Représentation, édition et exploitation de données multimodales : le cas des backchannels du corpus CID

Roxane Bertrand, Morgane Ader, Philippe Blache, Gaëlle Ferré, Robert Espesser, Stéphane Rauzy

Représentation, édition et exploitation de données multimodales : le cas des backchannels du corpus *CID*

R. Bertrand, M. Ader, P. Blache, G. Ferré, R. Espesser & S. Rauzy

Laboratoire Parole et Langage – Université de Provence
29, avenue Robert Schuman – 13621 Aix en Provence

INTRODUCTION GENERALE

La constitution et l'annotation de corpus multimodaux sont un enjeu essentiel pour la description linguistique et plus généralement la compréhension des mécanismes du langage. Il existe aujourd'hui plusieurs projets visant à constituer de telles ressources, et des conférences sont régulièrement organisées sur ce thème (cf. en particulier les workshops « Multimodal Corpora » lors des trois dernières éditions des conférences LREC [http://www.lrec-conf.org/]). Le recueil de corpus multimodaux pose en soi un certain nombre de problèmes, en particulier pour ce qui concerne la qualité du signal enregistré: chacune des modalités doit en effet pouvoir être analysée de façon précise, éventuellement à l'aide d'outils d'analyse automatiques. L'analyse du signal acoustique aux niveaux phonético-prosodiques requiert ainsi une qualité optimale imposant le plus souvent l'enregistrement de données en laboratoire. Mais les problèmes les plus importants résident dans l'annotation de telles données. Il n'existe à ce jour que très peu de projets se confrontant au problème d'une annotation aussi extensive que possible de corpus multimodaux, prenant en compte l'ensemble des domaines de l'analyse linguistique (phonétique, prosodie, morphologie, syntaxe, pragmatique, etc.) et des modalités différentes (gestes, parole). C'est ce que nous avons fait dans le cadre du corpus CID (*Corpus of Interactional Data*, Bertrand *et al.*, 2007) exposé dans cet article.

Notre objectif est de décrire, grâce à ce type de ressources, les modalités d'interaction existant entre les différents domaines linguistiques. Nous voulons plus précisément montrer l'intérêt de la prise en compte d'un ensemble important et varié d'annotations, dans chacun des domaines, pour expliquer ce type de phénomène. L'étude présentée ici porte sur les divers facteurs d'apparition (discursifs, prosodiques et gestuels) des signaux backchannels. Nous utilisons une méthode d'interrogation fondée sur l'outil *XSLT* qui nous permet, au travers de quelques requêtes, de mettre à jour un ensemble de régularités conditionnant leur apparition. Nos résultats confirment expérimentalement une partie de ceux de la littérature mais confirment surtout l'intérêt des ressources multimodales enrichies à divers niveaux.

1. LE CID (*CORPUS OF INTERACTIONAL DATA*): DISPOSITIF ET RECUEIL

Le CID, recueilli au LPL, compte à ce jour 8 x 1 heure de dialogue en français[1].

Chacun des dialogues met en présence 2 hommes ou 2 femmes. L'enregistrement, d'une durée d'une heure environ, est mené en chambre sourde : les sujets sont installés côte à côte, à environ 1 mètre de distance l'un de l'autre, ce qui représente une distance similaire à celle de locuteurs conversant librement. Chacun porte un micro-casque permettant de recueillir sa voix sur une piste séparée, ce qui garantit une qualité optimale favorisant l'analyse des données orales ainsi obtenues dans des logiciels de traitement de parole. Cela offre de surcroît la possibilité d'exploiter les phases de chevauchements de parole[2]. Enfin, les locuteurs sont filmés avec une caméra numérique (Canon XM2) en plan large et fixe (voir figures 1 et 3).

Les 16 locuteurs (10 femmes, 6 hommes), de langue maternelle française, sont pour la moitié d'entre eux natifs de la région PACA et pour l'autre moitié issus de diverses régions françaises. Au

[1] Le CID est voué à être enriché de nouveaux enregistrements et de nouvelles annotations.

[2] Celles-ci, en dépit de leur rôle dans les tours de parole, restent en effet souvent inexploitées car les logiciels de traitement du signal peinent encore à démêler les différentes voix.

moment de l'enregistrement, tous résident dans le Sud-Est de la France. Enfin, ils ont été choisis en fonction de leur familiarité avec le lieu d'expérimentation et d'un fort degré de connivence. L'expérimentateur leur précise qu'ils peuvent à tout moment se distancer du thème qui leur a été suggéré, et qui était surtout prétexte à faciliter la discussion, si nécessaire.

2. LES DIVERS NIVEAUX D'ANNOTATIONS DU CID

Le CID a été annoté sur plusieurs niveaux linguistiques dont l'exposé se limitera ici aux premières annotations sur lesquelles s'ancrent la plupart des autres (par ailleurs largement décrites dans Bertrand *et al.*, 2007) :

- 'transcription orthographique enrichie', dite TOE (effectuée par 2 experts), à partir de laquelle sont dérivées deux versions de transcription, l'une phonétique destinée aux niveaux phonético-prosodique et l'autre phonologique destinée aux niveaux morphologique et syntaxique
- phonétisation (DiCristo & DiCristo, 2001) et alignement (http://www.loria.fr/equipes/parole/) de la transcription phonétique avec le signal audio
- alignement des tokens orthographiques avec le signal

Ces 2 derniers niveaux sont cruciaux puisqu'ils servent de référence aux autres niveaux d'annotation en permettant notamment leur mise en relation ultérieure (entre autres temporelle).

- annotations aux différents niveaux morphosyntaxique, syntaxique, prosodique, pragmatique et gestuel, dont nous exposerons (en 4) seulement celles utilisées pour l'étude présente.

Selon les niveaux, les procédures d'annotations varient. Elles sont soit manuelles soit automatiques et elles impliquent l'utilisation d'outils et de logiciels différents. L'intérêt et l'objectif de ce projet est de proposer une approche intégrée permettant l'utilisation et l'exploitation simultanée de l'ensemble des informations disponibles, quel que soit leur mode de recueil par exemple.

Dans l'étude décrite ici, dont la problématique est clairement d'ordre linguistique, nous avançons quelques éléments de réponse à certaines difficultés que pose l'analyse de corpus multimodaux et nous présentons quelques-unes des solutions que nous avons adoptées pour permettre l'interrogation, dans un même formalisme, des multiples informations dont nous disposons.

Au préalable, le tableau 1 présente quelques données générales descriptives du CID permettant de mieux le caractériser:

Unit type	Total number	annotation process
Articulation duration	21.600 sec	automatic
Global pause duration	36.008 sec	automatic
*IPU**	13.873	automatic
words	103.034	semi-automatic
Acoustic word forms	110.521	semi-automatic
Speech overlaps	5.256	automatic

Tableau 1 : Statistiques descriptives pour les 8 heures du CID concernant la durée d'articulation avec et sans les pauses silencieuses, le nombre d'IPU (interpausal-units, blocs de parole compris entre deux pauses silencieuses), le nombre de mots (forme orthographique et forme phonétique) et le nombre de chevauchements de parole

Le tableau 2 comptabilise le nombre d'annotations du niveau prosodique. Chacune des catégories sera explicitée dans les points 4.5 et suivants.

Unit type	Total number	annotation process
Phonemes	272.166	automatic
AP	16.061	manual
IP	21.745	manual
Pitch C	25.997	manual

Tableau 2 : Nombre d'occurrences des unités phonético-prosodique (phonèmes, AP = accentual phrase, IP = Intonational Phrase, Pitch Contours = Contours intonatifs)

Le tableau 3 recense les informations du niveau morphosyntaxique, issues d'analyseurs automatiques.

Unit	type	Total number	Unit	type	Total number
Lexical categories	V	21.255	*Phrase types*	VP	16.046
	N	14.704		NP	25.687
	Adj	4.903		AP	3.517
	Adv	10.907		AdvP	6.989
	Prep	9.465		PP	9.465
	Coord	3.229		Relative clauses	2.198
	Det	5.368			
	Pro	8.704			

Tableau 3 : Statistiques du niveau morpho-syntaxique (à gauche les catégories lexicales V : verbe, N : nom, Adj : adjectif, Adv : adverbe, Prep : préposition ; Coord : coordination,, Det : déterminant, Pro : pronom ; à droite les syntagmes : VP : syntagme verbal ; NP : syntagme nominal, AP : syntagme adjectival, AdvP : syntagme adverbial, PP : syntagme propositionnel)

Le tableau 4 recense les informations du niveau gestuel (dont l'annotation est totalement manuelle) sur un 1/4 d'heure du corpus (2 locuteurs), à savoir l'extrait considéré ci-après pour l'étude des BC.

Movement / gesture type	Total number	Annotation process
Gaze	2390	manual
Head	2089	
Eyebrows	506	
Mouth	396	
Hands	281	

Tableau 4 : Statistiques du niveau gestuel concernant le regard, la tête, les sourcils, la bouche et les mains

3. REPRESENTATION ET EDITION DES DONNEES

Nous avons choisi le logiciel *ANVIL* (Kipp, 2003-2006) pour regrouper les différentes annotations existantes (cf. figures 2 et 3). *ANVIL* offre la possibilité d'importer des annotations réalisées sous d'autres logiciels tels que *Praat* (Boersma, 2005) par exemple pour les dimensions phonético-prosodique, ou l'étiqueteur morphosyntaxique développé au LPL (intégré à la chaîne de traitement LPLSuite, VanRullen, 2005). Comme souligné par Chen *et al*. (2006), l'intérêt de regrouper les différentes annotations dans *ANVIL* est double: non seulement le logiciel permet de rassembler et d'aligner des informations de nature différente, mais il permet aussi de mieux visualiser l'effet des différents paramètres sur le phénomène linguistique étudié.

