Penser tout haut. Analyse multimodale de fins de séquences conversationnelles

Gaëlle Ferré

Penser tout haut. Analyse multimodale de fins de séquences conversationnelles

Gaëlle Ferré

Université de Nantes - LLING
Chemin de la Censive du Tertre, BP 81227
44312 Nantes Cedex 3
Gaelle.Ferre@univ-nantes.fr

ABSTRACT

This paper proposes a multimodal analysis of thoughts spoken out loud by which speakers show some degree of inattention to the current conversation. The sequences under study have been detected in the video files of conversational *CID*, recorded at the LPL, mainly thanks to the unfocused and fixed gaze of speakers. An analysis showed that all the sequences under study share prosodic properties (in terms of F0 and intensity range and span ; presence of pause before and after the sequences). Gesturally speaking the speakers' fixedness of gaze is paralleled by a completely relaxed attitude of the body and absence of any hand gesture.

A discourse analysis of the utterances shows that they are all post-closing sequences (rather than closing sequences proper) and that they appear before a topic change.

Keywords: multimodality, gesture, prosody, conversational topics, closing sequences.

1. INTRODUCTION

Si l'inattention et le désengagement conversationnel des participants à une interaction a fait l'objet de très peu d'études jusqu'à présent, c'est sans doute parce qu'ils semblent difficiles à cerner. Ainsi, on trouve dans la littérature des positions telles que celle de Mueller et Dyer [14] qui affirment que le « daydream (...) is not manifested externally ». Pourtant, dans certains manuels d'ophtalmologie [1], il est précisé que l'inattention, la non-coopération ou la fatigue d'un patient peut fausser un diagnostique d'anormalité du champ visuel. C'est donc que l'inattention se traduit par des caractéristiques précises au niveau des yeux et du regard différentes de celles de l'attention. C'est d'ailleurs sur ces caractéristiques physiques du regard que se basent les nombreuses études appliquées à la sécurité routière qui se proposent de détecter automatiquement les phases d'inattention du conducteur au volant afin de les prévenir ([13] et [16] pour n'en citer que deux). Si l'inattention concerne également les participants à une interaction (corpus décrit en 2.1.), les paramètres du regard utilisés pour la repérer dans la conversation spontanée doivent cependant être adaptés par rapport à ceux utilisés pour les études de sécurité routière, car le regard joue également un rôle dans la gestion des tours de parole en conversation, et c'est donc ce que je présente en 2.2., le postulat initial étant que l'inattention des participants peut être repérée sur les enregistrements vidéo à partir du regard des locuteurs. Seules les séquences de « pensée à voix haute » ont été retenues dans cette étude et s'avèrent posséder des caractéristiques communes sur les plans discursif, prosodique et gestuel permettant de les analyser comme des « post-closing sequences » par opposition au « pre-closing sequences » proposées par Schegloff et Sacks [15] dans le domaine de l'analyse conversationnelle. Ces termes sont expliqués dans la discussion (section 4).

2. CORPUS ET ANNOTATION

2.1. Le corpus CID

Le corpus qui a été utilisé pour ce travail est le *Corpus of Interactional Data*, enregistré au Laboratoire Parole et Langage d'Aix en Provence[1] et qui a été décrit dans le détail dans Bertrand et al. [3]. Les enregistrements vidéo ont été réalisés par R. Bertrand et B. Priego-Valverde, et consistent en plusieurs heures de dialogues entre pairs (une heure de conversation par enregistrement). Les participants ont été filmés dans une chambre sourde. Ils se connaissaient bien et étaient familiers des lieux et des techniques d'enregistrement, ceci afin de garantir un plus grand naturel dans les interactions. Ils étaient assis l'un à côté de l'autre avec un plan de trois-quart face), étaient filmés par une seule caméra et enregistrés par deux micros serre-tête sur deux pistes son séparées pour une meilleure qualité du signal. Le corpus a été transcrit dans sa globalité et de nombreuses annotations ont été réalisées dans plusieurs champs linguistiques, notamment la phonétique / phonologie, la morphologie, la syntaxe, la prosodie. L'an passé, l'annotation des mouvements et des gestes

[1] Je remercie tout particulièrement P. Blache et R. Bertrand pour m'avoir autorisée à travailler sur ce corpus.

réalisés par les participants a même commencé. Cette annotation doit se faire de manière entièrement manuelle, aussi est-elle particulièrement longue et encore inachevée. Malgré cela, de nombreuses études ont déjà pu être réalisées sur le corpus, qui ne portaient cependant pas sur les séquences analysées ici.

