

Une caractérisation de la référence ostensive indirecte

Frédéric Landragin

DRAFT

Résumé :

Les mots ne désignent pas toujours les entités auxquelles renvoient leurs significations. Une « omelette au jambon » peut ainsi désigner le client d'un restaurant qui a commandé un tel plat. Ce phénomène, appelé référence indirecte, existe également lors de la production de gestes de désignation en lien avec les mots : le geste peut indiquer un objet qui n'est pas le référent de l'énoncé pris dans sa totalité. Ce phénomène, appelé ostension indirecte, met l'accent sur des problèmes fondamentaux de référence aux objets, et s'avère riche à la fois du point de vue de la théorie et de ses conséquences sur la réalisation de systèmes de compréhension automatique. Nous explorons l'hypothèse qu'un tel système rendra compte du maximum de phénomènes de référence en fonctionnant de manière ascendante, c'est-à-dire en considérant qu'un geste désigne avant tout un lieu, et que l'interprétation nécessite des décalages sémantiques ascendants partant du lieu et aboutissant au référent. Nous proposons une classification de ces décalages et des informations implicites qu'ils font intervenir. Nous donnons des arguments cognitifs et linguistiques pour appuyer notre hypothèse. Nous proposons enfin quelques pistes pour la réalisation d'un système capable de comprendre les subtilités de l'association du geste et du langage.

Title: A Characterization of Deferred Ostension

Abstract:

Words do not always refer to the entities linked to their meaning. Thus, a “ham omelette” can refer in a restaurant to the client who asked for such a dish. This phenomenon is called indirect reference and also appears when pointing gestures are produced together with words. In fact, the gesture can indicate an object which is not the referent of the whole utterance. This phenomenon is called deferred ostension and arises some fundamental problems of reference to objects. Studying it is useful from a theoretical point of view and from the point of view of the consequences on the design of automatic comprehension systems. We explore the hypothesis that such a system will treat the maximum of phenomena using a bottom-up interpretation process. We then consider that a gesture points out a place above all, and that semantic transfers are then necessary to move from the place to the referent. We propose a classification of semantic transfers and of the implicit information that intervenes. We give some cognitive and linguistic arguments to reinforce our hypothesis. We then lay the foundations for the realization of a system that is able to understand the subtle aspects of gesture and language association.

Introduction

Que ce soit en linguistique, en sciences cognitives ou en informatique, la référence aux objets faisant intervenir le geste en plus du langage est souvent considérée comme plus simple que la référence uniquement linguistique. On attend du geste – qui dans ce cadre ne peut être qu'un geste de pointage – qu'il complète une expression référentielle à laquelle aucun référent ne pourrait être attribué sans sa participation. Le geste est ainsi vu comme désignant directement le référent ciblé, et on n'imagine pas qu'il puisse introduire une ambiguïté supplémentaire. Nous voulons montrer dans cet article qu'une expression référentielle multimodale, c'est-à-dire l'association d'un geste de pointage et d'une expression référentielle linguistique, peut s'avérer beaucoup plus complexe, et peut inclure notamment des phénomènes de décalage interprétatif, de métonymie ou encore de polysémie. Le premier objectif que nous nous fixons est de décrire ces phénomènes de manière précise.

On trouve dans la littérature (Quine, 1968 ; Nunberg, 1979 ; Kaplan, 1989 ; Récanati, 1993 ; Kleiber, 1994 ; Corazza, 1995 ; Reimer, 1996 ; Powell, 1998 ; etc.) quelques analyses d'exemples où le geste désigne un objet qui n'est pas le référent. C'est ce phénomène que nous appelons référence ostensive indirecte. Les auteurs ne basent généralement pas leur modèle sur ces exemples, leur laissant un statut de marginalité. Il nous semble au contraire que c'est en s'intéressant en premier lieu à ces exemples complexes que l'on peut bâtir un modèle capable de rendre compte d'un maximum de phénomènes. Notre deuxième objectif est ainsi d'aboutir à un modèle pour l'interprétation des expressions référentielles multimodales.