Enfin, nous avons adopté *ANVIL* pour ses formats d'entrée et de sortie qui satisfont à la structure du format XML, standard actuel sur lequel sont développés les outils d'interrogation (voir 5).

4. EXPLOITATION MULTIMODALE DU CID : ETUDE EXPLORATOIRE DES SIGNAUX BACKCHANNEL (BC)

Nous avons choisi d'illustrer notre démarche autour des corpus en présentant une étude linguistique autour des *backchannel/signaux d'écoute*. Certes exploratoire, cette étude permet de rendre compte des difficultés mais aussi et surtout de l'intérêt d'un tel projet visant la constitution et l'exploitation multimodale de ressources orales dans le champ des sciences du langage.

L'étude porte sur un extrait des 15 premières minutes d'une interaction entre 2 hommes (cf. fig. 1).

4.1 Introduction

Le terme de *backchannel*[3] *(désormais BC)*, introduit par Yngve (1970), est employé de manière générique pour référer à l'ensemble des signaux verbaux, vocaux et gestuels, émis par l'interlocuteur d'un dialogue pour montrer son écoute, sa compréhension, son accord, etc. au discours produit.

En une décennie environ, les BC sont devenus un objet d'étude extrêmement investi aux différents niveaux de l'analyse linguistique (Schegloff, 1982 ; Cosnier, 1988 ; Koiso *et al.*, 1998; Ward & Tsukahara, 2000). D'abord plutôt étudiés par les deux courants de *l'Analyse Conversationnelle* (Sacks

[3] On recense également d'autres termes pour renvoyer à ces phénomènes, comme celui de *feedback* ou de *régulateurs* (Cosnier, 1988).

et al., 1974 ; Schegloff, 1982 ; Couper-Kuhlen & Selting, 1996, Couper-Kuhlen & Ford, 2004 parmi d'autres) et de *l'Analyse des Interactions* (Cosnier, Kerbrat-Orecchioni) dans les années 80-90, ils ont reçu depuis une attention sans cesse croissante de la part des différentes communautés de recherche sur la parole en raison de l'intérêt surgissant pour l'étude de la parole naturelle et spontanée en contexte de dialogues.

Cependant, bien que la littérature actuelle soit foisonnante sur ces phénomènes, elle présente aussi l'inconvénient majeur de la dispersion. Car si chacun s'est emparé de cet objet d'étude tant en prosodie (Bertrand, 1999 ; Bertrand & Espesser, 2003 ; Caspers, 1998, 2003; Cerrato & D'Imperio, 2003), qu'en psycholinguistique (Fox Tree, 1999, 2002), en reconnaissance automatique de la parole (Heldner & Edlund, 2006; Ward & Tsukahara, 2000) ou encore dans le domaine de l'acquisition des langues et des différences interculturelles (Stubbe, 1998 ; Allwood & Ahlsen, 1999), chacun l'a fait au sein de sa communauté, en prenant rarement en compte les résultats des travaux extérieurs à son propre champ d'investigation. Quelques études seulement ont tenté de mener des travaux dans une double perspective, notamment Koiso *et al*. (1998) qui ont cherché à rendre compte du poids respectif des indices syntaxiques et prosodiques dans l'apparition des backchannels verbo-vocaux, ou encore en prosodie dans le cadre de *l'Analyse Conversationnelle* (Caspers, 2003; Portes & Bertrand, 2006).

Par ailleurs, les BC n'ont pas reçu d'investigations systématiques concernant leur caractère multimodal : très peu d'études en effet se sont intéressées aux éventuelles différences entre les uns et les autres ou bien à la manière dont ils co-existent (Bertrand *et al.* 1995 ; Allwood & Cerrato, 2003). Nous expliquons ces lacunes précisément par l'absence de corpus favorisant de telles études.

L'analyse des BC interroge plus largement la question du fonctionnement des tours de parole. Les participants à une interaction ont à leur disposition diverses ressources grâce auxquelles ils projettent ou anticipent une fin de tour de parole (Ford & Thompson, 1996 ; Barkhuysen *et al.,* 2006 ; Auer, 1996, entre autres). De la même manière, divers facteurs sont impliqués dans l'apparition des BC : il peut s'agir d'une fin d'unité syntaxique, de contours intonatifs spécifiques, d'un changement d'orientation du regard, etc. De plus, si ces indices jouent un rôle dans l'apparition des BC, il s'avère selon nous indispensable de les prendre en compte simultanément afin d'en déterminer le rôle et le poids relatifs (Koiso *et al.*, 1998 ; Blache & Di Cristo, 2002 ; Di Cristo *et al.*, 2004). Aucune étude, sur le français de surcroît, n'a été menée en ce sens. Une ressource telle que le CID offre la possibilité de réaliser un tel travail, et ce à un niveau de description extrêmement fin pour chacun des niveaux d'annotation concernés (voir le détail des annotations plus loin).

4.2 Hypothèse

Les BC fournissent de l'information non seulement sur le processus d'écoute des interlocuteurs mais également sur le processus de production des discours des locuteurs (Fox Tree, 1999). En effet, ils marquent ou ponctuent des étapes importantes dans l'élaboration du discours. Ces étapes, signalées par divers indices, seront ratifiées à la seule condition de recevoir une réponse adaptée, c'est-à-dire attendue. Le locuteur produisant un contour intonatif typique créée ainsi une attente particulière qui peut être comblée par une réponse spécifique tel qu'un BC (Caspers, 1998, Marandin, 2004, Portes *et al.,* 2007). Les BC ont donc un réel impact sur le discours produit (Fox Tree, 1999), une attente non satisfaite pouvant donner lieu à diverses séquences parallèles s'achevant lorsque la réponse attendue a été obtenue (Kern, 2007).

4.3 Objectif

Bien que de plus en plus d'études s'attachent à décrire les corrélats prosodiques, syntaxiques ou gestuels des BC, il n'existe pas, en revanche, de travaux sur le français prenant en compte systématiquement l'ensemble des indices linguistiques dont l'étude contribuerait 1/ à améliorer les typologies existantes, et 2/ à mieux comprendre le fonctionnement des tours de parole en cernant davantage le rôle respectif et relatif des différentes ressources disponibles aux locuteurs.

Dans ce travail, nous nous centrons plus particulièrement sur le rôle des facteurs prosodiques, discursifs et conversationnels dans l'apparition des backchannels vocaux et gestuels.

4.4 Typologie générale des BCs

Il existe plusieurs classifications fonctionnelles des BC, parmi lesquelles celle de Schegloff (1982) par exemple qui distingue entre *continuers* et *assessments*. Les premiers ont une fonction d'*accusé-réception* : ils expriment l'attention mais aussi l'intérêt et la compréhension de l'interlocuteur. Les seconds ont une fonction de *prise de position :* l'interlocuteur montre son accord avec le locuteur.

Plus récemment sur le français québécois, outre les accusés-réception, Laforest (1992) distingue les régulateurs à fonction de *soutien* (proches des *assessments)*, et ceux à fonction de *support* qui renvoient aux attitudes de l'interlocuteur (exclamation, commentaire évaluatif). L'auteur identifie également une catégorie à fonction de *relance* destinée à encourager le locuteur à poursuivre même si ce dernier est prêt à céder son tour, et une dernière catégorie à fonction *indéterminable*. Elle oppose en outre les régulateurs *simples*, c'est-à-dire qui ne constituent pas un réel tour de parole, aux régulateurs *complexes*, qui renvoient aux divers cas de *reformulation*, *complétion*, *répétition* et *métaquestion* dont le statut de *non tour* s'avère plus délicat à établir.

Une autre typologie concerne le Japonais (Maynard, 1989), l'une des langues les plus étudiées, semble t-il, de ce point de vue. L'auteur distingue 6 fonctions : 1/ continuer, 2/ display of understanding of content, 3/ support toward the speaker's judgement, 4/ agreement, 5/ strong emotional response, 6/ minor addition, correction, or request of information. Les catégories adoptées dans ce travail, très similaires, sont les suivantes :

1/ ct *: Continuer* (prendre note minimalement)
2 udg : *Understanding* (j'ai bien compris mais sans notion d'adhésion, degré supérieur au ct)
3/ ack *: Acknowledgement* (support, adhésion à un propos)
4/ as : *Assessment* (évaluation, -rire par exemple-, jugement, déclaration d'attitude)
5/ (c)rt : R*equest/Confirmation request*
6/ *Complex*

4.5 Quels niveaux d'annotation ?

Un travail visant à améliorer les typologies formelles et fonctionnelles des backchannels nécessite le repérage des éléments verbaux, vocaux et gestuels susceptibles de fonctionner comme des BC mais qui comportent aussi, souvent, d'autres fonctions discursives. Le cas de *ouais* constitue en ce sens l'un des meilleurs exemples tant sa nature polysémique n'est plus à démontrer : il peut en effet fonctionner non seulement comme un BC mais aussi comme une simple réponse (dans le couple question/réponse), mais il peut aussi initier un tour (*turn-initiator*) ou encore une réparation (*self-repair initiator*). De la même manière, le geste est par nature polysémique : il peut fonctionner aussi comme simple réponse ou comme marqueur de prise de tour. Il peut également comporter une fonction de renforcement, visant soit à renforcer une focalisation intonative, soit à renforcer un autre geste. La nature polysémique des backchannels, non exclusive de ces derniers et qui concerne de nombreux phénomènes langagiers, rend la tâche d'identification très délicate et justifie le recours à plusieurs annotateurs qui permet parfois de réduire les cas d'incertitude.

