

Modélisation du sens et du contexte sur la base de représentations des objets référés Frédéric Landragin — DRAFT

La modélisation de l'interprétation linguistique pose non seulement le problème de la modélisation du sens des phrases ou des énoncés, mais aussi celui de la modélisation du contexte. Quand le langage est utilisé dans une situation de dialogue incluant un espace visuel partagé entre les interlocuteurs, le contexte inclut en première approximation deux facettes complémentaires : le contexte linguistique (ou historique du dialogue) et le contexte visuel. La modélisation de l'interprétation linguistique doit ainsi reposer sur ces deux facettes. Des éléments visuels aussi bien que linguistiques peuvent en effet être à l'origine d'inférences, de présuppositions, et bien entendu de références à des objets. On peut alors envisager un modèle d'interprétation qui raisonne sur ces éléments multimodaux (modalité visuelle et modalité linguistique), plus précisément sur des représentations liées aux objets référés.

1. Représentations référentielles et représentations du discours

Corblin a peut-être été le premier à faire le constat qu'une action de référence pouvait s'effectuer dans un *domaine de référence* dénotant un groupe de référents potentiels (ou antérieurs) dans lequel le référent courant est extrait et opposé aux autres (Corblin, 1995). En suivant le même principe, les travaux de Reboul puis de Salmon-Alt se sont attachés à modéliser la résolution de la référence à l'aide de représentations — *représentations mentales* (Reboul, Moeschler, 1998) ou *domaines de référence* (Salmon-Alt, 2001) — qui structurent un certain nombre d'informations sur les référents, sur les expressions utilisées pour y référer, et sur les relations qu'ils entretiennent les uns avec les autres. A chaque fois, l'accent est mis sur les oppositions (ou différenciations) entre les éléments des partitions caractérisant ces représentations référentielles. Avec comme objectif la résolution de la référence dans le dialogue incluant un support visuel, (Landragin, 2004) a ensuite étendu ce principe à la multimodalité avec l'intégration de domaines de référence visuels aux domaines linguistiques.

De leur côté, les théories de représentation de discours, avec comme exemples la SDRT (*Segmented Discourse Representation Theory*, cf. Asher, Lascarides, 2003) et la MDRT (*Multimodal Discourse Representation Theory*, cf. Pineda, Garza, 2000), se fondent sur des représentations logiques emboîtées pour modéliser le sens et l'interprétation linguistiques. Le point d'entrée est ici un ensemble de variables d'individu désignant en particulier les objets référés. Résoudre une anaphore revient à apparier deux variables, et interpréter consiste à opérer prédications et inférences avec ces variables comme arguments. A ces principes de base, la SDRT apporte une modélisation des relations de discours et des relations entre les échanges dialogiques, tandis que la MDRT apporte une modélisation du contexte visuel.

Les deux approches peuvent ainsi se distinguer selon l'unité utilisée pour les calculs : les théories de représentation du discours raisonnent sur des variables d'individu alors que la théorie des domaines de référence raisonne sur des représentations référentielles. Par conséquent, cette dernière se révèle plus fine au niveau de la résolution de la référence (dans sa version multimodale, la variété des phénomènes pris en compte, avec par exemple la saillance visuelle ou linguistique, permet de trouver le bon référent) mais plus éloignée de la sémantique de la phrase et du dialogue.

2. Modélisation du sens et du contexte sur la base de représentations référentielles

Un enjeu consiste ainsi à rapprocher les deux approches. Or, intégrer des domaines de référence dans les structures de représentation du discours, par exemple en remplaçant les variables d'individu par des domaines, s'avère pour le moins compliqué, à l'image des MDRS de (Pineda, Garza, 2000) dont le pouvoir d'expression étendu limite les traitements computationnels envisageables. Par contre, la structure d'un domaine de référence (à base de partitions des éléments et de traits décrivant les caractéristiques — type, cardinalité — de ces éléments) autorise plus facilement l'ajout de traits supplémentaires permettant de prendre en compte la sémantique verbale et les relations de dialogue. Comme l'interprétation se fait par unification des traits, l'ajout de traits n'a que peu de conséquences sur le modèle général.

