



HAL
open science

Analyse multimodale de la parole

Gaëlle Ferre

► **To cite this version:**

Gaëlle Ferre. Analyse multimodale de la parole. Rééducation orthophonique, 2011, 246, pp.73-85.
hal-00609124

HAL Id: hal-00609124

<https://hal.science/hal-00609124>

Submitted on 18 Jul 2011

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

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

Analyse multimodale de la parole

Gaëlle Ferré

Laboratoire de Linguistique (LLING)

Université de Nantes

Gaelle.Ferre@univ-nantes.fr

Résumé

Depuis quelques années, le champ de la pragmatique en linguistique, qui se consacre à l'étude des interactions, a pris conscience de l'importance de la production gestuelle dans certaines situations de communication. S'est alors développé un grand nombre d'études dans des domaines variés : on pense aux travaux en plein essor sur les langues des signes, intimement liées à la gestualité, mais aussi à un certain nombre d'études portant sur la multimodalité du langage, dont le but est de montrer les liens qui existent entre la gestualité et la parole en situation de communication. Le langage est alors considéré comme un système linguistique qui intègre parole et gestualité de manière organisée. Parmi ces différents travaux, certains se penchent sur l'acquisition du langage (acquisition simultanée de la parole et de la gestualité), d'autres sur les interactions entre parole et gestualité chez des locuteurs adultes. Enfin, des travaux intéressants portant sur les troubles du langage et/ou de la gestualité co-verbale voient également le jour. Dans l'optique de faire émerger ce type de recherches qui trouvent des applications dans le domaine de l'orthophonie, cet article propose de faire le point sur la multimodalité du langage en présentant certains des travaux réalisés à partir de corpus de parole spontanée.

Abstract

Pragmatics – as a particular field of linguistics especially concerned with the study of interactions – has recently become aware of the importance of gestures in some communication situations. Since then, numerous studies have been developed in a certain number of domains: research on Sign Languages which are intimately linked with gestures has of course become of major general interest, as well as a certain number of other works on the multimodality of language, which aim at establishing the links between gesture and speech in communication situations. In the latter type of research, language is considered as a linguistic system that combines speech and gesture in an organized way. While some of the studies examine language acquisition (in a perspective where speech and gesture are acquired in a simultaneous way), others focus on the interaction between speech and gesture in adults. There is also a growing body of research on pathologies affecting speech and/or gesture production, although they are not yet quite common in France. The objective of this paper is to encourage collaborative work on pathology and linguistics, which would find direct applications in speech therapy, by presenting the current research in multimodality involving corpora of spontaneous speech.

Mots-clés

Multimodalité, gestualité co-verbale, système linguistique

Introduction

La multimodalité s’inspire des travaux des anthropologues américains E.T. Hall (1971, 1981, 1984 pour les traductions françaises pour ne citer que quelques travaux) et R. Birdwhistell (1968) dans les années 1960. Ce n’est cependant que dans les années 1990, avec le développement technologique permettant le stockage et le traitement de données vidéos, que la multimodalité a connu un essor considérable dans le champs de la linguistique aux Etats-Unis. L’analyse multimodale consiste à mettre en relation des informations linguistiques produites dans différentes modalités, chacune d’elle contribuant à l’élaboration et à la perception du message communiqué. Ainsi, l’on peut distinguer la modalité verbale qui comporte plusieurs niveaux — phonèmes, choix du lexique, organisation syntaxique, organisation discursive — de la modalité orale — prosodie, qualité de voix — et enfin de la modalité visuelle — gestualité et expressions faciales. Encore assez peu développée au niveau international à cette époque (on citera néanmoins les travaux de Cosnier & Brossard, 1984, et Calbris & Porcher, 1989 en France), la multimodalité intéresse particulièrement les linguistes depuis le début des années 2000. Ce type d’analyse possède effectivement un grand nombre d’applications dans des domaines très variés parmi lesquels on peut citer le développement d’agents animés (Cassell et al., 1994, Pelachaud & Poggi, 1998, 2002, pour ne citer que quelques travaux dans un champs très productif) et la remédiation thérapeutique. Ces deux domaines se combinent parfois, ainsi que le montre l’étude de Lee & Cherney (2008), qui rendent compte du développement d’un logiciel figurant une tête parlante utilisée en thérapie par des patients aphasiques.