4.5.1 Annotation discursive et conversationnelle

C'est à ce niveau que les BC ont été annotés, par 2 experts (parmi les auteurs). Pour les BC vocaux, ils ont distingué à la suite de Laforest (1992) les BC simples (*ouais* et ses différents composés comme *ah ouais*, *eh ouais*, ainsi que *mh*, *ok*, *voilà*, *non*, *d'accord*, *ah bon*) des BC complexes. Une étiquette *autre* a été ajoutée pour des BC simples moins fréquents du type *ah.* L'ensemble des BC vocaux et gestuels ont été annotés ensuite d'un point de vue fonctionnel, conformément à la typologie exposée en 4.4. Cette annotation a été effectuée en se fondant notamment sur le contenu du discours précédent et en s'appuyant, si nécessaire, sur la prosodie associée au BC considéré. L'un des experts a annoté les seuls BC vocaux à partir du signal audio, l'autre les BC gestuels à partir du film. Ils ont ensuite confronté leurs annotations et tenté de s'accorder sur les cas problématiques. Des faits intéressants ont émergé durant cette phase de confrontation : par exemple, lors de la production simultanée d'un BC vocal et gestuel, les experts ont noté que les deux BC n'avaient pas nécessairement reçu la même valeur. Ces éventuelles divergences, illustrant une certaine indépendance des niveaux, ont été conservées si elles paraissaient légitimes. Elles ont permis également de mettre à jour une gradation (liée au degré de « prise en compte » par l'interlocuteur) dans les fonctions des BC (hormis *crt*) qui s'est notamment traduite par la modalité : pour des BC produits à la fois vocalement et gestuellement, le BC gestuel a souvent été considéré avec un degré supérieur. Pour un BC vocal annoté *ct*, le BC gestuel co-produit a été annoté *udg*. Ce décalage pourrait s'expliquer simplement par le fait que le film fournit un maximum d'informations qui, même lorsqu'on s'attache à une seule modalité, influent sur la perception des phénomènes.

A ce niveau, les marqueurs discursifs ont été également annotés selon la typologie suivante :

- *Connector:* mot(s) grammatical(aux) servant à relier entre elles deux unités discursives telles que les tours de parole par exemple (Calbris, 2002; Bouvet, 2001). Morel & Danon-Boileau (1998) les appellent *ligateurs*.
- *Punctuator:* mot(s) ou expression(s) apparaissant en fin d'énoncés tels que *quoi*, *bon* etc.
- *Phatic:* mots ou expressions telles que *hein*, *tu vois*, *tu sais* etc. faisant appel à l'interlocuteur.

Enfin, le CID a été annoté en unités conversationnelles telles qu'elles sont définies par les tenants de *l'Analyse Conversationnelle* (CA). Ces unités, communément appelées les 'unités de construction de tours' (*Turn-constructional units* ou *TCU*), sont définies comme 'les plus petites unités linguistiquement complètes, pertinentes au niveau interactionnel' (Selting, 2000). Nous adoptons cette définition et plus globalement la perspective de Selting qui a proposé des solutions intéressantes pour décrire des corpus tels que le CID. Celui-ci en effet, en raison de la consigne initiale donnée aux locuteurs, comporte de nombreuses séquences de narration ou d'explication que l'on peut décrire comme des cas d'unité complexes ('*multi-unit*'). Selting propose alors de distinguer les TCU des TRP (*Transition-Relevance-Place*). Le TCU n'est plus une unité de tour devant nécessairement se terminer dans une TRP mais il peut être une simple « partie » de tour, elle-même intégrée dans une unité plus complexe. Les TCU peuvent donc être 'finaux' (*TCU_f*), c'est-à-dire complets en terme syntaxique, prosodique et pragmatique, ou 'non-finaux' (*TCU_nf*) définis alors comme un des composants incomplets (d'un point de vue pragmatique par exemple) d'un tour complexe. Enfin, on peut trouver également des cas de *continuations* de tours (*turn-continuation*). Ceux-ci réfèrent aux cas pour lesquels le locuteur semble avoir atteint son objectif : il a donc produit un tour complet, et semble prêt à vouloir céder son tour, lorsqu'il commet une nouvelle proposition qui n'est pas un nouveau TCU dans la mesure où elle entretient encore un lien fort avec ce qui précède (Vorreiter, 2003).[4] Dans le corpus global, 3 heures ont été annotées en *TCU_f*, *TCU_nf* et *Cont*.

4.5.2 Annotation prosodique

La prosodie opère selon une organisation tripartite fondée sur trois axes : intonatif, temporel et métrique (Di Cristo *et al.*, 2004). Chacun fonctionne comme un sous-système selon des principes spécifiques, tout en étant fortement dépendant des deux autres. Il est donc fondamental de distinguer ces trois axes selon ce que l'on cherche à observer et de disposer des informations relatives à chacun d'eux. Nous avons considéré principalement, à ce jour, le niveau intonatif. Concrètement, trois niveaux d'annotation ont été effectués:

- le *phrasé prosodique :* renvoie au découpage du discours en unités prosodiques. Sont annotées les *unités intonatives* ('IP' = *Intonational Phrase*) et *accentuelles* ('AP' = *Accentual Phrase*) qui sont les plus communément admises pour le français (pour une revue voir Jun & Fougeron, 2002; D'Imperio *et al.*, 2007). Une troisième catégorie 'EP' (= *external phrase)* permet de recenser les éléments inclassables dans les deux premières catégories.

Ce niveau d'annotation, basé sur la perception, a été effectué par des experts sur 6 heures du CID.

- les *contours intonatifs* (*Pitch Contours*) qui réfèrent à des constructions associant une forme à une fonction[5] (Portes *et al.*, 2007) et dont les catégories sont les suivantes :
- fl : plat/*flat*
- F : descendant/*falling*
- mr : montée mineure/*minor rising*; m0 : autres contours mineurs/*other minors*
- RF1 : montant-descendant/*rising-falling*; RF2 : montant-descendant depuis l'avant dernière syllabe/*rising-falling from penultimate*
- RMC : montée de continuation majeure/*rising major continuation*; RT : montée terminale/*terminal rising*
- RQ : montée de question/ *question rising*
- ER : montée d'énumération/ *enumerative rising*; EF : descente d'énumération/*enumerative falling*

Les contours intonatifs ont été annotés manuellement par 2 experts sur 6 heures. L'annotation, fondée sur la perception, a nécessité aussi une vérification acoustique ultérieure, alourdissant considérablement le travail des annotateurs.

- Le troisième niveau d'annotation est effectué automatiquement grâce à un outil développé au LPL (*MOMEL-INTSINT*, Hirst *et al.*, 2000) qui permet de repérer et d'encoder automatiquement les cibles tonales grâce à un alphabet de 8 symboles :

[4] 'Increment', 'add-on', etc., relèvent également de cette catégorie (pour une revue, voir Vorreiter, 2003).

[5] Nous listons les contours intonatifs communément admis en français mais aussi des configurations telles que mr, ER, RQ qui sont utilisées dans le cadre de nos travaux sur les contours montants (Portes *et al.,* 2007).

*T*op, *Mid*dle et *B*ottom sont définis globalement par rapport au registre de chaque locuteur, *H*igher, *S*ame et *L*ower par rapport aux points précédents, *D*ownstepped et *U*pstepped également par rapport aux points précédents mais ils concernent des changements de plus faible ampleur.

4.5.3 Annotation gestuelle

Parmi l'ensemble des gestes annotés, nous avons retenu les gestes de la tête (et non pas la direction de la tête), les expressions faciales telles que les rires et les sourires, les mouvements des sourcils et la direction du regard de chaque locuteur.

L'annotation des gestes (bien entendu manuelle) s'est faite en deux temps : nous avons d'abord procédé à un repérage des gestes et des mouvements que nous avons bornés, puis nous leur avons attribué une valeur (phatique, de renforcement, backchannel, réponse, etc.), en adoptant la même typologie que pour le verbal.

Typologie des gestes :

- *Head :* les mouvements de tête que nous avons retenus sont quasi-similaires à ceux retenus par Cerrato & Skhiri pour le Suédois (2003a et b) et Allwood & Cerrato (2003):
  - *nod* – acquiescement
  - *shake* – geste de négation de la tête
  - *turn* – mouvement de la tête vers le côté (en précisant celui dont il s'agit)
  - *jerk* – mouvement de la tête vers l'arrière sur un plan vertical
  - *tilt* – inclinaison de la tête à droite ou à gauche
  - *waggle* – dodelinement
  - *other* – nous avons annoté dans cette même catégorie les mouvements de la tête vers l'avant ou l'arrière sur un plan horizontal, catégorie peu décrite dans la littérature.

Pour tous les mouvements de tête, nous avons noté si les mouvements sont uniques ou répétés (single vs. repeated), même si à ce stade de l'étude, nous n'avons pas exploité cette annotation.

- *Facial expressions :* parmi les expressions faciales, nous avons différencié les rires des sourires (*laughter* vs. *smiles*). Nous avions également prévu une étiquette pour noter l'air renfrogné (*scowl*) mais cette expression n'est pas apparue sur l'extrait considéré.
- *Eyebrow movements :* pour les mouvements des sourcils, nous avons noté les sourcils froncés (*frowning*) et les sourcils haussés (*raising*).