2.2. Repérage des séquences d'inattention

Dans le corpus CID, j'ai sélectionné trois heures de dialogue entre 6 locuteurs (4 femmes et 2 hommes) puis ai visionné les fichiers dans le logiciel d'annotation de la vidéo ELAN [7] afin de repérer les séquences d'inattention des participants, repérage basé essentiellement sur leur regard. Les études appliquées à la sécurité routière évaluent l'inattention du conducteur par la fixité de son regard (absence de *scanning*), par l'éloignement du regard par rapport à un objet distracteur latéral et par le degré d'ouverture de l'œil (l'inattention étant souvent associée à la fatigue dans ce type d'étude). En conversation, il est nécessaire d'adapter les indices de repérage puisque le regard joue une fonction dans la gestion des tours de parole par les participants (*cf.* le travail pionnier de Kendon [10]) : la fixité du regard seule, par exemple, ne pourrait constituer un indice dans la mesure où l'auditeur peut regarder le locuteur de manière relativement fixe pendant un tour de parole long. Cela ne signifie pas qu'il est inattentif, comme le montrent les nombreux backchannels émis dans de telles séquences (Bertrand, Ferré et al. [4]). Il en va de même pour l'éloignement du regard ; Lee et al. ([12]) ont montré que le locuteur ne regarde pas l'interlocuteur pendant la plus grande partie de son temps de parole. Cela ne signifie pas non plus qu'il est inattentif.

Les études portant sur l'inattention de manière générale prennent en compte une plus grande fixité du regard et une moins grande fréquence des clignements d'yeux (voir [2] et [9]).

Dans le corpus CID, j'ai retenu les paramètres suivants comme signes d'inattention : (1) le regard du locuteur est détourné de l'interlocuteur ; il est dans l'alignement du corps du locuteur (regard primaire) et n'est pas focalisé sur un objet quelconque, (2) le regard est parfaitement fixe pendant tout l'énoncé exprimé, (3) le locuteur ne cligne pas des yeux.

2.3. Annotation et traitement des séquences d'inattention

Dans ELAN, j'ai annoté les séquences d'inattention des participants uniquement lorsque celles-ci étaient accompagnées de parole. Dans une seconde tire, j'ai annoté les gestes manuels des participants sur ces séquences, en adoptant la typologie décrite dans Bertrand et al. ([3]). J'ai ensuite exporté les tires d'annotation dans Praat ([5]) que j'ai mises en relation avec la tire de transcription orthographique réalisée au LPL, afin d'effectuer une analyse prosodique des séquences retenues.

J'ai ensuite calculé la fréquence fondamentale de toutes les séquences avec un empan temporel de 0.01 s sur les parties voisées du signal, ainsi que l'intensité. Afin d'effectuer des comparaisons, j'ai effectué les mêmes calculs sur des séquences situées en milieu de tour de parole long pour chaque locuteur et où les participants ne montraient pas de signe d'inattention. Enfin, j'ai transformé les valeurs obtenues en valeurs logarithmiques afin de neutraliser les variations de plage intonative et de plage d'intensité entre les locuteurs, à l'instar de Couper-Kuhlen ([6]). Je n'ai en revanche pas calculé le débit de parole des locuteurs qui, perceptivement, ne semblait pas affecté dans la « pensée à voix haute ».