Ces travaux ont pour point commun de considérer que la gestualité joue un rôle dans la communication. On peut cependant aller plus loin et penser que la gestualité et la parole forment un système linguistique intégré, et que, comme chaque élément de ce système, la gestualité possède plusieurs fonctions intra- et inter-personnelles, ainsi que le formulent Morsella & Krauss (2004 : 421). Selon ces deux chercheurs, du point de vue de la production de la parole, “hand-arm movements seem to facilitate both spatial working memory and speech production. (...) participants gestured more when describing visual objects from memory and when describing objects that were difficult to remember and encode verbally”¹ (2004 : 412). De la même manière, dans la parole pathologique, Scharp et al. (2007 : 1) considèrent que : “aphasia researchers and clinicians consider gesture as both a means of communicative facilitation and compensation”². La gestualité facilite également la perception de la parole dans sa dimension verbale (Grandström & House, 2007) mais aussi orale (voir les travaux de Swerts & Krahmer, 2008, Al Moubayed et al., 2010 sur la perception de phénomènes prosodiques tels que les accents de mot). Ainsi, comme le suggèrent Moubayed et al. (op. cit.) : “visual cues of acoustic prominence can aid speech intelligibility.

¹ Les mouvements des bras et des mains semblent faciliter à la fois la mémoire de travail spatiale et la production de la parole. (...) Les participants gestualisaient plus lorsqu’ils décrivaient des objets visualisés puis mémorisés et lorsqu’ils décrivaient des objets dont la forme était difficile à retenir ou à encoder verbalement.

² Les chercheurs et le personnel médical travaillant sur l’aphasie considèrent que la gestualité permet à la fois de faciliter la communication et de compenser une communication défailante.

(...) Visualizing prominence may also provide information about speech rhythm and syllable boundaries of the underlying linguistic segment”³ (p. 308).

Si la multimodalité du langage est déjà prise en compte dans certaines études, comme c’est le cas de Goodwin (2000) qui étudie la communication multimodale d’un aphasique, ainsi que dans certains protocoles d’évaluation des troubles du langage (*Functional Communication Profile* et *Pragmatic Protocol*⁴, cités dans Prins & Bastiaanse, 2004), l’étude de la gestualité dans la communication non pathologique n’est sans doute pas encore suffisamment avancée pour qu’une collaboration totale soit possible à l’heure actuelle entre thérapeutes et linguistes. Il est cependant possible de croire que cette collaboration peut émerger dans les années qui viennent. Dans cet espoir, l’objectif de cet article est de produire un état des lieux des études linguistiques multimodales aujourd’hui, en se concentrant essentiellement sur la dimension visuelle de la parole, moins connue que les dimensions verbales et prosodiques.

1. Qu’est-ce qu’un geste ?

La définition qui est adoptée ici se veut la plus large possible : par geste, nous entendons tout mouvement de la face/tête, du buste, des bras/mains ou des épaules réalisés en dehors des changements de posture. Un changement de posture, quant à lui, est compris comme une modification de la position de repos (par exemple, déplacer un bras de l’accoudoir du siège pour le poser sur ses jambes en position assise). Alors que la tension musculaire nécessaire à un changement de posture n’a de but que d’atteindre une absence de tension, cela n’est pas le cas des autres mouvements du corps. Nous limitons les gestes à la partie supérieure du corps car ces mouvements sont plus susceptibles d’être en lien avec la dimension verbale et vocale de la communication.

1.1. Gestes non verbaux

Il existe deux types de gestes non verbaux : les gestes d’auto-contact ou adaptateurs, décrits par McNeill (1992, 2005) et les ‘gestes-action’. Les adaptateurs sont des gestes de confort réalisés par les personnes, comme remettre en place une mèche de cheveux, dans un contexte où ce geste n’est pas en relation avec la parole. Il y a en effet une différence entre toucher ses cheveux sans que le contexte verbal ait de lien avec les cheveux d’une manière ou d’une autre (adaptateur) et toucher ses cheveux tout en mentionnant le coiffeur par exemple. Dans ce dernier cas, le geste n’est plus un adaptateur mais permet d’établir un lien référentiel entre les cheveux et le coiffeur. Malgré leur caractère non verbal (“self-touching/grooming gestures are not tied to verbal expression, which sets them apart from communicative gesture”⁵, Scharp et al., 2007 : 5), les adaptateurs sont souvent pris en compte dans les

³ Les indices visuels de la saillance acoustique peuvent améliorer l’intelligibilité de la parole. (...) La visualisation de la proéminence accentuelle peut également apporter des informations sur le rythme et les frontières syllabiques d’un segment linguistique sous-jacent.