Enfin, nous avons annoté la direction du regard du locuteur et de l'interlocuteur, en adoptant une typologie légèrement différente car le regard est régulièrement utilisé par les locuteurs pour gérer les tours de parole. Nous avons donc noté dans un premier temps la direction du regard de chaque participant, puis nous lui avons attribué une fonction.

Typologie du regard :

- *Direction :* nous avons noté de manière systématique la direction du regard des interactants, mais pour cette étude précise, nous utilisons les *regards mutuels* ou *absence de regard mutuel* que nous ne codons pas directement dans l'annotation mais qu'il est aisé de retrouver par la suite à l'aide de simples requêtes.
  - *up* – vers le haut (de face, à droite ou à gauche)
  - *down* – vers le bas (de face, à droite ou à gauche)
  - *sideways* – vers le côté (droite ou gauche)
  - *front* – en face de soi
  - *wandering* – regard vagabond
  - *towards interlocutor* – vers l'interlocuteur
  - *towards object* – vers un objet concret ou abstrait
- *Fonction :* nous avons noté trois fonctions afin de rendre compte de ce que Chen *et al.* (2006) appellent « floor control ». Ainsi « speaking » marque le locuteur patenté et « listening » l'auditeur. Nous avons attribué une fonction phatique pour un regard de durée réduite du locuteur vers l'interlocuteur lorsque le rôle de ce contact oculaire est de susciter un BC de l'interlocuteur et non pas de céder le tour de parole.

4.5.4 Configuration de l'annotation sous ANVIL

Pour ce travail relatif au backchannel, nous avons configuré *ANVIL* afin qu'il n'affiche dans l'annotation que les événements que nous imaginons susceptibles d'avoir une incidence sur son apparition. Pour cette étude spécifique où la notion de complémentarité entre les deux locuteurs est donc essentielle, nous avons réalisé une annotation regroupant les deux locuteurs, ce qui n'est pas le cas des annotations du CID (utilisées pour d'autres études) qui sont généralement créées pour chaque locuteur dans des fichiers distincts.

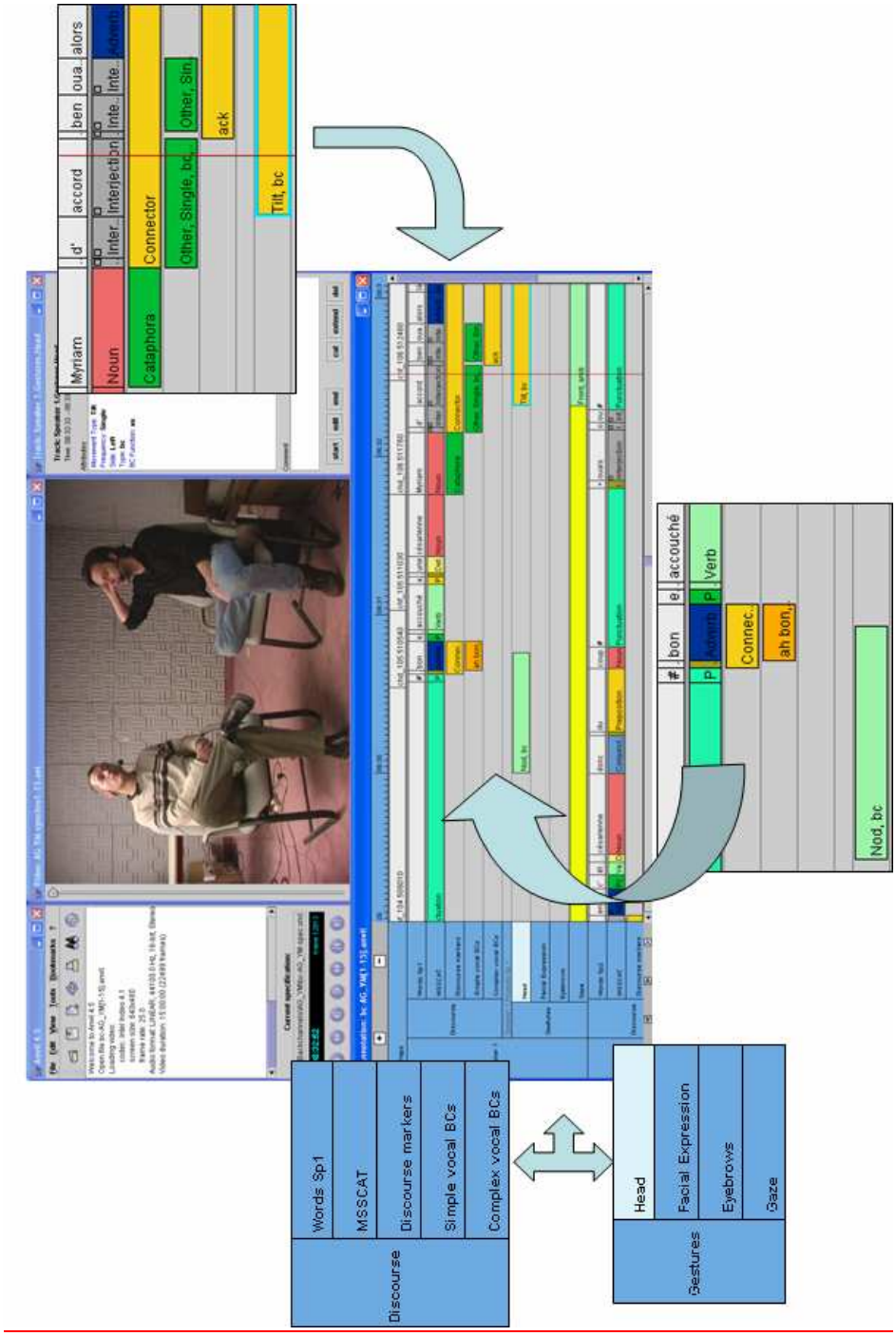

Figure 1. Annotation des backchannels sous ANVIL

La figure 1 montre comment s'organise l'annotation dans *ANVIL* : la fenêtre d'annotation est séparée en trois groupes : les phénomènes concernant le locuteur 1 (à gauche sur la vidéo) et ceux qui concernent le locuteur 2 (à droite sur la vidéo) forment les deux premiers groupes qui sont strictement identiques en ce qui concerne les pistes d'annotation et les étiquettes employées dans chaque piste. Le troisième groupe affiche les informations morphosyntaxiques.

Dans la piste supérieure est affichée la transcription des paroles prononcées par le locuteur. Cette annotation ainsi que celles des trois pistes suivantes (unités intonative et accentuelle, contours intonatifs, TCU finaux, non-finaux et de continuation) a été effectuée sous Praat.

Les trois pistes suivantes concernent l'annotation des marqueurs discursifs. Dans les deux pistes suivantes sont annotés les backchannels vocaux simples et complexes même si seuls les premiers seront examinés dans ce travail.

Les quatre pistes suivantes sont dédiées à l'annotation de certains mouvements et gestes de chacun des locuteurs.

Dans le dernier groupe enfin, les deux premières pistes comportent la transcription orthographique nécessaire aux niveaux morphosyntaxique et syntaxique. Les deux pistes suivantes concernent la morphosyntaxe pour le locuteur 1: la piste supérieure détaille les catégories morphosyntaxiques du type [pronom personnel 3ème personne masculin singulier], avec un nombre total d'étiquettes très élevé ; dans la piste inférieure nous avons, pour plus de lisibilité, réduit leur nombre à 11 catégories.

Avant de passer à l'étape d'interrogation du corpus, nous présentons quelques données chiffrées des annotations réalisées sur l'extrait examiné. Les tableaux 5 et 6 concernent le niveau prosodique pour chacun des deux locuteurs.

Type d'unité	**Nombre total**	
Unités prosodiques	**Loc1**	**Loc2**
IP	293	351
AP	184	261
EP	11	52
total	**488**	**664**

Tableau 5 : Nombre d'unités intonatives (IP), accentuelles (AP) et « externes » (EP), sur l'extrait observé

	Nombre d'occurrences des contours intonatifs												
	F	*RF1*	*RF2*	*RMC*	*RQ*	*RT*	*ER*	*fl*	*m0*	*mr*	*REM*[6]	*?*	***TOTAL***
locuteur 1	2	24	6	73	22	16	2	144	59	125	10	3	**486**
locuteur 2	3	30	7	110	8	25	1	210	99	162	8	3	**663**

Tableau 6 : Nombre de contours intonatifs sur l'extrait observé (F : descendant, RF1 et 2 : montant-descendant, RMC : montant de continuation, RQ : montant de question, ER : montant d'énumération,fl :plat, m : mineurs ? = ambigu)

Le tableau 7 présente les annotations du niveau conversationnel pour chacun des locuteurs.

Type d'unité	**Nombre total**	
	Loc1	**Loc2**
TCU_f	128	154
TCU_nf	57	86
Cont	27	36
total	**212**	**276**

Tableau 7 : Nombre d'unités conversationnelles sur l'extrait observé (tour de construction final, non final et continuation de tour)

Le tableau 8 présente le nombre de marqueurs discursifs pour chacun des locuteurs.

Type d'unité	**Nombre total**	
	Loc1	**Loc2**
Marqueurs discursifs		
connecteur	45	71
ponctuant	16	54
phatique	30	14
total	**91**	**139**

Tableau 8 : Nombre de marqueurs discursifs sur l'extrait observé

Avec près de 1000 occurrences d'unités prosodiques et de contours intonatifs, 500 unités conversationnelles et plus de 200 marqueurs discursifs pour les deux locuteurs confondus, nous pouvons effectuer, sur ce seul extrait d'1/4 d'heure, des premières analyses statistiques. Le fichier *ANVIL* résultant de ces diverses annotations est le fichier XML sur lequel nous allons réaliser les différentes requêtes.

