

Catégoriser les réponses aux interruptions dans les débats politiques

Brigitte Bigi¹ Cristel Portes¹ Agnès Steuckardt¹ Marion Tellier¹
(1) Laboratoire Parole & Langage, CNRS & Aix-Marseille Universités
5, avenue Pasteur, BP 80975, 13604 Aix en Provence, France

brigitte.bigi@lpl-aix.fr, cristel.portes@lpl-aix.fr, Agnes.Steuckardt@univ-provence.fr, marion.tellier@lpl-aix.fr

Résumé. Cet article traite de l'analyse de débats politiques selon une orientation multimodale. Nous étudions plus particulièrement les réponses aux interruptions lors d'un débat à l'Assemblée nationale. Nous proposons de procéder à l'analyse via des annotations systématiques de différentes modalités. L'analyse argumentative nous a amenée à proposer une typologie de ces réponses. Celle-ci a été mise à l'épreuve d'une classification automatique. La difficulté dans la construction d'un tel système réside dans la nature même des données : multimodales, parfois manquantes et incertaines.

Abstract. This work was conducted to analyze political debates, with a multimodal point of view. Particularly, we focus on the answers produced by a main speakers after he was disrupted. Our approach relies on the annotations of each modality and on their review. We propose a manual categorization of the observed disruptions. A categorization method was applied to validate the manual one. The difficulty is to deal with multimodality, missing values and uncertainty in the automatic classification system.

Mots-clés : corpus, annotations, multimodalité, classification supervisée.

Keywords: corpus, annotations, multimodality, classification.

1 Introduction

Quelque préparé qu'il ait pu être, le discours d'un député comporte, lorsqu'il est prononcé au sein de l'Assemblée nationale, une part d'improvisation. L'orateur y est exposé aux interruptions, rarement amènes, de ses collègues, prises de parole qui ne bénéficient pas d'une autorisation formelle du Président de séance. Le regard du député, alors, quitte les notes, l'intonation change, la parole spontanée prend le relai de la déclamation. Si l'analyse de discours s'est intéressée traditionnellement aux débats politiques retransmis par la télévision (Bonnafous & Tournier, 2001), en revanche les interactions à l'œuvre dans les assemblées délibératives (Assemblée nationale et Sénat), ont été moins étudiées, faute de matériau adéquat. Or l'Assemblée nationale met aujourd'hui à disposition sur internet la retransmission des vidéos des séances. Cette ressource nouvelle ouvre à la recherche l'opportunité de mieux comprendre les mécanismes de la délibération politique. Nous nous proposons d'aborder, selon une approche multimodale, les stratégies mises en œuvre pour contrer la répartition dans les débats politiques.

On retrouve dans de nombreux domaines les termes modalité et multimodalité. Cependant, leur usage et leur définition sont variables tant par le sens que par le formalisme de la définition. Par essence, cependant, la multimodalité est l'utilisation d'au moins deux des cinq sens pour l'échange d'informations. De cette définition, on peut directement déduire l'importance de la multimodalité dans la question du langage et de la parole. Les annotations multimodales de corpus ont soulevé de nombreuses questions. La création et l'annotation de corpus multimodaux présentent des difficultés liées aux choix de logiciel, à des décisions concernant la découpe et l'alignement parole/transcription, à la pertinence des méthodes automatiques, etc. Nous avons donc été confrontés aux grands problèmes posés par l'annotation de ce type de ressource. Cet article présente en premier lieu la méthodologie qui a été conduite pour l'annotation multimodale d'une partie de la séance publique du 4 mai 2010 à l'Assemblée nationale durant laquelle Yves Cochet intervient. Le corpus contient de nombreux niveaux d'annotations automatiques (éventuellement révisés manuellement), ou manuels. Cette approche s'appuie sur celle qui a été proposée pour le corpus CID - Corpus of Interactional Data (Blache *et al.*, 2010). À ces annotations, nous ajoutons l'indication des segments durant lesquels le locuteur répond (ou choisit de ne pas répondre) à une interpellation. En effet, c'est non pas aux interruptions elles-mêmes, peu audibles dans la vidéo et incomplètement enregistrées par

le compte rendu écrit, mais aux réponses à ces interruptions que l'on s'intéresse ici. L'analyse argumentative de ces données nous a amené à proposer une typologie de ces réponses aux interruptions. Trois grands types de stratégies ont été relevés : la décision d'ignorer l'interruption, la réponse sur la prédication avancée par celui qui interrompt, et la réplique, « pas de côté » méta-discursif ou méta-interactionnel.