⁴ *Profil de Communication Fonctionnel et Protocole Pragmatique*.

⁵ Les gestes d’auto-contact et les gestes réalisés à des fins de rectification esthétique d’une partie du corps ou des vêtements ne sont pas liés à l’expression verbale, ce qui les distingue des gestes communicatifs.

travaux sur corpus vidéos (cf. section 2), car ils ont un lien avec l'organisation séquentielle et interactionnelle dans un échange verbal. L'on remarque en effet que la densité d'adaptateurs produits est beaucoup plus importante lorsqu'un participant à une interaction est en position d'écoute que lorsque celui-ci prend un tour de parole. Ces gestes sont donc révélateurs de l'intention de prise de parole ou non par le participant et sont perçus comme tels par l'interlocuteur. La densité d'adaptateurs est également plus élevée dans des situations de recherche lexicale ou de recherche de formulation.

Le deuxième type de gestes non verbaux sont ce que l'on peut nommer les 'gestes-actions'. Il s'agit de tous les gestes réalisés au quotidien, dans un cadre personnel ou professionnel, dans le but de 'réaliser quelque chose' ; conduire, par exemple, suppose une série de 'gestes-action' comme tourner le volant, changer de vitesse, etc. Ces gestes ne sont en relation avec la parole que dans les cas de discours pédagogique (par exemple une leçon de conduite), ou encore dans les cas où le discours commente le geste pour en souligner l'inadéquation par exemple. Ces gestes sont très peu présents dans les corpus sur lesquels nous travaillons, qui sont orientés vers l'interaction. En revanche, d'autres laboratoires se penchent sur les interactions verbales en milieu naturel — interactions dans la vie quotidienne — comme c'est le cas du laboratoire ICAR à Lyon, et prennent en compte dans leurs analyses l'intégration des gestes-action dans l'interaction.

1.2. Gestes co-verbaux

Les gestes co-verbaux sont des mouvements qui peuvent être mis en relation avec la parole. Ils sont produits avec différents articulateurs :

- **Les sourcils** : les haussements ponctuels et rapides des sourcils sont traditionnellement associés aux proéminences accentuelles (Swerts & Kraemer, 2008 entre autres), même si, selon Al Moubayed et al. (2010) : "head-nods have been shown to be a stronger cue in the perception of prominence than eyebrows"⁶ (p. 303). Les mouvements des sourcils maintenus sur tout un énoncé ont plutôt une fonction de modalisateur au sens où ils apportent une modalité énonciative à l'énoncé.

- **La tête** : les mouvements de tête sont régulièrement utilisés comme mouvements emphatiques lorsqu'ils sont produits par le locuteur et comme backchannels⁷ lorsqu'ils sont produits par l'interlocuteur. Ils peuvent également avoir une fonction de pointage vers un espace référentiel ou jouer, à l'instar des sourcils, un rôle de modalisateur d'énoncé.

- **La bouche** : les mouvements de bouche produits hors articulation des sons permettent d'apporter une modalité appréciative à l'énoncé.

- **Le buste** : les mouvements de buste sont plutôt associés à l'organisation séquentielle de l'interaction en termes de prise de parole ou cession de tour de parole. Ils peuvent également marquer l'affiliation ou la non affiliation de l'interlocuteur avec ce qui lui a été dit précédemment.

- **Les mains** : Les gestes manuels jouent un rôle lexical ou grammatical. Certains gestes ont en effet pour fonction d'illustrer un objet ou de mimer une action et sont ainsi liés aux morphèmes lexicaux utilisés dans la parole, tandis que d'autres ont pour fonction d'établir un lien de co-référence entre des

⁶ Les gestes affirmatifs de la tête se sont révélés constituer de meilleurs indices de la proéminence accentuelle que les mouvements des sourcils.

⁷ Réponses minimales de l'interlocuteur.

unités verbales, d'indiquer l'organisation du discours en différentes unités, de renforcer une unité lexicale ou encore d'apporter une modalité et ont donc une fonction grammaticale. McNeill (1992, 2005) propose une classification typologique des différents gestes manuels qui est présentée dans la section 1.3.