[6] REM *n'est pas un contour au sens strict mais constitue une emphase (avec contour montant)*

5. OUTILS D'EXPLOITATION ET D'INTERROGATION DU CID

La constitution de telles ressources en termes d'annotations et leur synchronisation vise leur exploitation simultanée. Nous avons développé dans ce but des outils permettant l'interrogation des informations disponibles selon des requêtes plus ou moins simples, les plus simples étant celles requérant un seul niveau d'information, les plus délicates celles qui en requièrent plusieurs.

5.1 Premiers résultats concernant les fonctions des BC

A partir d'un premier jeu de requête extrêmement simple, nous avons extrait dans un premier temps les éléments pouvant fonctionner soit comme BC soit comme réponse, initiateur de tour, etc.

Nous avons ensuite recherché pour chaque BC la fonction qui lui a été assignée (cf. 4.4).

Le tableau 9 suivant présente les résultats pour les BC vocaux :

	BC VERBO-VOCAUX SIMPLES								
%	ouais	ah ouais	eh ouais	ok	voilà	mh	non	ah bon	d'accord
ct	5.6	0	0	0	0	16.7	0	0	0
udg	**33.3**	**33.3**	**75**	**100.0**	0	**50**	0	0	**100**
ack	**44.4**	**16.7**	**25**	0	**100**	**33.3**	**100**	0	0
as	7.4	**16.7**	0	0	0	0	0	0	0
rt	3.7	0	0	0	0	0	0	0	0
crt	5.6	**33.3**	0	0	0	0	0	**100**	0

Tableau 9 : % des BCs vocaux selon leur fonction discursive (ct : continuer, udg : understanding, ack : acknowledgement, as : assessment, rt : request, crt : confirmation request)

D'ores et déjà, et sous réserve de les valider sur l'ensemble du corpus, nous pouvons dégager quelques tendances. Certains BC vocaux semblent avoir une fonction unique : *voilà* et *non* ont en effet essentiellement une fonction de support (*ack*) ; *d'accord* et *ok* une fonction de compréhension (*udg*) et *ah bon* une fonction de demande de confirmation (*crt*). D'autres, comme attendu, sont plus polysémiques : s'il est essentiellement associé à *ack*, *ouais* apparaît toutefois dans une proportion importante en *udg*. C'est exactement l'inverse pour *eh ouais* et *mh*. Ce dernier est également l'un des seuls (avec *ouais,* dans une moindre mesure) à apparaître dans la catégorie de continueur (*ct*). Enfin, *ah ouais* est particulièrement polysémique puisqu'il apparaît de manière assez importante dans 4 catégories fonctionnelles différentes. Il est intéressant de constater aussi que la fonction *ct* traditionnellement attachée aux BC, est la plus faiblement représentée dans cet extrait. Ce point mériterait d'être validé sur l'ensemble du corpus afin de déterminer s'il s'agit d'une particularité de l'extrait, des locuteurs ou plus globalement des interactions du CID.

Dans le tableau 10 suivant, nous présentons les catégories gestuelles ayant été caractérisées comme des backchannels. Trois types de gestes ont été retenus : les mouvements de la tête, les expressions du visage et les mouvements des sourcils.

	HEAD MOVEMENTS							**FACIAL EXPRESSION**		**EYEBROW MOVEMENTS**	
%	jerk	nod	other	shake	tilt	turn	waggle	smile	laughter	frowning	raising
ct	0.0	**14.3**	0.0	0.0	0.0	0.0	0.0	3.7	0.0	0.0	0.0
udg, E	**29.7**	**34.7**	0.0	**14.3**	0.0	**60.0**	0.0	**11.1**	**35.0**	0.0	5.2
udg, U	**21.6**	6.1	0.0	0.0	0.0	**20.0**	0.0	3.7	10.0	0.0	**26.3**
ack	8.1	**34.7**	0.0	**14.3**	0.0	0.0	0.0	3.7	0.0	0.0	10.5
as	**18.9**	6.1	**50.0**	**28.6**	**90.0**	0.0	0.0	**70.4**	**50.0**	0.0	**52.6**
rt	**13.5**	2.0	**50.0**	14.3	0.0	0.0	0.0	0.0	0.0	**66.7**	0.0
crt	8.1	2.0	0.0	**28.6**	10.0	**20.0**	**100.0**	7.4	5.0	**33.3**	5.2

Tableau 10 : % des BC gestuels selon leur fonction discursive (on distingue ici entre UDG, U – Understanding, unexpected : une prise de conscience en décalage par rapport aux attentes initiales de l'interlocuteur, et UDG, E – Understanding, expected : l'expression de la compréhension sans qu'il y ait de décalage par rapport aux attentes initiales de l'interlocuteur)

Ces premiers résultats permettent là encore de dégager quelques tendances intéressantes : parmi celles-ci, certaines contredisent notamment des points relativement avérés dans la littérature, tels que les fonctions attribuées généralement aux *haussements de sourcil*s. En effet, ces derniers sont régulièrement observés en contexte d'interrogation. Or, nous constatons que pour nos 2 locuteurs, l'inverse se produit : ce sont ici les *froncements de sourcils* qui sont exclusivement employés pour marquer ce type d'acte de langage (66.7% dans le cas des interrogations -*rt*- et 33.3 % dans le cas des demandes de confirmation -crt-) alors que les *haussements de sourcils* sont employés pour marquer la compréhension -*udg*- (26.3 % et 5.2 %)[7], mais surtout l'évaluation –*as*- (52.6 %).

Les *expressions faciales* expriment quant à elles plutôt une évaluation de la part de l'interlocuteur (ceci est d'ailleurs plus vrai pour les sourires que pour les rires), mais un certain nombre d'entre elles expriment également une compréhension sans évaluation. Les *sourires* sont légèrement plus polysémiques que les *rires* dans la mesure où ils peuvent également marquer un retour sans prise de position marquée de l'interlocuteur, tel que *ct* et *ack*.

En ce qui concerne les *mouvements de tête*, le BC minimal est exclusivement exprimé par un simple *hochement de tête*, les autres mouvements exprimant toujours une part plus grande d'implication. Faute de données nous ne commenterons pas les rubriques *waggle* et *other*. Plus intéressants en revanche sont les *tilts* qui ont une fonction d'évaluation quasi systématique (90%) contre une fonction de demande de confirmation dans seulement 10% des cas, ce qui rend ce geste très peu équivoque. Parmi les mouvements très peu polysémiques, on trouve également les *turns* puisque la grande majorité d'entre eux sont associés à la compréhension (60 % d'*udg*, E et 20 % d'*udg,* U), les 20 % restants étant interprétés comme demande de confirmation. Les trois mouvements restants *jerks*, *nods* et *shakes* sont beaucoup plus polysémiques que les mouvements de tête décrits précédemment. Les *jerks* marquent majoritairement la compréhension (29.7 % d'*udg*, E et 21.6 % d'*udg*, U)[8]. Ils sont également employés pour marquer l'évaluation (18.9 % d'*as*) et dans une moindre mesure l'interrogation (13.5 % de *rt*). Les *nods*, dont on a dit plus haut qu'ils pouvaient marquer un retour minimal *(ct*), marquent également à pourcentage égal la compréhension (34.7 % *d'udg*, E) et le support (*ack).* Enfin, les *shakes* marquent en grande majorité l'interrogation (28.6 % de *crt* et 14.3 % de *rt*). Ils marquent aussi l'évaluation dans 28.6 % des cas. Dans une proportion plus faible, ils expriment le support et la compréhension (14.3 % respectivement). C'est donc le geste le plus polysémique et bien que nous n'ayons pas annoté le désaccord (tout simplement parce que nous n'en avons trouvé aucune occurrence en contexte de BC), il est clair que le *shake* est le geste le plus employé si l'on considère l'ensemble des contextes, et qu'il marque très rarement le désaccord. Comme backchannels, les mouvements de tête les plus fréquents sont les *nods*, les *jerks* et les *tilts.* (cf. le tableau 11 relatif aux divers mouvements de tête relevés dans l'extrait considéré):

nods	jerks	shake	tilts	waggles	turns	other
49	37	7	10	1	5	2

Tableau 11 : nombre d'occurrences des mouvements de tête à fonction de BC dans l'extrait examiné

5.2 Requêtes complexes permettant d'interroger plusieurs niveaux d'annotation

L'interrogation simultanée des différents niveaux d'annotation des corpus multimodaux nécessite le développement d'outils de génération de requêtes. Nous présentons ci-après la méthode utilisée sur le CID et quelques exemples de requêtes.

5.2.1 XSLT

Le fichier de sortie généré par *ANVIL* est au format XML. Il existe deux méthodes pour extraire des données d'un document XML : XSLT et XQuery. Ces outils sont quasiment identiques et s'appuient tous les deux sur XPath. Ils sont également reconnus par la norme W3C.

Un document XSLT est en fait une feuille de style que l'on applique à un fichier XML, qui porte l'extension .xsl. Cette feuille de style contient les règles que l'on souhaite suivre pour un ensemble des éléments du document XML. Le résultat d'une requête en XSLT peut être un nouveau document XML, un document HTML ou bien n'importe quel format texte.

[7] Et notamment une prise de conscience en décalage par rapport aux attentes préalables de l'interlocuteur (UDG, U), -ce qui serait à rapprocher des études liant les haussements des sourcils à l'expression de la surprise.