3. RÉSULTATS

3.1. Prosodie

En ce qui concerne la fréquence fondamentale, j'ai effectué deux types de calculs. Le premier consistait à savoir si la fréquence fondamentale moyenne était plus basse dans le cas de la pensée exprimée à voix haute que dans le cas du discours adressé à l'autre et le test ANOVA s'est révélé significatif me permettant de dire que c'est effectivement le cas (p-value = 3.65e-11 ; F=43.93, DF=6501, N=3251). Le deuxième calcul consistait à savoir si la plage de F0 était plus réduite dans le cas de la pensée à voix haute que dans le discours adressé à l'autre et c'est aussi le cas avec p=5.35e-90 (F=417.54, même degré de liberté et même nombre d'échantillons que le calcul précédent). Si l'on prend les locuteurs de manière individuelle, F0 s'est révélée significativement plus basse dans la pensée exprimée à haute voix que dans le discours adressé à l'autre pour 4 locuteurs sur les 6. On obtient les mêmes résultats pour l'écart moyen de F0 avec F0min (plage intonative).

En ce qui concerne l'intensité, j'ai effectué le même type de calculs que pour F0, souhaitant savoir dans un premier temps si l'intensité de la pensée exprimée à voix haute était moins élevée que l'intensité du discours adressé à l'autre. Pour l'ensemble des locuteurs, le test s'est révélé significatif (p=1.67e-68 ; F=309.4, DF=13641, N=6821) ce qui me permet de dire que l'intensité moyenne de la pensée exprimée à haute voix est plus basse que celle du discours adressé à l'autre. Enfin, de la même manière, le test statistique a montré que les écarts d'intensité sont moins élevés dans la pensée exprimée à voix haute que dans le discours adressé à l'autre (p=1.42e-123 ; F=570.57, même degré de liberté et même nombre d'échantillons que le calcul précédent). Ce paramètre est cependant beaucoup plus fiable que la fréquence fondamentale puisque le test s'est révélé significatif pour les 6

locuteurs pris de manière individuelle, tant en ce qui concerne l'intensité moyenne que les écarts d'intensité.

3.2. Gestes

En ce qui concerne les gestes, l'étude est beaucoup moins complète dans la mesure où l'annotation des gestes manuels des locuteurs est inachevée. En effet, si j'ai annoté les gestes manuels des locuteurs sur les séquences de pensée à voix haute, je ne possède pas d'annotations en nombre suffisant sur les séquences de discours adressé à l'autre pour pouvoir effectuer des comparaisons en termes de densité gestuelle par exemple. Néanmoins, il est apparu que sur les séquences de pensée à voix haute, les seuls gestes manuels produits par les locuteurs sont des adapteurs (aussi appelés gestes d'auto-contact, c'est-à-dire des gestes où la main touche une partie du corps du locuteur), et encore sont-ils peu fréquents. Le corps du locuteur qui exprime une pensée à voix haute est en position de repos (dos appuyé au dossier du siège, tête dans l'alignement du corps, bras reposant sur les accoudoirs du siège sauf en présence d'un adapteur).

4. DICUSSION

Ce qui est apparu dans les résultats de l'étude des séquences d'inattention du locuteur est que les paroles qui accompagnent ces séquences présentent des marques prosodiques et gestuelles congruentes. Sur le plan prosodique, elles sont produites dans la plage basse des locuteurs avec peu de variations mélodiques (plage compressée). Elles sont aussi produites avec une intensité basse (ce sont parfois des remarques à peine audibles) et avec peu de variations d'intensité. Sur le plan gestuel, le locuteur a un regard fixe et absent, sans clignement des yeux, ne produit pas de gestes manuels co-verbaux et le reste du corps est relâché. Autrement dit, tous les signes d'investissement dans la conversation sont absents, marquant ainsi le repli sur lui-même du locuteur. En dehors de ces paramètres physiques directement observables, les séquences de pensée à voix haute partagent aussi des caractéristiques sémantiques et structurelles.