L'ensemble de ces gestes, selon Kendon (cité par McNeill, 2005 : 5), s'inscrit dans un continuum allant du moins au plus conventionnel. Les gestes conventionnels sont partagés par l'ensemble des membres d'une même communauté linguistique et peuvent remplacer la parole, alors que les gestes non conventionnels sont produits par les locuteurs de manière plus ou moins idiosyncratique (plus ou moins car un geste peut avoir certaines caractéristiques conventionnelles, comme la forme de la main par exemple, mais être adaptés à la situation spécifique de discours) et ne peuvent donc pas se substituer à la parole.

1.3. Classification des gestes manuels (McNeill, 1992, 2005)

En 1992, McNeill propose une classification typologique des gestes manuels qui est largement utilisée au niveau international. C'est cette typologie que nous adoptons ici et qui consiste à classer les gestes manuels en fonction de leur relation avec la parole (cela exclut donc les gestes non verbaux) :

- **Battements** : les battements peuvent être des gestes indépendants, ou encore être réalisés au cours de la production d'un autre geste. Il s'agit d'un mouvement simple de la main qui monte puis redescend. Leur rôle est encore relativement obscur. Selon McNeill (1992), ils seraient utilisés pour marquer les différentes étapes du discours.

- **Butterworths** : gestes désordonnés utilisés lors des difficultés de formulation.

- **Déictiques** : gestes de pointage vers soi-même, l'interlocuteur, une tierce partie ou un objet, réalisés pour indiquer à qui ou de qui/quoi l'on parle. Ces gestes peuvent également être réalisés de manière plus « abstraite » si un référent a été posé au préalable dans une partie de l'espace et que l'on s'y réfère ensuite dans la suite du discours pour désigner ce référent.

- **Emblèmes** : gestes conventionnels (comme le geste que l'on réalise pour dire « on se téléphone »).

- **Iconiques** : ces gestes sont de deux natures différentes. Certains de ces gestes reprennent une caractéristique physique d'un objet ou d'une personne (pour parler d'un cylindre par exemple, les doigts sont mis dans une configuration arrondie, alors que pour décrire une forme rectangulaire ou carrée, le pouce et l'index forment un angle droit). Ces gestes peuvent également constituer un mime d'action (l'action de visser, par exemple, implique une rotation du poignet). Deux points de vue peuvent alors être adoptés : soit l'action est réalisée avec une partie du corps qui figure l'ensemble de l'actant (bouger les doigts pour figurer l'action de marcher) et le locuteur adopte alors un point de vue d'observateur, soit le locuteur utilise l'intégralité de son corps pour mimer l'action (bouger les coudes pour figurer l'action de courir) et il adopte alors un point de vue d'actant. La frontière est parfois mince entre les iconiques et les emblèmes car certains iconiques ont été complètement conventionalisés, alors que d'autres constituent de légères adaptations de gestes conventionnels, quand d'autres enfin ne sont réalisés que par un locuteur dans une situation spécifique.

- **Métaphoriques** : ces gestes figurent des idées abstraites, par exemple, déplacer la main vers la droite pour signifier « et ensuite ». Ils sont plus souvent liés à l'information grammaticale qu'à l'information lexicale du message.

L'ensemble de ces gestes se rencontrent dans les interactions spontanées et forment un tout avec la parole. Leur acquisition suit d'ailleurs le même développement que l'acquisition du langage chez les enfants, qui produisent d'abord des déictiques, puis des iconiques (en lien avec le développement de leur vocabulaire) et enfin, les battements et les métaphoriques (Colletta, 2004). L'acquisition des emblèmes est plus progressive.

1.4. Gestes bi-dimensionnels

Suite au travail sur corpus qu'un certain nombre de chercheurs ont mené dans la communauté scientifique internationale, McNeill (2005), reprenant les catégories définies en 1992, ajoute néanmoins un point capital : un geste peut relever de plusieurs catégories à la fois (dans notre propre analyse de corpus, nous n'avons pas trouvé de gestes relevant de plus de deux catégories simultanément). Cet aspect montre toute la complexité de la modalité visuelle.

Dans le cas d'un geste réalisé avec une seule main, il est possible en effet que le geste ne relève que d'une catégorie, mais il peut aussi relever de deux catégories différentes : par exemple, la configuration des doigts de la main peut figurer un type de déplacement (geste iconique) et la trajectoire du geste, une direction (geste déictique).

