

Université de Marne-La-Vallée  
Université Paris VII – Denis Diderot

Mémoire de D.E.A.  
(effectué au LORIA de mars à juillet 1999)

## Analyse de l'articulation entre parole et geste dans un corpus multimodal

**Auteur :** Frédéric LANDRAGIN

**Directeur de recherche :** M. Laurent ROMARY, Responsable Scientifique de l'équipe "Langue et Dialogue" du Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA).

### Résumé :

Une communication homme-machine véritablement naturelle se doit d'accepter et de comprendre les moyens naturels que possède l'homme pour communiquer, c'est-à-dire la parole et le geste venant en complément de celle-ci. Dans un système de dialogue homme-machine finalisé, les objets de l'application peuvent être désignés soit par un énoncé oral seul, soit par l'association d'un énoncé oral et d'un geste de désignation, ce qui constitue une interaction multimodale. Ce mémoire présente une étude des associations possibles de la voix et du geste spontanés, effectués sans contrainte, ainsi que de leurs traitements.

Dans l'état de l'art qui constitue la première partie de ce travail, nous nous focalisons sur le problème de la référence aux objets et nous montrons que les méthodes utilisées pour résoudre les références multimodales (c'est-à-dire pour faire le lien entre d'un côté les mots utilisés et les gestes effectués, de l'autre côté les objets de l'application) sont soit des extensions insuffisantes de méthodes utilisées classiquement dans le discours, soit des méthodes trop simplistes pour rendre compte de la complexité des énoncés multimodaux.

Dans la deuxième partie de ce travail, nous étudions un corpus multimodal selon une approche linguistique, complétant ainsi la première analyse de ce corpus effectuée par Frédéric Wolff et centrée sur le geste (thèse de doctorat soutenue en 1999). Nous identifions les informations contenues dans les énoncés oraux et nous les confrontons aux hypothèses de candidats des désignations gestuelles, en tenant compte des contextes discursif, perceptif et applicatif. Apparaît alors le problème consistant à établir les correspondances entre gestes et expressions référentielles lorsque  $p$  gestes sont effectués avec  $q$  expressions, phénomène formalisé sous le nom de référence multimodale combinée. Les informations temporelles, prosodiques et syntaxiques nécessaires à l'interprétation de ce type de référence sont explicitées, en tenant compte des spécificités de la langue parlée telles que les répétitions ou les auto-corrrections pour lesquelles il existe des équivalents dans la production du geste. Une modélisation du traitement de ces informations est alors proposée, puis une synthèse théorique de notre analyse, portant sur la prise en compte d'informations implicites pour la résolution des références multimodales. Cette synthèse permet d'aboutir à la proposition d'une classification des mécanismes de désignation et à des possibilités de prolongements de l'étude.