Cette typologie est mise à l'épreuve d'une classification automatique. On fait l'hypothèse que si une structure de classe existe au sein des nos données et qu'elle correspond à la typologie proposée, le système de classification automatique peut la caractériser et effectuer une catégorisation correcte. L'élaboration d'un tel classifieur implique de traiter des données hétérogènes et asynchrones, du fait de leur nature multimodale. De plus, certaines annotations ne peuvent pas être réalisées (caméra qui se déplace ou bruit de fond) car ce sont des données réelles (par opposition à des enregistrements de laboratoire). De ce fait également, il subsiste un certain degré d'incertitude sur les annotations produites qui mérite d'être pris en considération par le classifieur.

2 Corpus et annotations

2.1 Le corpus

Nous avons choisi d'exploiter une ressource rendue disponible par le site de l'Assemblée nationale¹ : la vidéo des séances publiques. Le débat sur le « Grenelle II de l'environnement » a été sélectionné en raison de la controverse importante qu'il a déclenchée. La première intervention sur le projet de loi du gouvernement a été présentée le 4 mai 2010 par le député Vert Yves Cochet, déçu de la timidité de ce Grenelle II ; dans son intervention, de 45 minutes, on a retenu le moment le plus vif de la controverse, où le député, fréquemment interrompu, quitte ses notes, et où la parole spontanée prend le relais de la déclamation. Le passage qui a été extrait dure 4 minutes et commence après 19 minutes de débat. Le fichier vidéo peut être téléchargé après demande d'autorisation². La vidéo est au format FLV (Flash Video), de qualité suffisante, mais de petite résolution (320x256). La piste audio a été extraite à partir de cette dernière. Un compte rendu au format HTML est également disponible. Durant le passage sélectionné, Yves Cochet prononce près de 800 mots, 136 contours prosodiques ont été relevés ainsi que 122 gestes principaux. De plus, il est interrompu à onze reprises.

2.2 Les annotations

Dans un premier temps, le signal de parole a été découpé automatiquement en unités inter-pausales (Interpausal Unit - désormais IPU). Les IPU sont des blocs de parole bornés par des pauses silencieuses d'au moins 200 ms.

Transcription Orthographique Enrichie (TOE) : Nous avons adopté les conventions de transcriptions établies pour le corpus CID. Si elle est essentiellement orthographique, la TOE spécifie toutefois les phénomènes typiques de l'oral tels que les pauses pleines (euh, hum, etc.), les faux-départs, les amorces, les mots tronqués, les répétitions, etc. La TOE a été choisie afin de garantir à la fois un meilleur rendement de l'alignement phonétique nécessaire à l'analyse du module phonétique, et le meilleur rendement pour les modules morpho-syntaxique et syntaxique qui ne peuvent être faits que sur des transcriptions orthographiques standard.

Phonétisation : La phonétisation consiste à déterminer les phonèmes qui ont été produits lors d'un énoncé oral. Le convertisseur graphème-phonème fournit une suite de mots phonétisés en SAMPA - Speech Assessment Methods Phonetic Alphabet. Les phonèmes sont ensuite alignés automatiquement sur le signal. Une vérification manuelle des alignements a été effectuée. La syllabation a été réalisée à partir des séquences de phonèmes avec le système décrit dans (Bigi *et al.*, 2010). Les syllabes ont également été révisées manuellement.

Syntaxe : Les annotations syntaxiques et morpho-syntaxiques ont été réalisées automatiquement par le système décrit dans (Blache & Rauzy, 2008).

Auto-répétitions : Elles ont été réalisées manuellement. Trois catégories sont associées au répétable et au répété : DIS (énoncé disfluent) ; RHE (énoncé de type rhétorique) ; LIST (effet de liste).