Dans le cas d'un geste bi-manuel, si les deux mains sont symétriques, on se situe dans le même cas de figure qu'avec un geste réalisé avec une seule main, mais il est également possible que le geste soit réalisé avec deux mains en configuration asymétrique. Dans ce cas, il n'est pas impossible qu'une main serve de point de référence à l'autre main qui va réaliser le mouvement. Ainsi, dans un corpus, pour l'expression « longer la route », le locuteur place sa main gauche à plat et effectue un mouvement figurant « avancer auprès de » avec la main droite qui frôle la main gauche. Les deux mains réalisent alors un geste iconique mais la main gauche a aussi une fonction de référent déictique.

Ceci a une influence en termes d'annotation de corpus, qui doit pouvoir rendre compte de la multi-dimensionnalité des gestes manuels, sachant que l'on peut également rencontrer deux gestes produits en chevauchement partiel ou total par chacune des deux mains.

2. Corpus d'annotation multimodale

Afin de pouvoir enrichir l'analyse multimodale de la parole, un certain nombre de projets se sont développés, dont l'objectif est d'annoter des corpus vidéos de manière systématique. L'ANR a ainsi financé en 2005 le projet « L'acquisition et les troubles du langage au regard de la multimodalité de la communication parlée », décrit dans Colletta et al. (2008). L'ANR a également financé le projet « Outils pour le traitement des informations multimodales », décrit dans Bertrand et al. (2008), et sur une partie duquel repose cet article. Ce projet a pour but de constituer un corpus de français parlé, annoté linguistiquement dans différents domaines : phonétique, prosodie, morphologie, syntaxe, discours et gestualité. Le but est de pouvoir mettre en relation des informations linguistiques de différents niveaux, annotées indépendamment les unes des autres, et de pouvoir appliquer la méthodologie d'annotation à

d'autres types de corpus (il est très envisageable par exemple, avec des modifications mineures, d'annoter des corpus de parole pathologique et de comparer ensuite ces corpus à de la parole spontanée non pathologique). L'annotation est réalisée grâce aux logiciels Praat (Boersma & Weenink, 2009) pour ce qui concerne les annotations phonétiques et prosodiques et Anvil (Kipp, 2001) pour l'annotation de la gestualité. Il serait trop long ici de donner le détail du schéma d'encodage utilisé pour la gestualité, qui est décrit de manière assez détaillée dans Tan et al. (2010). Il nous semble en revanche plus intéressant de voir quelles observations peuvent permettre ce type de corpus — sachant que l'annotation du corpus CID est toujours en cours car il s'agit d'un processus complexe où l'annotation ne peut se faire que manuellement à l'heure actuelle.

2.1 Variabilité des gestes

La première question qui se pose sur les gestes concerne leur variabilité et donc la difficulté d'effectuer des généralisations sur les corpus. En effet, McNeill (1992) mentionne les différences culturelles qu'il peut y avoir d'une communauté linguistique à une autre, tant dans la densité gestuelle que dans la forme des gestes produits. Mais qu'en est-il de la variabilité au sein d'une même communauté linguistique ? A partir du corpus CID, pour lequel nous avons à l'heure actuelle annoté 1795 gestes manuels produits par 6 locuteurs, il apparaît que le nombre de gestes produits par chaque locuteur varie du simple au double et qu'il y a également une variation en termes d'amplitude des gestes produits selon le locuteur. En revanche, un test ANOVA révèle que le nombre de gestes moyen produit par les locuteurs ne varie pas de manière significative ($F(2.62)=1.39, p=0.26$), non plus que le nombre moyen par type de geste ($F(2.75)=2.03, p=0.12$). Ce qui varie en revanche, c'est la répartition des types de gestes pour chaque locuteur. Ainsi, comme l'indique la Figure 1 ci-dessous, le locuteur 2 du corpus produit moins d'adaptateurs que le locuteur 1 au profit des métaphoriques. Les autres types de gestes sont produits dans des proportions similaires par les deux locuteurs. Il serait d'ailleurs intéressant de voir si ces variations ont un lien avec le temps d'articulation de chaque locuteur.

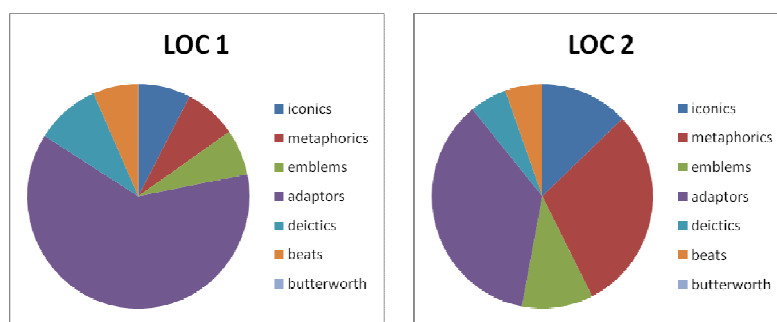


Figure 1 : Proportion de gestes manuels pour deux locuteurs du corpus CID.