[8] Ce sont d'ailleurs quasiment les seuls gestes, avec les *turns*, à marquer une compréhension en décalage par rapport aux attentes préalables de l'interlocuteur.

L'intérêt du XSLT est que l'on peut exécuter les requêtes sans outil particulier, avec un simple navigateur web. Il suffit pour cela d'insérer une ligne au début du document XML qui va faire référence à la feuille de style utilisée, comme le montre l'exemple ci-dessous :

```
<?xml-stylesheet type="text/xsl" href="nom_de_la_feuille_de_style.xsl" ?>
```

Le navigateur sait interpréter XSLT et va donc transformer l'affichage du document XML en suivant les règles énoncées dans la feuille de style.

Il existe néanmoins des outils qui permettent d'exécuter des requêtes et de créer des documents de sortie, au lieu de simplement modifier l'affichage du document XML.

5.2.2 Adaptation des fichiers générés par ANVIL

La majorité des requêtes que nous souhaitions exécuter portait sur le temps des évènements. Or, le fichier de sortie généré par *ANVIL* ne donne pas le même format de temps pour les pistes primaires et pour les pistes secondaires.

Dans le cas des pistes primaires, le temps indiqué pour chaque évènement correspond bien au temps absolu. Cependant, sur les pistes secondaires, le temps donné pour chaque événement correspond en fait à l'index de l'événement de la piste primaire qui lui est associé.

Par exemple, si les informations d'un événement d'une piste secondaire sont de la forme :

`<el start= « 12 » end = « 15 »>`, cela signifie en réalité que cet événement commence en même temps que l'événement 12 de la piste primaire associée et se termine avec l'évènement 15.

Nous avons donc choisi de transformer le fichier généré par *ANVIL* pour que les temps donnés dans les pistes secondaires soient des temps absolus et non des références aux pistes primaires.

5.2.3 Détail des requêtes

Plusieurs requêtes s'appliquant au document *ANVIL* ont été effectuées, chaque requête faisant l'objet d'une feuille de style.

- La première requête a permis par exemple de récupérer les temps absolus des évènements des pistes secondaires.
- La seconde requête a permis de compter les différents types de phatiques.
- La troisième requête a listé tous les phatiques suivis d'un backchannel.
- La quatrième requête a répertorié tous les contours intonatifs apparaissant avant un backchannel dans un délai de 400 ms.

Comme nous l'avons mentionné précédemment, les temps donnés aux éléments des pistes secondaires dans le fichier de sortie généré par *ANVIL*, ne sont pas des temps mais des références aux éléments de la piste primaire associée. La première requête nous a donc permis de modifier le résultat d'*ANVIL* pour obtenir un document ne comportant que des temps absolus. Le principe de cette requête était donc de récupérer le nom de la piste primaire associée à chaque piste secondaire, puis d'aller chercher les temps de début et de fin des éléments de la piste primaire sur lesquels étaient alignés ceux des pistes secondaires.

Cette requête a plus été une préparation aux requêtes futures qu'une analyse des évènements présents dans la séquence vidéo, le résultat de cette requête ayant pour seul but de simplifier les prochaines requêtes portant sur le temps des évènements.

Une fois cette étape passée, il était intéressant de connaître le nombre et les types de phatiques contenus dans notre séquence vidéo. La seconde requête a donc consisté d'une part à compter le nombre total de phatiques et d'autre part le nombre de phatiques d'un type donné.

Sachant cela, nous nous sommes intéressés à l'interaction entre les phatiques et les backchannels. Le résultat de la troisième requête représente la liste des couples (phatique, backchannel) pour les phatiques suivis d'un backchannel dans un délai de 400 millisecondes. Ce laps de temps doit être compté à partir du début du phatique. La représentation de ce résultat n'est autre qu'un tableau qui contient sur chaque ligne les temps de début et de fin du phatique et du backchannel qui le suit, ainsi que le type du phatique et la fonction du backchannel.

Dans le même type de requête, nous avons voulu établir la liste des couples (contours intonatifs, backchannels) sous la condition que le backchannel devait commencer encore moins de 400 millisecondes après le début du contour intonatif.

5.3 Exploitation statistique des requêtes et premiers éléments de discussion

5.3.1 Résultats

Nous avons testé plus précisément les entités suivantes:

- Unités intonatives : IP, AP, EP
- Unités conversationnelles : TCU_f, TCU_nf, Cont
- Contours intonatifs : F ; REM ; ER ; RF1 ; RF2 ; RMC ; RQ ; RT ; fl ; m0 ; mr

Pour chaque entité, nous avons testé l'effet du type de l'entité sur la présence d'un BC dans les 400 ms suivantes, par un test de proportion. Par exemple, si la proportion des contours intonatifs de type *RMC* suivis d'un BC est supérieure à la proportion de ces mêmes contours dans tout le dialogue, alors le contour en question augmente la probabilité d'apparition d'un BC.
Nous avons utilisé un test unilatéral (u) lorsque nous avions une hypothèse a priori (proportion attendue plus grande ou plus petite), et un test bilatéral (b) dans le cas contraire.

Les prédictions étaient les suivantes :

- Les IP privilégient l'apparition d'un BC.
- Les TCU_f privilégient l'apparition d'un BC.
- Les contours intonatifs terminaux, à savoir RF1, RF2 et RT attirent préférentiellement les BC
- Les contours mineurs (m0, mr) ne privilégient pas l'apparition des BC

Le tableau suivant (12) présente les résultats :

	Type	**% global**	**% suivi de BC**	**Chi2, p**
Unités intonatives	ap	38.6	24.8	16 < 0.001 (u)
	ep	5.5	6.9	0.54 NS (b)
	ip	55.9	68.3	12.3 < 0.001 (u)
Unités conversationnelles	cont	13.1	5.5	5.6 0.0089 (u)
	f	57.7	74.5	12.6 < 0.001 (u)
	nf	29.2	20	4.4 0.018 (u)
Contours intonatifs	R	21.3	27.6	3.2 0.036 (u)
	RMC	15.9	18.3	0.53 NS (b)
	RT	3.6	6.5	2.83 0.09 (b)
	fl	30.8	31.9	0.04 NS (b)
	m0	13.7	9.7	1.9 0.08 (u)
	mr	25	18.4	3.4 0.03 (u)

Tableau 12 : Résultats concernant la probabilité d'apparition des BC après les unités intonatives, conversationnelles et les contours intonatifs. La dernière colonne indique la valeur du chi2, la pvalue, et le test effectué (u=unilateral, b=bilateral)

Les résultats confirment les hypothèses concernant les unités intonatives et les unités conversationnelles. Les BC apparaissent de manière significative majoritairement après une unité intonative (IP) au détriment des unités accentuelles (AP). La catégorie EP quant à elle ne varie pas significativement.

Les résultats relatifs aux unités conversationnelles confirment également les prédictions. Les BC apparaissent de manière significative après les TCU_f au détriment des TCU_nf et des Cont.

Les résultats concernant les contours intonatifs indiquent des taux de significativité plus marginaux mais nous pouvons parler de tendances possibles. La catégorie R, regroupement dans une seule et même catégorie de tous les contours montants majeurs, privilégie l'apparition d'un BC. Pris catégorie par catégorie en revanche, RMC (montant continuatif) et RT (montant terminal) ne semblent pas avoir d'influence sur la présence d'un BC. m0 (continuation mineure) et fl (plat) n'influent pas non plus sur la présence d'un BC. Quant à mr (mineur montant), il tend à être significativement moins suivi d'un BC.

Etant établi que les contours intonatifs jouent (ou non) un rôle dans l'apparition d'un BC, nous nous sommes interrogés sur une relation éventuelle entre certains de ces contours intonatifs et la modalité de réalisation des BC qui les suivent.

Dans la figure 2 suivante, nous présentons la répartition des différents BC en fonction de leur modalité vocale, gestuelle ou voco-gestuelle par contours intonatifs.

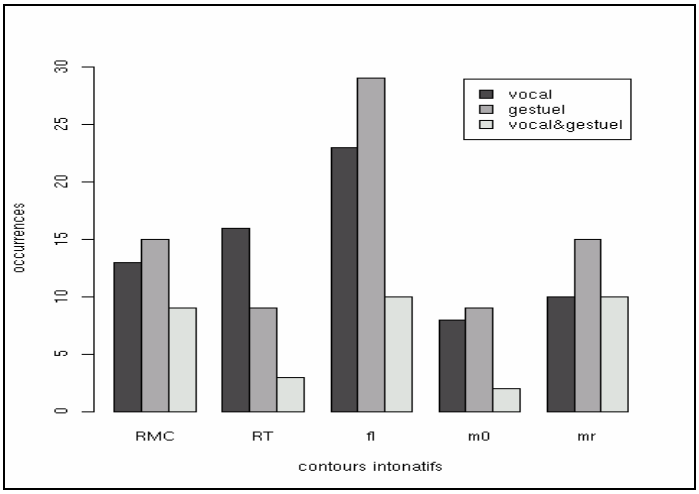

Figure 2 : Répartition des BC selon leur modalité vocale, gestuelle ou voco-gestuelle par type de contours intonatifs

Si la faiblesse des effectifs ne nous a pas permis de tester statistiquement les données, celles-ci présentent quelques tendances intéressantes. Nous avons regroupé l'ensemble des contours terminaux (RF1, RF2 et RT) dans une seule et même catégorie (RT) pour les comparer aux autres contours de continuation (RMC) et aux contours mineurs (mr, m0, fl). Nous avons distingué en outre les différents BC selon qu'ils étaient produits dans une modalité uniquement vocale, gestuelle ou voco-gestuelle. Nous remarquons que les contours RT favorisent fortement l'apparition de BC vocaux tandis que RMC et les autres contours favorisent davantage les BC gestuels.

