

Les systèmes multimodaux sont des systèmes interactifs qui autorisent l'interaction à travers des modes et des médias divers, qui permettent des interactions parallèles et qui, pour certains d'entre eux (systèmes composés) gèrent sémantiquement et temporellement des informations variées pour reconstituer des messages fractionnés à travers les modes. Toutes ces propriétés rendent nécessaires des architectures logicielles adaptées, qu'il reste encore à l'heure actuelle à définir.

En effet, il existe actuellement, encore très peu de données expérimentales qui puissent permettre de dégager des règles et des concepts stables pour la communication homme-machine multimodale. Nous avons donc adopté une approche expérimentale pour mieux cerner les problèmes spécifiques à ce type de communication. Nous avons pour cela mis au point une petite application de création d'un document administratif lié à la gestion de la scolarité. Nous avons limité notre champ d'étude à l'association de la parole écrite avec des gestes de désignation avec la souris. Ce type d'interaction met en oeuvre trois modes ou média qui nous sont familiers : la parole dans un langage quasi-naturel, les gestes de désignation et la vision.

Nous avons donc implémenté le module qui traite du langage naturel écrit, et bien sûr celui qui traite des gestes de désignation avec la souris dans le contexte de l'application.

Sur le plan technique, se pose le problème de l'intégration des diverses modalités qui participent à l'interaction : un système multimodal doit disposer des connaissances propres à chaque modalité et pouvoir faire le lien entre les informations qui contribuent au sens du message. En effet, résoudre un problème de communication multimodale ne consiste pas en la simple juxtaposition des différentes modalités mais plutôt en leur coopération au cours d'un même échange avec la machine<sup>[16]</sup>.

La solution pour laquelle nous avons opté est celle de combiner la représentation en CMR du MMI2 et celle de la vue en couches du modèle Seeheim. Le langage naturel et les gestes de désignation avec la souris sont traités seuls avant de subir un fusionnement pour compléter les informations d'un tableau d'analyse plus général.

Le langage naturel, subit les traitements lexical et syntaxique avant d'en tirer les informations pertinentes.

Nous avons ensuite étudié les cas de multimodalité synergique ou composée dans les cas où un même message est exprimé suivant les deux modalités. La vérification du comportement de notre système dans un pareil état est assez ardue ; le choix que nous avons été contraint de prendre ne semble pas être le plus adéquat. C'est pour cela qu'une intégration de la parole orale est la chose la plus urgente qu'on serait à même d'entreprendre.