Un autre source possible de variation des gestes tient dans le genre discursif. Ceci ne peut pas être testé sur le CID qui ne comporte que des enregistrements de parole conversationnelle, avec une tâche qui favorise le genre narratif, mais une observation superficielle d'autres types d'enregistrements semble montrer que le type de gestes produits par les locuteurs varie effectivement selon le genre discursif :

- **Dialogue vs. monologue** : les adaptateurs seront beaucoup moins présents dans des situations de monologue que dans des situations de dialogue.

- **Formel vs. informel** : les iconiques seront moins présents dans un discours formel que dans un discours informel au profit des métaphoriques.

- **Intention communicationnelle** : dans un discours injonctif ou pédagogique, il est vraisemblable que la proportion de déictiques sera plus élevée que dans d'autres types de discours, alors que dans un discours argumentatif, on trouvera plus de métaphoriques et de battements et dans un discours narratif, plus d'iconiques.

S'il peut y avoir une grande variabilité selon le genre discursif pour ce qui concerne les gestes manuels, il est en revanche peu probable que l'on observe la même variabilité pour les autres types de gestes (mouvements de tête, des sourcils) qui jouent un rôle plus grammatical dans le discours. De plus, si l'on considère les caractéristiques des gestes manuels produits, l'on constate que ce sont régulièrement les mêmes configurations manuelles (forme des doigts), trajectoires et types de mouvements qui reviennent régulièrement. En conclusion, la variabilité observée sur le plan de la gestualité n'est sans doute pas plus importante que la variabilité existant dans d'autres domaines linguistiques (phonétique, syntaxe...).

2.2 Alignement gestes / parole

Si certains gestes sont produits en complémentarité de la parole, d'autres en revanche apportent une information qui est redondante par rapport au message verbal, même si cette redondance n'est jamais totale dans la mesure où le simple fait de produire un geste redondant non obligatoire peut modifier la perception du message (en termes d'emphase par exemple). Ces gestes redondants, en particulier certains iconiques, sont particulièrement utiles dans la mesure où ils permettent d'étudier très précisément leur alignement avec la parole, puisqu'il est possible pour ce type de gestes de les associer à ce que l'on nomme un 'affilié lexical' — un mot du discours illustré par le geste. Dans une étude précédente (Ferré, 2010), nous avons ainsi pu étudier l'alignement geste/parole pour 244 gestes iconiques produits par les 6 locuteurs du corpus CID. Loehr (2004) avait trouvé une synchronisation entre l'apogée des gestes (tous gestes confondus) et les accents lexicaux (tous accents confondus), même s'il adoptait une fourchette temporelle de plus ou moins 275 ms, ce qui nous semble relativement large. Aussi n'avons-nous pas concentré nos efforts sur l'accent. En revanche, nous souhaitions avoir plus de précisions sur l'alignement temporel du geste avec le groupe intonatif tel qu'il est défini dans la théorie de Selkirk (1978), ainsi que l'alignement de la phase dynamique du geste avec l'affilié lexical correspondant. Dans ce but, nous avons annoté chaque phase de réalisation des gestes manuels en adoptant les phases gestuelles suggérées par Kendon en 1980 (cité dans McNeill, 1992) : le geste commence par une *préparation* (mise en place des articulateurs), puis vient une phase de *réalisation* (le geste lui-même), qui peut être suivie d'une *tenue* (les articulateurs font une pause) avant la *rétraction* (retour des articulateurs à une position de repos). Nous avons ensuite mis en relation le début de la phase de préparation avec le début du groupe intonatif et le début de la phase de réalisation avec le début de l'affilié lexical. Nous avons aussi mesuré la durée de l'affilié lexical et du groupe intonatif d'une part, que nous avons comparée avec la durée du geste dans son ensemble et de la phase de réalisation d'autre part.

Sur le plan de la durée, il est apparu que le geste dans son ensemble est plus long que le groupe intonatif comportant l'affilié attaché au geste. De la même manière, la phase dynamique du geste (réalisation) est plus longue que l'affilié lexical auquel elle correspond.