Plus précisément, nous développons l'hypothèse que l'interprétation de toute expression référentielle multimodale commence par l'identification d'un lieu (celui indiqué par geste) et met ensuite en jeu des mécanismes de transfert sémantique permettant de passer de ce lieu à l'objet physique qui s'y trouve, et ainsi au référent. Cette hypothèse peut être qualifiée d'ascendante (bottom-up), dans la mesure où elle part du matériau le plus proche du phénomène physique observable et intègre petit à petit des considérations plus conceptuelles. L'ascendance s'opère par des transferts sémantiques rendus possibles par les connaissances que les deux interlocuteurs partagent, connaissances d'ordre général ou dépendant de la situation de communication. Dans le but d'identifier et de caractériser les mécanismes ascendants, nous décrivons la nature de ce terrain commun et de ces transferts sémantiques.

Il ne s'agit pas d'appliquer notre hypothèse ascendante au fonctionnement cognitif humain, car les transferts sémantiques dont il est question n'ont pas souvent de matérialisation linguistique, et car nous ne savons pas comment ils peuvent s'effectuer dans le cerveau. Nous ne pouvons donc que donner des arguments pour la plausibilité cognitive de cette hypothèse. Nous voulons également la prendre en compte dans l'élaboration d'un modèle computationnel, c'est-à-dire montrer que l'interprétation automatique d'expressions référentielles multimodales (par exemple dans un système de dialogue homme-machine) a tout intérêt à fonctionner de manière ascendante. Nous développons nos arguments cognitifs dans ce sens, et nous tirons les conséquences de notre hypothèse en termes de représentation sémantique formelle et de description des processus pragmatiques d'identification des référents dans un système de compréhension automatique.

Définitions, problèmes et méthodologie

La direction indiquée par un geste de pointage (ou encore geste de désignation, geste déictique ou geste ostensif) est souvent imprécise, et cette imprécision physique s'ajoute au décalage de point de vue entre locuteur et interlocuteur pour rendre l'identification de la cible parfois difficile. Lorsque plusieurs cibles potentielles sont possibles, seuls les termes linguistiques – eux-mêmes imprécis (déterminant, catégorie, modifieurs éventuels) – qui sont utilisés conjointement au geste permettent de lever l'ambiguïté. Lorsque les cibles potentielles se ressemblent au point que les mots du langage sont insuffisants à les différencier, l'ambiguïté peut rester et provoquer une incompréhension. L'association entre langage et geste repose donc sur des imprécisions pouvant conduire à l'échec. Cette idée se trouve dans de nombreux travaux, et amène par exemple Roberts (2002) à faire une distinction entre le geste ostensif (au sens physique) et la démonstration, qui doit garantir que l'interlocuteur va retrouver l'objet désigné.

Cette première considération montre que l'interprétation d'une expression référentielle multimodale n'est pas un problème forcément simple. Dans un même ordre d'idée, abordons maintenant le cas où le geste désigne un objet qui n'est pas le référent. Quine (1968, 194) est peut-être le premier à utiliser un terme pour ce phénomène (« deferred ostension »). Plus tard, Reimer (1996, 136) parlera de « deferred reference ». Nous utiliserons pour notre part les termes de référence ostensive indirecte, ostension indirecte ou encore ostension non coréférente, la coréférence caractérisant le fait qu'il y a identité entre l'entité du discours (le référent) et l'entité du geste (le demonstratum). On trouve cette distinction entre référent et demonstratum dans quelques travaux en philosophie du langage, avec une terminologie variable. Par exemple, Récanati (1993) parle d'indice pour le demonstratum, ainsi que de PPP (« primary pragmatic processes ») pour mettre en correspondance référent et indice. Cette mise en correspondance est appelée « referring function » par Nunberg (1979) qui utilise comme nous les termes de référent et de demonstratum. Ce dernier est parfois appelé index. Corazza (1995) distingue le référent intentionnel du référent sémantique, le premier correspondant à peu près au référent et le second au demonstratum. D'autre part, le terme de Quine a été recyclé pour des phénomènes proches, mais uniquement linguistiques. Ainsi, Nunberg (2002) parle de « deferred interpretation » ou « deference » pour des phénomènes linguistiques tels que la métonymie et la synecdoque. Pour sa part, Webber (1991) utilise le terme pour la déixis textuelle, plus particulièrement pour le pronom « *this* » quand son interprétation s'étend à tout ou partie d'une proposition.