5.3.2 Premiers éléments de discussion

Cette étude préliminaire présente quelques résultats intéressants qui méritent d'être développés et testés plus précisément. Nous confirmons les résultats issus de la littérature selon lesquels les BC ont tendance à apparaître après des unités potentiellement achevées puisqu'ils sont majoritairement présents après un tour final (TCU_f). Par ailleurs, nous confirmons également le rôle de la dimension prosodique puisque nous constatons que les unités intonatives (IP) attirent préférentiellement les BC. La question se pose de savoir maintenant quel est le poids de ce critère prosodique dans l'apparition des BC. Rappelons que les TCU sont construits à partir de trois critères, syntaxique, prosodique et pragmatique parmi lesquels il s'agira de déterminer le poids éventuel de chacun dans l'apparition des BC, et plus globalement dans le système de tours de parole.

Enfin, ces premiers résultats nous encouragent à poursuivre l'analyse des contours intonatifs, en termes formels et fonctionnels notamment, pour rendre compte plus finement de l'apparition des BC, étant établi que les contours montants notamment favorisent l'apparition d'un BC à l'opposé des contours dits mineurs par exemple.

Plus remarquable, ces résultats permettent de conclure à l'importance cruciale qu'il y a à analyser le discours dans ses aspects multimodaux tout en conservant un niveau d'analyse extrêmement fin au sein de chacun des domaines considérés. En effet, les contours terminaux par exemple affichent une tendance à privilégier plutôt l'apparition des BC à modalité vocale. Si l'on considère les contours de continuation (RMC) qui pour leur part, ne semblent pas représenter le contour qui attire le plus de BC, on note en revanche que les BC qui les suivent sont davantage produits dans une modalité gestuelle. Cette tendance, qui mériterait d'être validée statistiquement sur l'ensemble du corpus, pourrait d'ores et déjà être interprétée comme suit : après un contour de continuation, l'interlocuteur marque qu'il a bien pris en compte le désir du locuteur de poursuivre son tour, en utilisant un BC gestuel susceptible de moins interférer avec le discours produit (Cosnier, 1988). Ceci irait dans le sens de Kern (2007) qui signale, pour l'allemand, que le contour de continuation marque une attente du locuteur qui crée un lieu potentiel dans son discours, pour une réponse minimale (type BC) de l'interlocuteur. La figure 3 ci-après illustre précisément le cas d'un backchannel gestuel (hochement de tête -nod-) qui ponctue une séquence interactionnelle spécifique (dont nous reproduisons les derniers termes seulement): le locuteur s'engage dans un récit qu'il a beaucoup de mal à initier (nombreuses interruptions et reprises, accompagnées de gestes métaphoriques qui s'enchaînent directement). Durant cette séquence, l'interlocuteur le regarde sans intervenir, que ce soit gestuellement ou verbalement :

```
YM :et des fois ça m’arrivait //
    quand //
    en fait c’est bon //
    quand j’allais à l’école
```

En posant « en fait c’est bon », le locuteur semble enfin avoir réussi à formuler mentalement le début de son anecdote qui commence réellement avec « quand j’allais à l’école ». Ce TCU non-final, achevé dans un contour montant de continuation (RMC) (cf. courbe de fréquence fondamentale exportée dans ANVIL et les cibles tonales L (Low) et H (High) du codage *INTSINT*) auquel s’ajoute le changement d’orientation du regard du locuteur vers son interlocuteur (rôle phatique) semblent se conjuguer pour susciter le BC gestuel (nod) de l’interlocuteur.

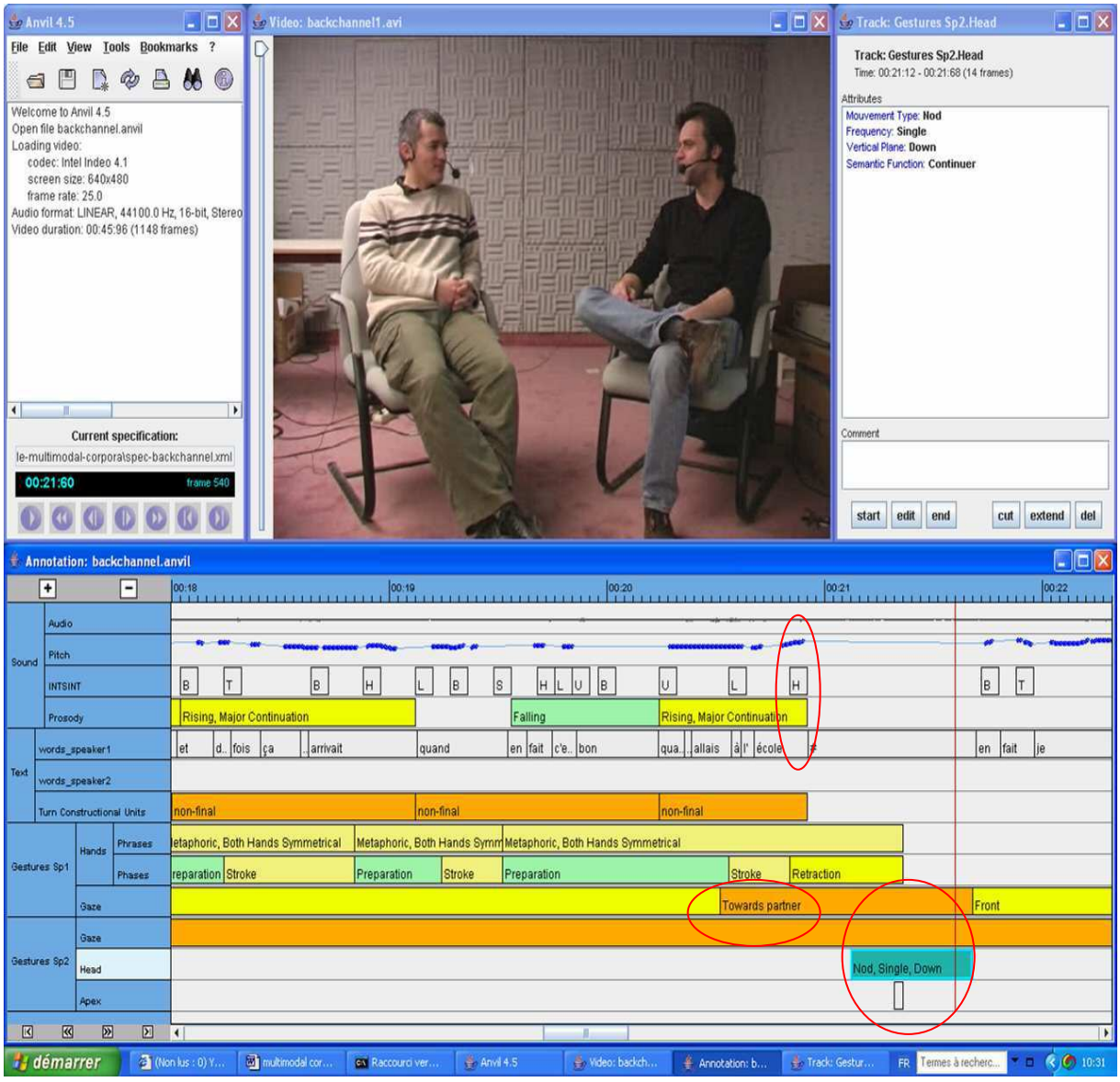

Figure 3. Illustration d’un cas de contour de continuation (RMC) suivi d’un BC gestuel.

CONCLUSION

Nous avons présenté dans ce travail une étude exploratoire sur les signaux backchannel en vue d'améliorer les typologies formelles et fonctionnelles existantes tout en cherchant plus globalement à montrer le rôle des facteurs prosodique, discursif et conversationnel dans leur apparition.

Cette étude a été plus largement prétexte à décrire les principes généraux d'annotation et d'exploitation du CID (*Corpus of Interactional Data*). Ce corpus constitue une ressource particulièrement importante en France étant entendu qu'il existe très peu de corpus similaire permettant une exploitation multimodale des phénomènes linguistiques. Un tel objectif nécessite non seulement la constitution des annotations elles-mêmes, dont certaines ont été présentées ici, mais également le développement d'outils adaptés permettant notamment leur mise en relation. Nous avons proposé par ailleurs des solutions en ce sens.

Concernant la question cruciale de la représentation et de l'édition de nos données, nous avons adopté le logiciel *ANVIL* qui, entre autres fonctionnalités, permet non seulement d'importer des annotations issues d'autres logiciels, mais utilise un format d'entrée et de sortie XML. Actuellement, ce format est devenu le standard sur lequel il est possible d'interroger et d'exploiter les corpus. Nous avons en ce sens présenté quelques exemples de requêtes plus ou moins complexes, c'est-à-dire interrogeant simultanément plusieurs niveaux d'annotations.

Enfin, le but ultime de notre projet est de proposer une plateforme intégrée permettant d'exploiter et d'interroger de manière optimale des données multimodales, et à terme de les partager. La question de la multimodalité est aujourd'hui au centre des recherches en linguistique. Elle repose sur l'analyse et l'interprétation de corpus annotés. La question de l'accès à ce type de données est donc cruciale, de même que, plus généralement, l'accès à toute sorte de données linguistiques. Nous avons donc décidé de nous inscrire dans la démarche de mise à disposition des données, corpus et ressources initiée par le CNRS dans le cadre du *CRDO* (Centre de ressource pour les données orales, http://www.crdo.fr). Notre corpus, en même temps que d'autres ressources, sera ainsi mis à disposition de la communauté à travers le centre.