Sur le plan de l’alignement temporel, il est apparu que le geste dans son ensemble commence avant le groupe intonatif et se termine légèrement après le groupe intonatif. De même, la phase dynamique du geste commence avant l’affilié lexical auquel elle correspond et se termine légèrement après celui-ci (les chiffres précis sont donnés dans Ferré, 2010). Nous pouvons ainsi proposer l’organisation temporelle suivante :

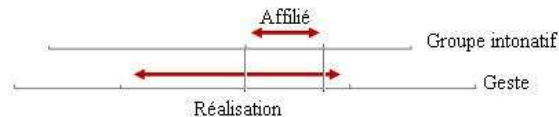


Figure 2 : Organisation temporelle des iconiques redondants par rapport à la parole.

Dans une proposition (Ferré, soumis), nous allons systématiser l’annotation des affiliés afin de pouvoir établir des liens entre les pointages (gestes déictiques) et leur affilié lexical, mais aussi entre les pointages et d’autres types de gestes. Cela permettra de comprendre comment se construit la référenciation au fil du discours, car s’il y a un lien évident entre les gestes et la parole qu’ils accompagnent, les locuteurs établissent également des liens entre certains gestes et d’autres gestes produits plus tôt dans le discours en se référant à un espace déterminé, ou de manière simultanée (pointer du regard vers un geste réalisé avec les mains). D’autres études sont en cours sur le rôle du geste dans la réalisation de la focalisation ou la construction de la modalité, et sur l’effet de la recherche lexicale sur le geste. Ce dernier point est particulièrement intéressant. Ainsi, Scharp et al., citant une étude de Mayberry & Jaques réalisée en 2000, mentionnent le fait que les personnes bègues stoppent la phase de réalisation du geste tout en maintenant les mains dans la même configuration jusqu’à la fin de la dysfluente (2007 : 7). Si l’on compare ce résultat avec les premières observations faites sur corpus de parole spontanée chez des locuteurs ne présentant pas de pathologie, il apparaît que les choses se passent différemment lors de la recherche lexicale : le geste est d’abord réalisé dans son intégralité, puis répété, la plupart du temps dans des formes atténuées (avec une moins grande amplitude le plus souvent), jusqu’à ce que le mot juste soit produit. Ces premières observations mériteraient grandement d’être vérifiées de manière systématique.

Conclusion

“Verbal communication and gesture are produced in parallel, and gesture is a potentially equal participant in the conceptual/planning stages. Communicative expression thus occurs via both verbal and spatial means, providing a temporally linked, non-redundant, multidimensional, content-rich message.”⁸ (Scharp et al., 2007 : 2) Cette citation montre tout l’intérêt des études multimodales à la fois pour décrire la parole non pathologique et ainsi constituer une référence pour l’évaluation des troubles du

⁸ Communication verbale et gestualité sont produites en parallèle, et la gestualité participe tout autant que la parole aux étapes de conceptualisation et de planification du message. L’expression communicative se réalise aussi bien par des moyens verbaux que par des moyens spatiaux, donnant jour à un message non-redondant, multidimensionnel, riche sur le plan du contenu et dans lequel s’expriment des relations temporelles.

langage, mais aussi pour la remédiation thérapeutique. Si les auteurs de la citation considèrent que la prise en compte de la gestualité est un atout dans la thérapie, cette prise en compte suppose que l'on ait une bonne connaissance du fonctionnement de la gestualité co-verbale dans la parole dite normale. L'objectif de cet article a donc été de présenter un bref état de l'art des études multimodales, en mettant l'accent sur la constitution et l'annotation de corpus de parole spontanée. Ces annotations permettent d'ores et déjà d'effectuer des mesures sur l'alignement temporel entre geste et parole et rendent possibles d'autres études sur le système linguistique à l'œuvre.