¹<http://www.assembleenationale.fr/13/seance/vod/grenelle-2-20100504-2.asp>

²La demande peut être faite auprès de la Cellule GILDA - Gestion des Images en Ligne des Débats de l'Assemblée, de l'Assemblée nationale, via la Division de l'information multimédia (en charge de la mise en ligne des documents parlementaires) : dim@assemblee-nationale.fr.

Prosodie : Le terme de prosodie renvoie à l'étude linguistique de ce qui peut être conçu intuitivement comme le rythme et la mélodie de la parole. Les variations métriques (rythme) et intonatives (mélodie) de la parole remplissent d'abord une fonction d'organisation du flux de parole en associant plusieurs mots dans des « groupes de sens » (Di Cristo, 1999). Dans cette étude, nous nous sommes concentrés sur les contours intonatifs. Ces schémas mélodiques associés aux groupes de sens diffèrent selon les langues mais la plupart des langues possèdent un paradigme de contours contrastifs. Ceux-ci jouent un rôle pragmatique important, notamment en véhiculant l'attitude des interlocuteurs vis-à-vis des contenus transmis (Beyssade & Marandin, 2007). Leur importance argumentative est donc avérée. Pour le français, nous avons adapté l'inventaire proposé dans (Bertrand *et al.*, 2007). Cet inventaire non exhaustif comporte deux catégories de contour mineur (montant et descendant), généralement associés à des constituants de l'ordre du syntagme, et quatre catégories de contours majeurs (descendant, montant, montant-descendant, descendant depuis la pénultième), subordonnant les contours mineurs dans la structure prosodique, et correspondant plutôt au niveau de la phrase ou de l'énoncé. Au sein des contours majeurs montants, notre inventaire distingue en outre trois variétés fonctionnelles : montant continuatif, montant de liste, montant final. Leur annotation repose sur une identification auditive réalisée par des experts.

Gestes : Les gestes co-verbaux (ici les gestes des mains qui accompagnent spontanément la parole) ont été annotés manuellement. Pour classer les co-verbaux, nous avons utilisé la typologie de McNeill (McNeill, 1992, 2005) qui définit quatre dimensions du geste : les gestes iconiques (qui illustrent des concepts concrets), les gestes métaphoriques (qui illustrent des concepts abstraits), les déictiques (qui pointent vers un référent présent ou absent) et les battements (qui rythment le discours, sans valeur sémantique). À cette typologie, la plus fréquemment utilisée dans les études de la gestuelle aujourd'hui, nous avons ajouté deux types de gestes : les gestes interactifs (qui ont une fonction de régulation de l'interaction selon (Bavelas *et al.*, 1995) et les emblèmes (qui appartiennent à un répertoire fixe et propre à une culture, un peu comme des expressions idiomatiques). Enfin, nous avons également annoté les gestes avortés, c'est-à-dire esquissés mais non achevés. Certains gestes pouvant parfois présenter deux dimensions nous avons ajouté une annotation « type de geste secondaire ». De plus, pour l'analyse de ce corpus particulier, nous avons choisi d'annoter la manualité, c'est-à-dire la main utilisée pour former le geste selon 4 catégories : main droite, main gauche, deux mains asymétriques, deux mains symétriques.

Réponses aux interruptions : Leur repérage a consisté à mettre en correspondance ce que nous observons sur la vidéo et l'audio, la transcription et le compte rendu, dans le but de déterminer avec précision les segments de la transcription qui correspondent aux moments où Y. Cochet répond (ou choisi de ne pas répondre) aux interruptions.

3 Typologie des réponses aux interruptions

Dans le passage analysé, Yves Cochet soutient la thèse que les voitures électriques, loin de constituer une panacée écologique, ne font que déplacer le problème énergétique vers le nucléaire, dont il souhaite que la France sorte. Cette attaque des voitures électriques suscite des réactions hostiles de la part des députés et des ministres de la majorité. Interrompu, le député vert met en œuvre plusieurs stratégies pour retrouver sa position d'orateur et reprendre le fil de son argumentation.