La notion de *domaine de référence sous-spécifié* est à la base de la représentation du sens d'un énoncé. Il s'agit de traduire les expressions référentielles en contraintes pour la recherche des référents, ces contraintes se formalisant à l'aide d'un domaine de référence structuré mais pas totalement instancié. Ainsi, l'expression démonstrative *ce N* se traduit par l'existence d'une partition dans laquelle le référent est focalisé, singulier et opposé à d'autres N pour justifier le contraste induit par le démonstratif (Salmon-Alt, 2001). Il s'agit maintenant d'intégrer d'autres aspects sémantiques et pragmatiques. Nous proposons ainsi de :

- intégrer la sémantique verbale via les objets concernés par le prédicat. En particulier, la catégorie sémantique du verbe a une influence sur les focalisations résultantes dans les partitions du domaine sous-spécifié : objet affecté par l'action pour un verbe de mouvement ou de changement d'état ; objet sur lequel s'appliquent les conséquences de l'événement pour un verbe décrivant un événement ; receveur pour un verbe de transfert ; etc. D'autre part, le contenu sémantique du verbe est soit formalisé à l'aide d'un trait supplémentaire, soit intégré dans le critère de différenciation des partitions du domaine sous-spécifié.

- intégrer les actes de langage et les aspects prosodiques. En particulier, l'intonation impérative dans l'exemple *la porte !* conduit à ajouter dans le domaine sous-spécifié un trait traduisant l'action requise par une telle exclamation (à fermer si la porte est ouverte, à ouvrir si la porte est fermée, sachant que ce sera le domaine correspondant au contexte visuel qui permettra de choisir la bonne interprétation).

- intégrer les présuppositions. Les domaines de référence ne rendent pas compte de cet aspect. Or la référence peut être contrainte par des présuppositions (dans *rangez la chaise*, le référent est supposé être une chaise non rangée). Une telle contrainte se formalise très facilement avec un trait venant compléter le trait « type » du référent attendu. D'autres présuppositions comme *Jean est marié* et *Jean battait sa femme* avec l'exemple classique *Jean a arrêté de battre sa femme*, se formalisent de même avec un trait *marié* dans la représentation référentielle de Jean et fut battue dans celle de sa femme.

- dans cette lignée, intégrer l'historique des actions effectuées et des énoncés prononcés, en les rattachant aux objets concernés. L'objectif est cette fois la modélisation du contexte, ce qui se réalise en stockant d'une part les domaines linguistiques avec les étapes de leur construction et de leurs mises à jour, et d'autre part les domaines visuels. Plus que deux simples piles de domaines, il s'agit d'un stock structuré dans la mesure où l'ordre et la nature des extractions sont conservés et lient entre eux les différents domaines. Cette structuration à la fois conceptuelle et temporelle constitue d'ailleurs un avantage de la théorie des domaines de référence par rapport aux approches telles que celle des domaines de quantification.

Que l'on prenne en compte uniquement les actions de référence ou l'ensemble des aspects sémantiques d'un énoncé, il semble donc possible de partir des objets référés et de conserver une modélisation relativement simple à base de domaines, de partitions, de traits, de différenciations et de focalisations. C'est un point fondamental pour deux raisons principales : d'une part parce que, le sens et le contexte d'interprétation étant modélisés de la même façon (à l'aide de domaines de référence), il est possible de les faire interagir, de les confronter ; d'autre part parce que cela permet de maintenir la simplicité des implémentations (basées sur l'unification de structures de traits) tout en prenant en compte le maximum de phénomènes interprétatifs. Il reste toutefois à noter que cette approche semble surtout réalisable pour le dialogue incluant un support visuel, c'est-à-dire dans des situations privilégiant les références.

Références

- ASHER N., LASCARIDES A. (2003) *Logics of Conversation*, Cambridge University Press.
- CORBLIN F. (1995) *Les formes de reprise dans le discours*, Presses Universitaires de Rennes.
- LANDRAGIN F. (2004) *Dialogue homme-machine multimodal*, Hermès.
- PINEDA L., GARZA G. (2000) A Model for Multimodal Reference Resolution, *Computational Linguistics* 26 (2), pp. 139-193.
- REBOUL A., MOESCHLER J. (1998) *Pragmatique du discours*, Armand Colin.
- SALMON-ALT S. (2001) *Référence et dialogue finalisé : de la linguistique à un modèle opérationnel*, Thèse de doctorat, Université Henri Poincaré de Nancy.