Dans une ostension indirecte telle qu'on la considère, le décalage entre le demonstratum et le référent est forcément implicite, c'est-à-dire qu'aucune marque linguistique ne vient donner d'indice permettant son identification. Ainsi, l'association de l'expression référentielle « *j'aime sa couleur* » à un geste désignant un fauteuil rouge ne constitue pas une véritable ostension indirecte. Cet exemple fait en effet intervenir un possessif, qui, même s'il ne donne à l'interlocuteur aucune information sur le demonstratum, signale un transfert entre ce demonstratum et sa propriété de couleur. Dans « *j'aime la couleur de ce fauteuil* », le processus est encore plus explicite car l'interlocuteur a tous les éléments pour identifier le référent. L'expression référentielle « *la couleur de ce fauteuil* » comprend un complément du nom qui décrit le demonstratum, le décalage étant marqué par la préposition. Il s'agit en fait d'une expression référentielle à l'intérieur d'une autre, et aucun problème d'ostension indirecte ne se pose : on associe « *ce fauteuil* » au geste, on résout la coréférence ainsi formée, puis on prend en compte le décalage explicite pour aboutir au référent final qu'est la couleur. Le processus est le même pour les locutions prépositionnelles telles que les références

spatiales (l'ostension s'applique au site, et le référent final est la cible : « l'étagère derrière ce fauteuil », « les livres à côté de celui-ci », « l'eau dans cette bouteille », etc.), ainsi que pour certaines subordonnées relatives (dans « la chaise qui se trouvait là », l'ostension s'applique à « là » et permet d'aboutir au référent).

Nous nous intéressons dans cet article aux cas tels que l'association de « *j'aime cette couleur* » avec un geste désignant un fauteuil rouge, où le décalage entre demonstratum et référent est totalement implicite. Selon notre hypothèse, l'interlocuteur identifie tout d'abord le lieu désigné, passe ensuite de ce lieu à l'objet qui s'y trouve (le fauteuil), puis de l'objet à sa propriété de couleur qui constitue le référent final (rouge). Tenant compte de gestes vers du concret, nous nous intéressons à la référence aux objets, aux groupes d'objets, aux classes ou espèces d'objets, aux qualités ou caractéristiques d'objets, aux personnes, et aux lieux.

Rôles et fonctionnement du geste

Un geste ostensif indique une direction dans laquelle l'interlocuteur va regarder. En ce sens, tout geste désigne un point de l'espace, c'est-à-dire un lieu. Même lorsque la direction est imprécise, par exemple dans le cas de l'association de « *cette chaise* » avec un geste désignant un point éloigné d'un bon mètre de la chaise ciblée, le mécanisme d'extension à un espace incluant le référent se fait très bien. La réussite est due à la catégorisation opérée par l'expression linguistique. Si l'on ne considère que le geste, il est impossible de déterminer le demonstratum de manière sûre. Ce manque d'autonomie du geste appuie notre hypothèse de transfert sémantique ascendant : le point de départ semble bien être l'identification du lieu désigné.

Le fonctionnement contrastif du démonstratif va dans ce sens. Le composant linguistique le plus naturel pour une référence multimodale est en effet le groupe nominal ou le pronom démonstratif. Or, en association avec un geste ostensif, le démonstratif réalise forcément un contraste spatial. L'interprétation commence donc par la prise en compte de considérations spatiales. L'idée que le geste désigne forcément un lieu se retrouve d'ailleurs dans la littérature, en particulier de la part d'un certain Bennett cité par Kaplan (1989, 527) sans référence bibliographique : « Michael Bennett has proposed that only places be demonstrata and that we require an explicit or implicit common noun phrase to accompany the demonstrative, so that: 'that' [pointing at a person] becomes 'dthat' [the person who is there [pointing at a place]] ».