La stratégie du mépris : Elle est utilisée deux fois dans le corpus. Sur la vidéo, la décision d'ignorer peut être perçue par des indices convergents : une courte pause, un mouvement de tête vers le côté droit d'où provient l'interruption, suivi d'un recentrage, une mimique de mépris (fermeture prolongée des yeux, pincement des lèvres). Ces interruptions sont enregistrées par le compte rendu écrit : avec l'interruption "Passons au vélo !", le député Maxime Bono caricature la pensée d'Yves Cochet en poussant ses conséquences à des extrémités susceptibles de faire rire l'hémicycle ; la seconde interruption est enregistrée par la didascalie "Rires et exclamations sur les bancs de l'UMP".

La réponse frontale : La deuxième stratégie oppose à l'interruption une réponse sur la prédication. Deux sous-catégories apparaissent dans le corpus. Dans le premier cas, Yves Cochet se contente d'une négation de la prédication de l'adversaire. Avec ce degré zéro de la réponse contradictoire, aucune réfutation n'est engagée, soit parce que l'orateur est arrêté par une nouvelle interruption, soit parce que le député se contente d'asserter une prédication opposée à la précédente. La seconde sous-catégorie de réponse sur la prédication se décompose en deux moments. Dans un premier temps, l'orateur reprend, en dialogisme interlocutif immédiat l'énoncé de son contradicteur. Cette reprise s'opère avec une modalité prosodique spécifique : l'intonation d'arrière-plan. À cette hétéro-répétition, le locuteur oppose ensuite un contre-argument.

Le pas de côté : La troisième stratégie consiste à répliquer à l'interruption par un commentaire méta-discursif ou méta-interactionnel, au lieu de répondre sur le fond. Les études interactionnelles appellent ce type de réplique, « intervention réactive portant sur l'énonciation et non sur l'énoncé de l'intervention précédente » (Traverso, 2002). Les répliques méta-discursives d'Yves Cochet évaluent la qualité du propos de l'adversaire : l'évaluation positive sert à retourner l'énoncé de l'adversaire contre lui, sur le mode d'un vous ne croyez pas si bien dire. Les répliques méta-interactionnelles s'adressent non à l'hémicycle mais à la seule Présidente. Lassé d'être continuellement interrompu, Yves Cochet tente de renégocier le cadre de l'interaction et demande l'appui de la Présidente par deux fois. Ces répliques-là sont prononcées avec un niveau de registre tonal bas et une orientation du corps en direction du « perchoir ». La renégociation du cadre interactionnel fait partie des techniques de débat utilisées par les politiques : la scénette jouée en aparté avec la Présidente permet à l'orateur de se poser en victime d'adversaires irrespectueux des règles de l'Assemblée nationale.

À l'examen de détail, la typologie argumentative paraît pouvoir être mise en correspondance avec la prosodie et le geste, sans toutefois que les descriptions prosodiques et gestuelles dégagent des caractères absolument spécifiques pour chaque type argumentatif ; il faudrait, pour parvenir à une convergence des trois approches, raffiner la typologie argumentative, au risque de la rendre peu lisible. Malgré cette relative approximation, on se propose de mettre la classification argumentative à l'épreuve du traitement automatique.

4 Classification d'annotations multimodales

On procède à une classification automatique des données en vue d'une objectivation de l'analyse manuelle, en confrontant les interprétations de l'annotateur avec une classification automatique des réponses du locuteur aux interruptions. La classification automatique ainsi établie vise à valider/vérifier la classification manuelle des réponses aux interruptions en ne faisant pas intervenir la subjectivité de l'expérimentateur par autre chose que le choix des représentations qui sont utilisées. Nous rappelons que la classification automatique supervisée consiste à caractériser des groupes d'objets ayant un comportement homogène. On doit y définir un ensemble T d'exemples caractérisés par des descripteurs, et y associer la classe durant la phase d'apprentissage. La construction d'un classifieur dans l'ensemble C des catégories consiste en la définition d'une fonction qui, pour chaque exemple $s_t, t \in T$, renvoie une catégorie $c_i \in C$ à laquelle il est assigné. Les résultats présentés dans (Lavesson & Davidsson, 2006) indiquent que les paramètres (et leur optimisation) sont souvent plus importants que le choix de l'algorithme de classification. Nous supposons que cet assertion est également pertinente en cas de données multimodales. Nous avons choisi le classifieur basé sur l'algorithme C4.5, reposant sur des arbres de décisions, proposé dans (Quinlan, 1993) et implémenté dans le logiciel Weka (Holmes *et al.*, 1994). Les expériences sont menées selon le principe de la validation croisée, avec un processus basé sur 100 itérations.