Si l'on regarde les exemples de pensées à voix haute suivants :

> *c'était complètement hallucinant comme situation*
> *j'aime bien ces trucs-là quoi*
> *toute façon c'est pas l'pire quand même*
> *ah il était bien c't'appart*
> *non c'était drôle*

on remarque que ce sont des séquences relativement brèves dans lesquelles le locuteur exprime toujours une évaluation sur ce qui a été dit précédemment. Ceci peut évoquer la phase d'évaluation des séquences narratives (Labov & Waletzky [11], Ferré [8]). Cependant, je ne pense pas qu'il s'agisse du même type de séquence pour deux raisons. La première raison est que les pensées exprimées à voix haute peuvent apparaître après des séquences narratives monologales, mais elles peuvent également apparaître après des séquences dialogales avec un même sémantisme d'évaluation. Elles ne sont donc pas spécifiques au récit comme peuvent l'être d'autres phases laboviennes telles que l'*orientation*, la *complication*, etc. D'autre part, lorsqu'elles interviennent après une séquence narrative, elles sont séparées du reste du récit par une pause silencieuse et reprennent l'évaluation donnée en fin de narration, comme dans l'exemple suivant :

> **apogée du récit :** *on arrivait pas avancer au moment on arrivait au col on s'est rendu compte qu'y avait plus d' vent et qu'on pouvait s' redresser et à c' moment-là me suis trouvée face à face avec des gens qui nous regardaient comme ça*
>
> **évaluation du récit :** *je m'suis dit mon Dieu mais quelle horreur*
>
> (0.318)
>
> **sotto voce :** *quelle horreur ah ouais non là ç'avait été complèt'ment ridicule aussi comme situation*

Dans cet exemple, je donne l'apogée du récit qui est immédiatement suivie de la phase d'évaluation. Cette phase est elle-même suivie d'une pause silencieuse assez brève, puis la locutrice se fige et prononce une deuxième évaluation à voix beaucoup plus basse et avec une mélodie basse et plate. Sémantiquement parlant, il n'y a rien de nouveau dans ce qu'elle dit puisque le ridicule et l'horreur de la situation ont tous les deux été mentionnés dans le récit lui-même. De ce fait, si cette pensée exprimée *sotto voce* peut se rattacher au récit qui la précède, elle s'en distingue néanmoins prosodiquement.

Un autre point mérite d'être mentionné. Les séquences évaluatives des récits sont entérinées par le feedback de l'interlocuteur. Les pensées exprimées à voix hautes ne le sont en revanche jamais. Elles ne sont pas destinées à l'interlocuteur, même si le fait de dire quelque chose marque une certaine prise en compte de l'autre, et ne trouvent aucune réponse chez l'interlocuteur, même sous la forme d'un backchannel gestuel ([4]). De plus, après une pensée exprimée à voix haute, l'un ou l'autre des participants prend la parole en changeant le sujet de conversation. Ainsi, si une pensée exprimée à voix haute intervient après que le locuteur a raconté un récit, jamais celui-ci ne développe son récit après la pensée exprimée à voix haute. Soit il poursuit la conversation en initiant une nouvelle séquence narrative, soit il engage une séquence dialogale.

A la lumière des exemples rencontrés dans le corpus, on peut donc dire que la pensée exprimée à voix haute ne forme pas une clôture de séquence (la clôture serait la phase d'évaluation dans une séquence narrative, par exemple), mais une post-clôture ou une clôture de *topic*. A l'instar de Schegloff & Sacks [15] qui voient

dans certains énoncés des *pre-closing sequences* (énoncés annonçant une fermeture de séquence), il me semble que dans la pensée exprimée à voix haute, on peut voir une *post-closing sequence*, où le locuteur prend le temps d'un repli sur lui-même entre deux séquences thématiques. Pendant cet instant, il n'est plus attentif à l'interaction en cours, inattention marquée par la prosodie et la gestualité.

5. CONCLUSION

Si la pensée exprimée à voix haute où le locuteur montre une certaine inattention dans la conversation a fait l'objet de si peu d'études jusqu'à présent, c'est principalement parce que les études – prosodiques notamment – reposent sur des enregistrements audio de conversations spontanées. Or, j'ai montré dans cet article que l'inattention du locuteur est marquée par un faisceau d'indices auditifs et visuels. Une fréquence fondamentale ou une intensité basse ne sont pas en elles-mêmes des marqueurs d'inattention. Par contre, lorsque ces marques sont associées à un regard primaire fixe, sans clignement des yeux, mais aussi à une absence de gestes manuels autres que les adapteurs et à un relâchement du corps du locuteur, alors, on peut penser que le locuteur manifeste un certain désengagement conversationnel, d'où l'impression d'inattention. D'autres paramètres jouent certainement un rôle également : dans un futur travail, il serait intéressant par exemple de voir si ces séquences sont hypo-articulées par rapport au reste du discours.