Une étude de corpus va nous donner quelques éléments en faveur de cette idée. Le corpus Magnét'Oz (Wolff, 1999) est un corpus multimodal obtenu suite à une expérience de type Magicien d'Oz, c'est-à-dire qu'il s'agit d'une simulation d'un dialogue homme-machine, la machine étant en fait contrôlée à distance par le magicien (qui dans notre cas est l'expérimentateur lui-même). Au cours d'une séance, le sujet est placé devant un poste informatique comprenant un microphone et un écran tactile. Une scène comportant des objets géométriques est affichée à l'écran, et le but de la séance, à savoir le rangement de ces objets dans des boîtes appropriées, est expliqué au sujet. On lui suggère de s'adresser à la machine spontanément, en désignant s'il le souhaite les objets à l'écran. Les énoncés produits ainsi que les gestes et les états successifs de la scène sont enregistrés et constitueront le corpus. Plusieurs heures de dialogue multimodal ont été obtenues. La variété des gestes produits et des caractéristiques de leur interprétation en fonction de la disposition des objets et des

expressions linguistiques a déjà fait l'objet de plusieurs analyses. Ce qui nous intéresse pour le moment, c'est de comparer les gestes effectués pour la désignation d'un objet et ceux effectués pour la désignation d'un lieu. Nous observons rapidement que les formes de trajectoires sont exactement les mêmes quand elles s'appliquent à un objet ou à un lieu : les pointages ponctuels aussi bien que les entourages sont produits de la même façon dans les deux cas. Nous observons également que les accès individuels (un geste par demonstratum) et les accès pluriels (un geste global à plusieurs demonstrata) apparaissent également dans les deux cas. De plus, les sources d'ambiguïté – en particulier l'ambiguïté de portée d'un geste, cf. l'exemple « *mets de la moquette ici* » de (Romary, 1993), où le geste de pointage porte en fait sur toute une surface – sont également comparables. Ainsi, rien ne vient dissocier le geste portant sur un objet du geste portant sur un lieu. Le test suivant vient confirmer ce point de vue : lors de la phase de spécification d'une interface multimodale dans le cadre d'un projet européen, nous avons été amené à imaginer deux paradigmes d'interaction différents pour la désignation d'un objet et pour la désignation d'un lieu. Il a vite été évident que ce principe allait totalement à l'encontre de la spontanéité de la communication homme-machine. Le mécanisme de production de geste n'est pas lié à la nature du demonstratum.

A ce stade de notre étude, nous ne pouvons pas encore valider notre hypothèse. Nous pouvons par contre soulever la question de la différence entre objet et lieu : en fin de compte, est-il nécessaire de distinguer si catégoriquement un objet et un lieu ? Comme un objet, un lieu a des limites physiques qui lui donnent consistance. Tout lieu est donc aussi objet. D'autre part, tout objet peut prendre une fonction locative, soit parce qu'il est composé de surfaces qui peuvent servir de support, soit parce qu'il peut servir de contenant. En quelque sorte, tout objet peut donc aussi être lieu. Il existe d'ailleurs des termes qui dénotent à la fois l'objet et le lieu : la *chaise*, car on peut la déplacer comme un objet ou s'asseoir dessus ; la *cuisine*, car on peut l'acheter comme un objet ou y entrer pour manger. Il existe même des termes pour lesquels s'est fait un transfert de sens : le *bureau* dans lequel on s'installe pour travailler ou sur lequel on pose ses affaires. Sans aller plus loin dans ces questions ontologiques, nous noterons que rien ne s'oppose à ce que nous considérions le geste comme désignant avant tout un lieu.

Caractérisation des mécanismes d'ostension indirecte

Notre principal argument consiste en fait en une description systématique des formes de décalage entre demonstratum et référent. Le tableau 1 récapitule de manière synthétique une vingtaine de mécanismes différents. Les exemples qui y sont donnés sont parfois inspirés par l'un des auteurs déjà cités. Ainsi, le décalage entre le chapeau (demonstratum) et l'homme qui l'a porté (référent) vient de Kleiber (1994). Le décalage entre deux personnes liées par un lien affectif est tiré d'un exemple de Powell (1998) : « Spoken while pointing at a girl child to identify her father: "He is in real estate" ». Le lien nécessaire est ici familial.