En cas de corpus multimodal, la difficulté vient de l'hétérogénéité, et de l'absence d'alignement entre les données (ici, des annotations). La première étape consiste donc à choisir comment opérer la fusion des descripteurs afin d'obtenir un vecteur de caractéristiques multimodal. La littérature - notamment (Snoek *et al.*, 2005), est essentiellement fondée sur l'opposition entre deux stratégies de fusion des données multimodales : la fusion de décisions (ou fusion tardive), la fusion de descripteurs (autrement nommée fusion précoce). La fusion de décisions consiste à créer autant de systèmes de classification que de modalités et à fusionner les décisions prises séparément pour chacune de ces modalités. Dans cette technique le caractère multimodal de l'information n'est que peu pris en compte. L'objectif de la fusion de descripteurs est d'obtenir, à partir de vecteurs de descripteurs extraits de chacune des modalités impliquées, un vecteur de caractéristiques multimodal. La fusion précoce implique d'effectuer le choix d'une méthode pour fusionner les descripteurs et obtenir ainsi un vecteur multimodal. Le plus simple, qui est assez largement utilisé, consiste à simplement concaténer les vecteurs unimodaux (La Cascia *et al.*, 1998). D'autre part, compte tenu de l'aspect asynchrone des données, à la manière de (Zhi *et al.*, 2001), nous avons choisi de suréchantillonner les annotations multimodales en choisissant de générer un vecteur de descripteurs pour chaque phonème car le phonème est la plus petite de nos unités annotées. Notre corpus est alors constitué de 2488 exemples. À chaque exemple, on associe 0 ou 1 annotation par modalité, dans la liste suivante :

1. Contour Prosodique : aucun, F, RMC, RF1, mc, RT, RL, RF2, L
2. Type de geste principal : aucun, beat, metaphoric, deictic, interactive, beat_beat, iconic, emblem, aborted
3. Type de geste secondaire : aucun, beat, interactive, emblem, deictic, metaphoric, iconic
4. Manualité : aucune, right_hand, both_hands_symmetric, left_hand, both_hands_non_symmetric

5. Répétition : aucune, RHEreptable, RHEreptition, DISreptable, DISreptition, LISTreptable, LISTreptition
6. Catégorie morpho-syntaxique : aucune, adverb, determiner, noun, adjective, preposition, pronoun, verb, conjunction, interjection, auxiliary
7. Groupe syntaxique : aucun, GR, GN, GA, GP, NV, PV
8. Classe : Discours, Prédication, MetaInteractionnel, Ignorée, MetaDiscursif

Soit un exemple s_t , et \mathbf{s}_t son vecteur de descripteurs multimodaux tel que : $\mathbf{s}_t = (s_t^1, s_t^2, \dots, s_t^m, \dots, s_t^M)$. Chaque modalité m de l'exemple s_t^m est représentée par un vecteur d'annotations \mathbf{s}_t^m tel que : $\mathbf{s}_t^m = (a_t^{m,1}, a_t^{m,2}, \dots, a_t^{m,n}, \dots, a_t^{m,N^m})$ où $a_t^{m,n} = 1$ si l'annotation est affectée à cet exemple dans cette modalité et $a_t^{m,n} = 0$ sinon. Le taux de bonnes classifications de ce classifieur est de 90,63 %.

Ce premier système, s'il obtient déjà des résultats encourageants, souffre de deux problèmes, liés au fait que nous traitons des données réelles. Le premier point concerne le fait qu'il ne considère pas différemment l'absence d'annotation car aucun événement ne se produit, et l'absence d'annotation car aucun événement ne peut être observé (cas des données manquantes). (Saar-Tsechansky & Provost, 2007) comparent de nombreuses solutions de méthodes d'emputations pour l'algorithme C4.5 ; cependant, celle proposée dans Weka ne permet qu'une faible amélioration des performances (91,08 %). Pourtant, dans nos données, 117 exemples ne contiennent pas d'annotation des contours prosodiques, car le bruit de fond ne permet pas cette annotation. De même, 74 exemples n'ont aucune des trois modalités gestuelles car la caméra n'est pas dirigée vers l'orateur mais vers l'Assemblée. Pour traiter ces données, nous proposons simplement d'affecter un poids égal à chacune des annotations de la modalité manquante : $a_t^{m,n} = \frac{1}{|N^m|}$. Le score de bonnes classifications en est nettement amélioré, passant à 94,33 %.

