 1967. *Intonation and Grammar in British English*. Mouton. The Hague.

- Halliday, M.A.K., and R. Hasan. 1976. *Cohesion in English*. Longman. London.
- Hirschberg, J., and B. Grosz. 1992. "Intonational Features of Local and Global Discourse Structure." In *Proceedings of the Workshop on Speech and Natural Language*, Harriman, NY., 441–46.
- Hirschberg, J., and J. Pierrehumbert. 1986. "The Intonational Structuring of Discourse." In *Proceedings of the 24th Annual Meeting on Association for Computational Linguistics*, Bonn, Germany, 136–44.
- Hirst, D. J. 2007. "A Praat Plugin for Momel and INTSINT with Improved Algorithms for Modelling and Coding Intonation." In *Proceedings of the XVIth International Conference of Phonetic Sciences*, Saarbrücken, Germany, 1233–36.
- Kendon, A. 1972. "Some Relationships between Body Motion and Speech." In A. W. Siegman and B. Pope (eds): *Studies in Dyadic Communication*, Pergamon, New York: 177–210.
- . 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.
- Kita, S., I. Van Gijn, and H. Van der Hulst. 1998. "Movement Phases in Signs and Co-Speech Gestures, and Their Transcription by Human Coders." In *Gesture and Sign Language in Human-Computer Interaction* (1371): 23–35.
- Lascarides, A., and M. Stone. 2009. "Discourse Coherence and Gesture Interpretation." *Gesture* 9 (2): 147–80.
- Laursen, L. 2005. "Towards an Embodied Grammar: Gesture in Tying Practices Constructing Obvious Cohesion." In *Interacting Bodies: Online Proceedings of the 2nd ISGS Conference*. Lyon.
- Lee, Y. 2007. "On the Role of Intonation and Pauses in the Interpretation of English Parenthetical Adverbs." *The Linguistic Association of Korea Journal* 15: 85–107.
- Local, J. 1992. "Continuing and Restarting." In Auer, P. & Di Luzio, A. (eds): *The Contextualization of Language*, John Benjamins, Amsterdam and Philadelphia: 273–96.
- Local, J. 2007. "Phonetic Detail and the Organisation of Talk-in-Interaction." *Proceedings of the 16th ICPHS, Saarbrücken, Germany*.
- Mazeland, H. 2007. "Parenthetical Sequences." *Journal of Pragmatics* 39 (10): 1816–69.
- McCawley, James D. 1982. "Parentheticals and Discontinuous Constituent Structure." *Linguistic Inquiry* 13 (1): 91–106.
- McNeill, D. 1992. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press. Chicago.

- . 2005. *Gesture and Thought*. University of Chicago Press. Chicago.
- Mondada, L. 2004. “Temporalité, Séquentialité et Multimodalité Au Fondement de L’organisation de L’interaction: Le Pointage Comme Pratique de Prise de Tour.” *Cahiers de Linguistique Française* 26: 169–92.
- Mondada, L., and F. Zay. 1999. “Parenthèses et Processus de Configuration Thématique: Vers Une Redéfinition de La Notion de Topic.” In *Pragmatics in 1998: Selected Papers from the 6th International Pragmatics Conference, 2*. Antwerp: 396–411.
- Morel, M-A., and L. Danon-Boileau. 1998. *Grammaire de l’intonation*. Paris: Editions Ophrys.
- Norris, S. 2004. *Analyzing Multimodal Interaction: A Methodological Framework*. Routledge. New York, London.
- . 2011. “Three Hierarchical Positions of Deictic Gesture in Relation to Spoken Language: A Multimodal Interaction Analysis.” *Visual Communication* 10 (2): 129–47.
- Payà, M. 2003. “Prosody and Pragmatics in Parenthetical Insertions in Catalan.” *Catalan Journal of Linguistics* (2): 207–27.
- Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. Longman. London.
- Reichman, R. 1978. “Conversational Coherency.” *Cognitive Science* 2 (4): 283-327.
- Rodrigues, I. G. 2005. “Verbal and Nonverbal Signals in Face-to-Face Interaction: A Theoretical Framework for a Holistic Micro-Analysis: The Example of a Parenthesis.” *Interacting Bodies*, Lyon: 1–20.
- Sacks, H. 1995. *Lectures on Conversation*. Vol. I & II. Oxford Blackwell.
- Schegloff, E.A. 1984. “On Some Gestures’ Relation to Talk.” In Atkinson, J.M., Heritage, J. (eds): *Structures of Social Action*, Cambridge University Press, Cambridge: 266–98.
- Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Mass.: MIT Press.
- Sloetjes, H., and P. Wittenburg. 2008. “Annotation by Category: ELAN and ISO DCR.” In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. <http://www.lat-mpi.eu/tools/elan/>.
- Tannen, D. 1993. *Framing in Discourse*. Oxford University Press. New York, Oxford.
- Wells, J.C. 2006. *English Intonation PB and Audio CD: An Introduction*. Cambridge University Press. Cambridge.
- Wichmann, A. 2000. *Intonation in Text and Discourse*. Longman. London.

———. 2001. “Spoken Parentheticals.” *Gothenburg Studies in English* (81): 177–93.