– Tableau 1 à insérer à peu près ici –

La dernière colonne du tableau montre les différents types de décalage entre demonstratum et référent, avec parfois une ou plusieurs entités intermédiaires. Dans les chaînes représentant l'aspect ascendant de l'interprétation, chaque flèche indique un transfert sémantique. La relation justifiant la flèche peut être :

(i) une relation caractérisante, comme pour le passage d'un objet vers sa couleur ou d'un humain vers sa taille ;

(ii) une relation prototypique, comme pour certains passages d'une propriété à un objet ou à une classe d'objets particulièrement susceptible de posséder cette propriété ;

(iii) une relation de similarité, comme pour certains passages d'un objet à un groupe d'objets similaires et accessibles dans le contexte ;

(iv) une relation spatiale, telle qu'on peut en voir dans les trois premiers exemples du tableau, ou encore dans les exemples faisant intervenir une relation de proximité ou de contenance (l'odeur dans la salle fermée ou l'eau dans la bouteille) ;

(v) une relation historique, comme pour certains passages d'un objet vers une personne, quand la personne et l'objet sont liés par une histoire quelconque (c'est le cas également d'un exemple de Kaplan faisant intervenir une fleur et un homme : « a background story can be provided that will make pointing at the flower a contextually appropriate, though deviant, way of pointing at a man; for example, if we are talking about great hybridizers ») ;

(vi) une relation culturelle, telle que celle nécessaire pour passer de Dieu au ciel.

L'établissement de ces relations fait appel, suivant les cas, à des traits sémantiques tels que ceux décrits par la sémantique componentielle, à des propriétés dans la base de données d'un système, à la situation d'énonciation, à l'historique de l'interaction, ou encore à des connaissances encyclopédiques sur les objets ou sur les personnes en cause. Un grand nombre d'aspects de la sémantique et de la pragmatique sont ainsi mis en jeu.

Nous remarquons que la plupart des chaînes ascendantes énumérées comportent un ou deux transferts sémantiques (une ou deux flèches). Deux exemples font intervenir trois transferts. Il s'agit d'une part de la référence à un groupe d'objets visibles à partir de la désignation gestuelle d'un seul membre de ce groupe (« *enlève ces fauteuils* » avec un geste désignant un seul des fauteuils similaires dans la scène), et d'autre part de la référence à une classe d'objets à partir de la désignation de plusieurs éléments distincts de cette classe (« *j'aime bien ces fauteuils* » avec un geste désignant deux fauteuils qui diffèrent sur certaines propriétés, par exemple la forme, mais qui sont similaires sur d'autres propriétés, par exemple la matière). Dans le premier cas, l'interlocuteur, après avoir identifié le lieu désigné et lui avoir rattaché un objet, doit extraire de cet objet une ou plusieurs qualités afin de déterminer l'ensemble de qualités nécessaire à l'identification des autres membres du groupe dont il est question. Dans le deuxième cas, l'extraction de qualités se fait sur plusieurs objets non similaires, c'est-à-dire qu'elle doit conduire à l'identification d'un ensemble minimal de qualités communes (la matière dans notre exemple), ensemble nécessaire à l'identification de la classe référée. Ces mécanismes, on le voit, sont totalement implicites, mais sont nécessaires pour l'interprétation des expressions référentielles multimodales. Sans leur prise en compte, l'association de l'expression « *ces fauteuils* » avec un geste désignant un seul fauteuil ne peut que conduire à

une incompréhension, due à l'incohérence entre le pluriel et le geste individuel. Ce type d'expression multimodale est pourtant courant, et nous semble devoir être intégré à tout modèle d'interprétation.

Dans le tableau 1 n'apparaissent pas les passages d'une personne vers un lieu, ni d'une personne vers un objet, tels que « *sa maison est un vrai château* » ou « *son bureau est toujours mal rangé* » avec un geste désignant un homme. La raison est que ces passages ne nous semblent possibles qu'avec l'utilisation d'un possessif, et ne relèvent donc pas de l'ostension indirecte.