Le second point concerne la notion d'incertitude, qui est un problème très récemment abordé en classification. Nous citerons les travaux présentés dans (Chau *et al.*, 2005) avec un classifieur à base des k plus proches voisins, ceux de (Ren *et al.*, 2009) avec des « Naives Bayes », (Qin *et al.*, 2010) avec un système à base de règles, ou (Liang *et al.*, 2010) avec des arbres de décision. Néanmoins, dans (Chau *et al.*, 2005; Liang *et al.*, 2010), l'incertitude est simulée. De plus, tous ces travaux concernent une seule modalité. Nous proposons d'introduire un degré d'incertitude au niveau des paramètres de chaque modalité. On considère ainsi que chaque annotation peut exister, même si elle n'est pas mentionnée par l'annotateur (expert humain ou système automatique). On affecte une valeur ϵ à chaque annotation non-observée. Dans un système multimodal, il va de soi que cette valeur doit être spécifiée différemment selon la modalité. En l'absence d'un corpus de développement qui permettrait d'estimer les valeurs optimales, nous avons attribué une valeur empirique ϵ_m . Celle-ci a été fixée à $\epsilon_m = 0,001$ pour les modalités annotées manuellement, et $\epsilon_m = 0,015$ pour les modalités annotées automatiquement. Puisqu'on affecte une valeur aux annotations non observées, les valeurs des annotations observées en sont d'autant réduites. Ainsi $a_t^{m,n} = \epsilon_m$ en l'absence d'annotation, sinon $a_t^{m,n} = a_t^{m,n} - \lambda_m \epsilon_m$ où λ_m est le nombre des annotations non observées pour la modalité m de l'exemple s_t . Ce système permet une amélioration de la classification avec un score de 94,98 %, soit un gain relatif de 11,46 % par rapport au système précédent.

5 Conclusion

Cet article se situe dans le cadre de l'étude multimodale sur corpus des stratégies mises en œuvre pour contrer la répartie dans les débats politiques. Nous avons sélectionné et richement annoté 4 minutes de l'intervention d'Y. Cochet à l'Assemblée nationale le 4 mai 2010, dans un débat sur « Le Grenelle II de l'environnement ». L'analyse argumentative, mise en correspondance avec l'analyse prosodique et gestuelle nous ont amenées à proposer une typologie des réponses aux interruptions. Nous avons validé cette hypothèse par une classification automatique basée sur les annotations multimodales. Un paramétrage particulier a été proposé pour prendre en considération la spécificité de ces données : multimodales, parfois manquantes et incertaines. Au terme de cette exploration méthodologique, apparaît clairement la possibilité d'asseoir la compréhension des fonctionnements argumentatifs à l'œuvre dans le débat politique sur une analyse multimodale systématique. Bien que les annotations manuelles soient fastidieuses, nous nous efforçons de reproduire cette approche afin de valider nos propositions sur un corpus plus important. Le corpus, les annotations réalisées et une description détaillée sont d'ores et déjà diffusés librement sur le site internet du Centre de Ressources pour la Description de l'Oral d'Aix-en-Provence³.