L'article montre donc tout l'intérêt d'une étude multimodale pour ce type de *post-closing sequences*. Elle a cependant ses limites : il ne s'agit en effet que d'étudier les signes extérieurs de la pensée à voix haute, pas de savoir ce qui se passe dans la tête du locuteur à ce moment-là. Des techniques comme l'EEG permettraient peut-être de creuser la question, mais l'appareillage qu'elles nécessitent remet directement en cause la spontanéité des données obtenues et donc la présence de *post-closing sequences* dans une conversation. De plus, bien qu'analysant un nombre limité de cas, ce travail repose sur une annotation manuelle longue du corpus dont l'objectivité serait cependant renforcée si elle était réalisée par plusieurs annotateurs, ce qui n'est pas encore le cas.

BIBLIOGRAPHIE

[1] *Anomalie du champ visuel* [Par Dr Tchaplyguine F, Dr Gain P (CHU de Saint-Etienne, Hôpital Bellevue). Site éditeur Université Jean Monnet de Saint-Etienne, Faculté de médecine Jacques Lisfranc ; rubrique DCEM2 et 4/Ophtalmologie

[2] Antrobus, J.S. & Singer, J.L., Eye movements accompanying daydreaming, visual imagery, and thought suppression. *Journal of Abnormal and Social Psychology* 69 : 244–252, 1964.

[3] Bertrand, R., P. Blache, et al. Le CID – Corpus of Interactional Data – : protocoles, conventions, annotations. *Travaux Interdisciplinaires du Laboratoire Parole et Langage (TIPA) 25* : 31-60, 2006.

[4] Bertrand, R., G. Ferré, et al. Backchannels revisited from a multimodal perspective. *Proceedings of Auditory-visual Speech Processing*, Hilvarenbeek, The Netherlands. Non paginé, 2007.

[5] Boersma P. & Weenick D. Praat: A system for doing phonetics by computer. http://www.praat.org

[6] Couper-Kuhlen E. The prosody of repetition: On quoting and mimicry. In: E. Couper-Kuhlen and M. Selting, eds., *Prosody in Conversation: Interactional studies*. Cambridge: Cambridge University Press, 366-405, 1996.

[7] Elan (H. Sloetjes): http://www.lat-mpi.eu/tools/elan/

[8] Ferré, G. Récits de femmes – Analyse multimodale du récit conversationnel en français : une étude de cas. Accepté au *Congrès Mondial de Linguistique Française*. Paris. Juillet 2008.

[9] Gaarder, K. Fine Eye Movements during Inattention, *Nature* 209 : 83 – 84, 1966.

[10] Kendon, A. Some functions of gaze direction in social interaction. *Acta Psychologia*, 26: 1-47, 1967.

[11] Labov, W. & Waletzky J.. Narrative analysis: Oral versions of personal experience. In J. Helm (Ed.), *Essays on the verbal and visual arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. Seattle : University of Washington Press, 12-44, 1967.

[12] Lee, S. P., Badler, I. N. & Badler, J. Eyes Alive, ACM Transaction on Graphics, 21(3): 637-. 644. 2002.

[13] Lim, C., Sayed T. & Navin F. A driver visual attention model. Part 1. Conceptual framework, *Can. J. Civ. Eng*. 31: 463–472, 2004.

[14] Mueller, Erik T., & Dyer, Michael G. (1985). Towards a computational theory of human daydreaming. *Proceedings of the Seventh Annual Conference of the Cognitive Science Society* (pp. 120-129). Hillsdale, NJ: Lawrence Erlbaum.

[15] Schegloff, E. & Sacks, H., Opening up Closings, *Semiotica*, 8/4 : 289-327, 1974.

[16] Underwood G., Chapman P., et al., Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers, *Ergonomics*, 46 : 629-646, 2003.