Figures de rhétorique et autres cas particuliers

La métonymie consiste à évoquer une entité pour référer à une autre entité qui lui est liée par un rapport logique (ou rationnel). Elle peut s'accompagner d'un geste désignant l'une ou l'autre entité, comme dans « *je vais boire ce verre* » où le geste désigne un verre (supposé opaque) alors que le référent est son contenu, et dans « *ce Bordeaux est meilleur que celui-ci* » où les deux gestes désignent soit des verres soit directement les vins, mais en aucun cas la ville de Bordeaux. Dans le premier cas, nous pouvons imaginer que la résolution de la référence multimodale portant sur le verre est antérieure à l'interprétation de la métonymie et à l'identification du référent final. Cette antériorité va dans le sens de notre hypothèse ascendante. Dans les deux cas, nous noterons cependant que la nature exacte du demonstratum – verre ou vin – n'a que peu d'importance pour l'identification du référent, en tout cas moins que l'expression linguistique. Le verbe « *boire* » et le démonstratif « *ce* » devant « *Bordeaux* » suffisent en effet à fixer la nature du référent. Autrement dit, les mots imposent un mécanisme descendant (du concept vers le demonstratum) qui vient à l'encontre de notre hypothèse ascendante. Nous constatons la même chose lorsque le demonstratum et le référent ont une relation partie-tout, par exemple dans « *cette voile à l'horizon* » avec un geste désignant un bateau, ou plus simplement « *ce bateau* » avec un geste désignant la voile d'un bateau : dans les deux cas, le référent intentionnel correspond au tout et non à la partie. Même si d'un point de vue cognitif le mécanisme ascendant semble remis en question, nous considérons que ce n'est pas forcément le cas d'un modèle d'interprétation. En effet, les exemples de « *boire ce verre* » et de « *ce Bordeaux* » peuvent se décomposer selon la chaîne ascendante suivante : passage du lieu désigné au verre de vin, puis, avec la prise en compte de « *boire* » ou de « *Bordeaux* », passage du verre au vin lui-même. Pour « *cette voile* », la prise en compte de la métonymie (plus précisément de la synecdoque) arrête le processus à l'identification du bateau, donc soit après un passage direct du lieu désigné au bateau qui s'y trouve, soit après un passage du lieu à la voile, passage auquel s'ajoute une extension de la voile au bateau.

En ce qui concerne la polysémie, nous avons déjà évoqué le cas du mot *bureau*. Un autre exemple bien connu est celui du mot *livre*, avec les deux interprétations qu'il peut entraîner, celle liée à l'exemplaire physique et celle liée au contenu. Ainsi, lorsque l'énoncé « *ne lis pas ce livre* » est associé à un geste désignant un livre, le demonstratum est forcément le livre au sens de l'exemplaire, alors que le référent, c'est-à-dire le livre qui peut être lu, est forcément le livre au sens du contenu. Un mécanisme ascendant peut parfaitement rendre compte de cet exemple : passage du lieu désigné au livre-exemplaire qui s'y trouve, puis passage du livre-exemplaire au livre-contenu qui constitue le référent. Dans « *ce livre est abîmé* » avec le même geste, ce dernier passage n'est bien entendu pas effectué. Le cas de « *ce livre est sur*

ma table de chevet » est plus complexe. Le référent est en effet un livre-exemplaire qui n'est pas le même exemplaire que le livre désigné, le lien entre ces deux exemplaires se faisant par l'intermédiaire du livre-contenu. Ici encore, nos mécanismes de passage permettent de matérialiser les étapes de l'interprétation.

Dans le cas particulier d'un référent immatériel, par exemple la référence ostensive à un trou comme dans l'association de « *il faut boucher ce trou* » à un geste désignant le milieu d'un large trou circulaire dans un mur (donc désignant du vide), nous ne pouvons pas considérer que le *demonstratum* est le trou lui-même, puisqu'un trou est immatériel et se définit par ses bords (qui justement ne sont pas désignés par le geste). Le *demonstratum* est donc forcément le lieu où se trouve le trou, et la référence ostensive est bien indirecte : le lieu désigné est étendu spatialement jusqu'à atteindre les bords du trou, donc jusqu'à considérer l'objet, dans notre cas le mur, qui contient le trou et lui donne existence. Le fonctionnement ascendant rend donc bien compte de ce cas particulier.

Dans le cas d'absence physique du référent, comme dans l'association de l'expression référentielle « *la chaise* » à un geste désignant un bureau sans chaise, le geste indique clairement un lieu. Le fonctionnement ascendant partant du lieu vers l'objet rend bien compte de l'interprétation, que celle-ci soit « *la chaise qui était ici et que je visualise encore* », « *la chaise qui devrait être ici* » ou encore « *la chaise que j'imagine à cette place* ». La même remarque est valable lorsque le référent est caché par un objet perturbateur, comme dans « *ce char* » avec un geste désignant un pont sous lequel un char est caché. On part du lieu, on passe au pont, puis au char dont on sait ou dont on accepte le fait qu'il est caché.