³Fiche 000729 à l'adresse : <http://www.crdo.fr/>

Références

- BAVELAS J.-B., CHOVIL N., COATES L. & ROE L. (1995). Gestures specialized for dialogue. *Personality and Social Psychology Bulletin*, **21**, 394–405.
- BERTRAND R., PORTES C. & SABIO F. (2007). Distribution syntaxique, discursive et interactionnelle des contours intonatifs du français dans un corpus de conversation. *Travaux neuchâtelois de linguistique*, **47**, 59–77.
- BEYSSADE C. & MARANDIN J.-M. (2007). French intonation and attitude attribution. In *Texas Linguistics Society Conference : Issues at the Semantics-Pragmatics Interface*, Somerville, MA : Denis et al. (eds.), Cascadilla Press.
- BIGI B., MEUNIER C., NESTERENKO I. & BERTRAND R. (2010). Automatic detection of syllable boundaries in spontaneous speech. In *Language Resource and Evaluation Conference*, La Valetta, Malte.
- BLACHE P., BERTRAND R., BIGI B., BRUNO E., CELA E., ESPESSER R., FERRÉ G., GUARDIOLA M., HIRST D., MAGRO E.-P., MARTIN J.-C., MEUNIER C., MOREL M.-A., MURISASCO E., NESTERENKO I., NOCERA P., PALLAUD B., PRÉVOT L., PRIEGO-VALVERDE B., SEINTURIER J., TAN N., TELLIER M. & RAUZY S. (2010). Multimodal annotation of conversational data. In *The Fourth Linguistic Annotation Workshop*, Uppsala, Suède.
- BLACHE P. & RAUZY S. (2008). Influence de la qualité de l'étiquetage sur le chunking : une corrélation dépendant de la taille des chunks. In *Actes de TALN*, p. 290–299, Avignon.
- BONNAFOUS S. & TOURNIER M. (2001). Discours et gestes télévisés : quelles méthodes ? *Mots. Les langages du politique*, **67**, 110–128.
- CHAU M., CHENG R. & KAO B. (2005). Uncertain data mining : A new research direction. In *Workshop on the Sciences of the Artificial*, Hualien, Taiwan.
- DI CRISTO A. (1999). Le cadre accentuel du français contemporain. *Langues*, 3(2), 184–205, **Langues**, **4(2)**, 258–267.
- HOLMES G., DONKIN A. & WITTEN I.-H. (1994). Weka : a machine learning workbench. In *Second Australian and New Zealand Conference on Intelligent Information Systems*, p. 357–361 : Intelligent Information Systems.
- LA CASCIA M., SETHI S. & SCLAROFF S. (1998). Combining textual and visual cues for content-based image retrieval on the world wide web. In *IEEE Workshop on Content - Based Access of Image and Video Libraries*, Washington, DC, USA.
- LAVESSON N. & DAVIDSSON P. (2006). Quantifying the impact of learning algorithm parameter tuning. In *The Twenty-First National Conference on Artificial Intelligence*.
- LIANG C., ZHANG Y. & SONG Q. (2010). Decision tree for dynamic and uncertain data streams. In *2nd Asian Conference on Machine Learning*, volume 3, p. 209–224, Tokyo, Japon.
- MCNEILL D. (1992). *Hand and Mind : What gestures reveal about thought*. Chicago : The University of Chicago Press.
- MCNEILL D. (2005). *Gesture & thought*. Chicago : The University of Chicago Press.
- QIN B., XIA Y., SATHYESH R., PRABHAKAR S. & TU Y. (2010). urule : A rule-based classification system for uncertain data. In *10th IEEE International Conference on Data Mining Workshops*, p. 1415–1418, Sydney, Australia.
- QUINLAN J.-R. (1993). *C4.5 : Programs for Machine Learning*. Morgan Kaufman ed.
- REN J., LEE S.-D., CHEN X., KAO B., CHENG R. & CHEUNG D.-W.-L. (2009). Naive bayes classification of uncertain data. In *Ninth IEEE International Conference on Data Mining*, p. 944–949.
- SAAR-TSECHANSKY M. & PROVOST F. (2007). Handling missing values when applying classification models. *Journal of Machine Learning Research*, **8**, 1625–1657.
- SNOEK C.-G.-M., WORRING M. & SMEULDERS A.-W.-M. (2005). Early versus late fusion in semantic video analysis. In *13th annual ACM international conference on Multimedia*, p. 399–402, New York, USA.
- TRAVERSO V. (2002). *Réplique*, p. 502. Dictionnaire d'analyse de discours. Patrick Charaudeau et Dominique Maingueneau (éds), Paris : Seuil.
- ZHI Q., KAYNAK M. N., SENGUPTA K., CHEOK A.-D. & KO C.-C. (2001). HMM modeling for audio-visual speech recognition. In *IEEE International Conference on Multimedia and Expo (ICME'01)*, p. 136–139, Los Alamitos, CA, USA : IEEE Computer Society.