Références

Allwood J. & Ahlsen, E., 1999, Learning how to manage communication, with special reference to the acquisition of linguistic feedback, *Journal of Pragmatics*, 31, pp. 1353-1389.

Allwood J. & Cerrato L., 2003, A study of gestural feedback expressions, in P. Paggio, K. Jokinen, A. Jönsson (eds.), *First Nordic Symposium on Multimodal Communication*, Copenaghen, 23-24 September 2003, pp. 7-22.

Auer P., 1996, On the prosody and syntax of turn-continuations, in E. Couper-Kuhlen & M. Selting (Eds), *Prosody in Conversation*, Cambridge: Cambridge University Press, pp. 57-101.

Barkhuysen P, Krahmer A, Swerts M, 2006, The interplay between auditory and visual cues for end-of-utterance detection, *Communication and Cognition*, Tilburg University.

Bertrand R., Boyer J., Cavé C., Guaïtella I. et Santi S., 1995, Relationship between gestures and voices in verbal interaction : prosodic and kinesic aspects of back-channel signals, in *Proceedings of XIIIth ICPhS 95* (1995 : Stockholm, Suède), pp. 746-749.

Bertrand R., 1999, De l'Hétérogénéite de la Parole. Analyse énonciative de phénomènes prosodiques et kinésiques dans l'interaction interindividuelle. *Thèse de doctorat de Sciences du Langage*, Université Aix-Marseille I.

Bertrand R. & Espesser R., 2003, Prosodic cues of back-channel signals in French conversational speech, *Prosody and Pragmatics International Congress* NWCL (6th : 2003 novembre 14-16, Preston, United Kingdom), communication orale.

Bertrand R., Blache P., Espesser R., Ferré G., Meunier C., Priego-Valverde B. et Rauzy, S., 2007, Le CID - Corpus of Interactional Data -: protocoles, conventions, annotations, *Travaux Interdisciplinaires du Laboratoire Parole et Langage d'Aix-en-Provence* (TIPA), vol. 25. 2007, p. 25-55.

Blache P. et Di Cristo A., 2002, Variabilité et dépendances des composants linguistiques, in *Proceedings of The Conference Traitement Automatique des Langues Naturelles* (TALN), pp. 205-214.

Boersma P. & Weenink D., 2005, Praat : doing phonetics by computer (version 4.3.14). Logiciel téléchargé le 26 mai 2005; http://www.praat.org/

Bouvet D., 2001, *La dimension corporelle de la parole. Les marques posturo-mimo-gestuelles de la parole, leurs aspects métonymiques et métaphoriques et leur rôle au cours d'un récit*, Paris: Peeters.

Calbris G., 2002, Sémantisme des connecteurs : nuancement du verbal par le gestuel, *Lidil* 26, pp. 139-153.
Caspers J., 1998, Who's next? The melodic Marking of Questions versus Continuation in Dutch, *Language and Speech*, 41, pp. 375-398.
Caspers J., 2003, Local speech melody as a limiting factor in the turn-taking system in Dutch, *Journal of Phonetics* 31, pp. 251-276.
Cerrato L. & Skhiri M., 2003a, Analysis and Measurement of communicative gestures in human dialogues, *Proceedings of AVSP '03*, St Jorioz, France, pp. 251-256.
Cerrato L. & Skhiri M., 2003b, in P. Paggio, K. Jokinen & A. Jönsson (eds.), *Proceedings of the First Nordic Symposium on Multimodal Communication*, Copenhague, 23-24 Septembre 2003, pp. 43-52.
Cerrato, L.; D'Imperio M., 2003, Duration and tonal characteristics of short expressions in Italian, in*Proceedings of 15th International Congress of Phonetic Sciences* (15 : 2003 août 3-9 : Barcelone, Espagne), pp. 1213-1216.
Chen L., Harper M., Franklin A., et al., 2006, A Multimodal Analysis of Floor Control in Meetings, *3rd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms (MLMI)*, Washington DC.
Cosnier J., 1988, Grands tours et petits tours, In Cosnier, Gelas, Kerbrat-Orecchioni (eds), *Echanges sur la conversation*, Editions du CNRS, pp; 175-184.
Couper-Kuhlen & Selting M., 1996, *Prosody in Conversation*, Cambridge: Cambridge University Press.
Couper-Kuhlen E. & Ford C. E., 2004, *Sound Patterns in Interaction*. Cross-linguistic studies from conversation, Amsterdam: John Benjamins Publishing Company.
Di Cristo A. & Di Cristo P., 2001, Syntaix, une approche métrique-autosegmentale de la prosodie, *TAL*, 42(1), pp. 69–111
Di Cristo A., Auran C., Bertrand R.**,** Chanet C., Portes C. et Regnier A., 2004, Outils prosodiques et analyse du discours, in A.C. Simon, A. Auchlin et A. Grobet (eds), *Cahiers de Linguistique de Louvain 30/1-3,* Louvain-la-neuve: Peeters, vol. 28, pp. 27-84
D'Imperio M., Bertrand R., Di Cristo A. et Portes C., 2007, Investigating phrasing levels in French : Is there a difference between nuclear and prenuclear accents? *Linguistic Symposium on Romance Languages* (LSRL) (36 : 2006 mars 31-avril 2 : Rutgers University, New Brunswick), [à paraître].
Ford C. E. & Thompson S. A., 1996, Interactional Units in Conversation : syntactic, intonational and pragmatic resources for the management of turns, in E. Ochs, E. A. Schegloff & S. A. Thompson (eds), *Interaction and Grammar*, Cambridge: CUP, pp. 134-184.
Fox Tree J.E.,1999, Listening in on Monologues and Dialogues, Discourse Processes, 27, 1, pp. 35-53.
Fox Tree J. E., 2002, Interpreting pauses and ums at turn exchanges, *Discourse Processes*, 34(1), pp. 37-55.
Heldner M., & Edlund J., 2006, Prosodic cues for interaction control in spoken dialogue systems. In *Working Papers 52: Proceedings of Fonetik 2006***.** Lund, Sweden: Lund University, Centre for Languages & Literature, Dept. of Linguistics & Phonetics, p. 53-56.
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, pp. 51-87.
Kipp, M., 2003-2006. *ANVIL 4.0. Annotation of Video and Spoken Language.* http://www.dfki.de/~kipp/*ANVIL*
Kern F., 2007, Prosody as a resource in children's game explanations: some aspects of turn construction and recipiency, *Journal of Pragmatics* 39, pp. 111-133.
Jun S.-A. & Fougeron C., 2002, Realizations of accentual phrase in French intonation, *Probus* 14, pp. 147-172.
Koiso H., Horiuchi Y., Ichikawa A., and Den Y., 1998, An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogs, *Language and Speech*, Vol. 41, pp. 295–321.
Laforest M., 1992, Le back-channel en situation d'entrevue, in *Recherches Sociolinguistiques, 2*., Québec : Université Laval, CIRAL.
Marandin J.-M., 2004, Contours as constructions, *ICCG-3 The third international Conference on construction grammars*, 7-10 juillet 2004, Marseille, non paginé.
Maynard S., 1989, *Japanese Conversation: Self-Contextualization through Structure and Interactional Management*, Ablex, Norwood, NJ.
Morel M.-A. & Danon-Boileau L., 1998, *La grammaire de l'intonation. L'exemple du français*, Paris, Gap: Ophrys.

Portes C. et Bertrand R., 2006, Some cues about the interactional value of the ‘continuation’ contour in French, in Actes *Discours et Prosodie comme Interface Complexe* (IDP) [Cederom, 14 pages].
Portes C., Bertrand R., Espesser, R. 2007, Contribution to a grammar of intonation in French. Form and function of three rising patterns», *Nouveaux Cahiers de Linguistique Française*, n°28, 2007, p. 155-162.

Sacks H., Schegloff E. A. & Jefferson G., 1974, A simplest systematics for the organization of turn-taking for conversation, *Language*, Vol. 50, pp. 696–735.
Schegloff E.A., 1982, Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences, in D. Tannen (ed), *Analyzing discourse: Text and talk*, Washington, DC: Georgetown University Press, pp. 71-93.
Selting M., 1998, TCUs and TRPs: the construction of 'units' in conversational talk, *InLiSt (Interaction and Linguistic Structures)*, Vol. 4, pp. 1–48.
Stubbe M., 1998, Are you listening ? Cultural influences on the use of supportive verbal feedback in conversation, *Journal of Pragmatics* 29, pp. 257-289.
Van Rullen T., 2005, Vers une analyse syntaxique à granularité variable, *Thèse de Doctorat*, Université Aix-Marseille I, Décembre 2005.
Vorreiter S., 2003, Turn continuations: towards a cross-linguistic classification, *InLiSt [Interaction and Linguistic Structures]*, No. 39, URL: http://www.uni-potsdam.de/u/inlist/issues/39/index.htm
Ward N., 1996, Using Prosodic Clues to Decide When to Produce Back-Channel Utterances, in *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)*, pp. 1724-1727.
Ward N. & Tsukahara W., 2000, Prosodic Features which Cue Back-channel Responses in English and Japanese, *Journal of Pragmatics*, 23, pp. 1177-1207.
Yngve V., 1970, On getting a word in edgewise, in *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society*, pp. 567-578.