Enfin, nous noterons un dernier cas particulier de référence ostensive indirecte, celui de la référence à un type par l'intermédiaire d'un geste désignant une instance de ce type. C'est le cas de « *ce fauteuil* » avec un geste désignant un fauteuil dans une scène qui en regroupe une certaine variété, l'expression linguistique s'interprétant comme « *ce type de fauteuil* ». Le problème avec cet exemple est qu'un type, par définition, ne doit pas être instancié. Or c'est *a priori* le rôle que prend le geste. Nous considérons qu'après le premier passage du lieu désigné au fauteuil qui s'y trouve, l'interprétation consiste en l'abstraction du fauteuil en un type. Ce mécanisme constitue un exemple de plus de décalage interprétatif ascendant.

Analyse de quelques reprises linguistiques

Des tests d'acceptabilité linguistique permettent d'avancer quelques arguments en faveur de l'hypothèse d'existence de décalages interprétatifs ascendants. L'étude des contraintes que pose une référence ostensive vis-à-vis des énoncés ultérieurs est particulièrement intéressante. Comme premier exemple, considérons les possibilités de continuation du dialogue en termes de marqueurs interrogatifs. Après l'énoncé « *tu dois regarder ce tableau* » associé à un geste désignant une peinture dans un musée, on peut, en cas d'impossibilité à résoudre la référence, poser une question aussi bien sur le lieu que sur l'objet. Des exemples de questions portant sur le lieu sont : « *où ça ?* », « *où dois-je regarder ?* » ; des exemples de questions portant sur l'objet sont : « *lequel ?* », « *que dois-je regarder ?* ». Ceci illustre bien l'existence d'un lieu en deçà de toute ostension vers un objet.

Suivons le même principe avec l'énoncé « *tu dois regarder là-bas* » associé à un geste similaire. En cas d'impossibilité à identifier le référent, on peut poser une question sur le lieu

(« où ça ? », « où dois-je regarder ? »), mais beaucoup plus difficilement sur l'objet (*« lequel ? », *« que dois-je regarder ? »). Ceci illustre qu'aucun objet n'intervient dans l'identification d'un lieu, et notre hypothèse ascendante en est ainsi appuyée.

Comme deuxième exemple, considérons le fonctionnement des marqueurs déictiques « -ci » et « -là ». Selon les grammaires du français (Riegel *et al.*, 1994 ; Grevisse, 1997), ces marqueurs ne peuvent apparaître qu'avec les démonstratifs (« celui-ci », « ce fauteuil-là »), le marqueur « -ci » dénotant une plus grande proximité que le marqueur « -là », proximité sur un plan linguistique ou spatial (lorsqu'un geste ostensif est produit). Cette prise en compte d'une proximité spatiale nous semble intéressante pour notre problème, et nous a conduit à procéder à une étude de corpus. Nous avons ainsi repris le corpus Magnét'Oz (Wolff, 1999) pour y analyser en détail les occurrences de marqueurs déictiques. Comme la scène visuelle représente des formes géométriques disposées sur un plan, donc en deux dimensions, la notion de proximité spatiale n'intervient pas lors de la production d'un geste sur l'écran tactile. Nous ne pouvons donc rien conclure sur la différence purement spatiale entre « -ci » et « -là ».

Par contre, nous remarquons très rapidement que l'un et l'autre de ces marqueurs apparaissent fréquemment avec un geste associé à un groupe nominal défini : « l'objet-ci », « la boîte-là », « la forme-ci », « la forme-là ». Ce corpus a été enregistré en Lorraine et il est vraisemblable qu'il s'agisse d'une particularité régionale. Néanmoins, la présence de marqueurs déictiques aussi bien avec les définis qu'avec les démonstratifs nous incite à nous poser quelques questions relatives à notre problème d'interprétation ascendante. L'expression référentielle multimodale associant un geste ostensif à « le N-ci » où N est un substantif, fait penser soit à « ce N-ci », soit à « le N ici ». Le premier cas est la forme renforcée classique, et le second, qui s'interprète comme « le N qui se trouve ici », correspond à un geste vers un lieu, geste associé à un déictique pur et auquel s'ajoute un passage du lieu à l'objet, avec la description définie comme seule marque de décalage.