Références :

- Al Moubayed, S., et al., 2010. Auditory visual prominence. From intelligibility to behavior. *Journal of Multimodal User Interfaces* 3, 299-309.
- Bertrand, R., et al., 2008. Le CID - Corpus of Interactional Data - Annotation et Exploitation Multimodale de Parole Conversationnelle. *Traitement Automatique des Langues* 49, 105-133.
- Birdwhistell, R.R., 1968. L'analyse kinésique. *Langages* 10, "Pratiques et langages gestuels", 101-106.
- Boersma P., & Weenink, D. 2009. Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from <http://www.praat.org/>
- Calbris, G., Porcher, L., 1989. *Geste et communication*, Hatier ed. CREDIF, Paris.
- Cassell, J., et al., 1994. Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In *Proceedings of SIGGRAPH '94*, Orlando, Florida, pp. 413-420.
- Colletta, J.-M., 2004. *Le développement de la parole chez l'enfant âgé de 6 à 11 ans. Corps, langage et cognition*. Pierre Mardaga Éditeur, Sprimont (Belgique).
- Colletta J.-M., et al. 2008. Multitrack annotation of child language and gestures, *LREC 2008 Workshop on Multimodal Corpora*, Marrakech, Disponible à l'adresse : http://embots.dfki.de/doc/MMC2008_proceedings.pdf
- Cosnier, J., Brossard, A., 1984. *La communication non verbale*. Delachaux & Niestlé, Neuchâtel, Paris.
- Ferré, G., 2010. Timing Relationships between Speech and Co-Verbal Gestures in Spontaneous French. In *Proceedings of LREC: Workshop on Multimodal Corpora*, Valetta, Malta, pp. 86-91.
- Ferré, G., soumis. Modification du schéma d'encodage d'OTIM pour l'annotation et l'interprétation des pointages, *Colloque TALN, Atelier DEGELS (Défi Geste Langue des Signes)*, Montpellier, 1^{er} juillet 2011.
- Goodwin, C., 2000. Gesture, aphasia, and interaction. In: D. McNeill (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, pp. 84-98.
- Grandström, B., House, D., 2007. Inside out - Acoustic and visual aspects of verbal and non-verbal communication In *Proceedings of 16th International Congress of Phonetic Sciences (ICPhS)*, Saarbrücken, Germany, pp. 11-18.
- Hall, E.T., 1971. *La dimension cachée*, Editions du Seuil ed, Paris.
- Hall, E.T., 1981. Proxémique. In: J.-L. Giribone (Ed.), *La nouvelle communication*, Editions du seuil ed. Points, Paris, pp. 191-221.
- Hall, E.T., 1984. *Le langage silencieux*, Editions du seuil ed. Point, Paris.
- ICAR, Interactions, Corpus, Apprentissages, Représentations, Lyon, Unité Mixte de Recherche -UMR-5191, <http://icar.univ-lyon2.fr/>

- Kipp, M. (2001) Anvil - A Generic Annotation Tool for Multimodal Dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech)*, pp. 1367-1370.
- Lee, J.B., Cherney, L.R., 2008. The Changing "Face" of Aphasia Therapy. *Perspect Neurophysiol Neurogenic Speech Lang Disord* 18(1), 15-23.
- Loehr, D., 2004. *Gesture and Intonation*. PhD thesis, Georgetown University, Georgetown.
- McNeill, D., 1992. *Hand and Mind : What Gestures Reveal about Thought*. The University of Chicago Press, Chicago and London.
- McNeill, D., 2005. *Gesture & Thought*. The University of Chicago Press, Chicago and London.
- Morsella, E., Krauss, R.M., 2004. The Role of Gestures in Spatial Working Memory and Speech. *The American Journal of Psychology* 117(3), 411-424.
- Pelachaud, C., Poggi, I., 1998. Talking Faces that Communicate by Eyes. In: S. Santi, et al. (Eds.), *Oralité et Gestualité (ORAGE) - Communication multimodale, interaction*. L'Harmattan, Besançon, pp. 211-216.
- Pelachaud, C., Poggi, I., 2002. Subtleties of Facial Expressions in Embodied Agents. *Journal of Visualization and Computer Animation (JVCA)* 13, 301-312.
- Prins, R., Bastiaanse, R., 2004. Review. Analysing the spontaneous speech of aphasic speakers. *Aphasiology* 18(12), 1075-1091.
- Scharp, V.L., et al., 2007. Gesture and aphasia: Helping hands? *Aphasiology* 21, 1-9.
- Selkirk, E., 1978. On Prosodic Structure and its Relation to Syntactic Structure. In: T. Fretheim (Ed.), *Nordic Prosody II*. Tapir, Trondheim, pp. 111-140.
- Swerts, M., Kraemer, E., 2008. Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics* 36, 219-238.
- Tan, N., et al., 2010. Multi-level Annotations of Nonverbal Behaviors in French Spontaneous Conversations. In *Proceedings of LREC: Workshop on Multimodal Corpora*, Valetta, Malta, pp. 74-79.