Avec un point de vue diachronique, nous pouvons considérer deux hypothèses : d'une part celle de l'altération lexicale qui, partant de « ce N-ci », a conduit à « le N-ci » tout en gardant le caractère démonstratif (premier cas précédemment cité) ; et d'autre part celle de l'altération co-articulatoire qui, partant de « le N ici », a conduit à la disparition du premier *i* de « ici » et à un groupe nominal défini (deuxième cas). Le problème revient donc à déterminer si « le N-ci » et « le N-là » se comportent comme des démonstratifs ou comme des définis.

Dans le sens de la première hypothèse, le corpus Magnét'Oz montre que les gestes effectués avec une expression telle que « l'objet-là » sont les mêmes qu'avec « cet objet » ou « cet objet-là », que ce soit au niveau de la forme des trajectoires, du type d'accès aux démonstrata (individuel ou pluriel, comme nous avons déjà eu l'occasion d'en discuter), et des sources possibles d'ambiguïté. Le corpus présente également quelques cas de dissociation entre le marqueur déictique attaché au groupe nominal et la désignation autonome d'un lieu, comme dans « la boîte-là, là-bas » avec un geste synchronisé à « là-bas ». Cet exemple, dont la prosodie nous confirme qu'il ne s'agit ni d'une reprise ni d'une réparation, tend à considérer le premier « -là » comme un marqueur déictique, et non comme un déictique pur tel que le conçoit la deuxième hypothèse (ce rôle étant pris ici par « là-bas »). Le test le plus important est cependant la comparaison entre d'un côté le fonctionnement de « le N-ci », et de l'autre côté celui de « ce N-ci », « celui-ci » et « ceci », particulièrement en termes de reprises.

Comparons ainsi :

- (i) « *la boîte-ci et cette boîte-là* » avec « *cette boîte-ci et cette boîte-là* » ;
- (ii) « *la boîte-ci et celle-là* » avec « *cette boîte-ci et celle-là* » ;
- (iii) « *la boîte-ci et cette autre-là* » avec « *cette boîte-ci et cette autre-là* » ;
- (iv) « *la boîte-ci et ce coffre-là* » avec « *cette boîte-ci et ce coffre-là* ».

Pour chacune de ces paires, l'acceptabilité des deux expressions n'est jamais vraiment la même. Bien qu'il soit difficile d'affirmer que telle ou telle expression est clairement inacceptable, nous remarquons aisément que les deux formes « *le N-ci* » et « *ce N-ci* » ont du mal à cohabiter. La première hypothèse semble donc assez aléatoire.

Dans le sens de la deuxième hypothèse, qui suppose que la forme « *le N-ci* » fonctionne comme « *le N qui se trouve ici* », le corpus Magnét'Oz fournit des arguments plus convaincants. Il s'agit tout d'abord de la fréquente synchronisation temporelle du geste ostensif avec le déictique pur « *ici* » ou « *là* » : même si des écarts temporels sont parfois perceptibles, la synchronisation sur le déictique pur est significativement plus fréquente que la synchronisation sur l'ensemble du groupe nominal. Il s'agit également de la présence de certaines formes qui tendent à séparer « *le N* » de « *ici* », par exemple « *les trois là* », et « *la même boîte, là aussi* », ainsi que « *la boîte ici* » qui est très proche du fonctionnement que nous attendons. Il est normal que ces formes apparaissent, car l'explication diachronique implique que nous sommes dans une phase de changement, et autorise donc la coexistence des deux formes « *le N-ci* » et « *le N ici* ». Le corpus présente même les cas suivants : « *les formes claires ci* » et « *l'ensemble des trois formes ci* », où le dernier mot prononcé est bien « *ci* » et non « *ici* ».

De même que précédemment, le test le plus important reste la comparaison entre les fonctionnements de « *le N-ci* » et de « *le N [qui se trouve] ici* » en termes de reprises. Comparons ainsi :

- (i) « *la boîte ci et la boîte là* » avec « *la boîte qui se trouve ici et la boîte là* » ;
- (ii) « *la boîte ci et la boîte qui se trouve là* » avec « *la boîte qui se trouve ici et la boîte qui se trouve là* » ;
- (iii) « *la boîte ci et l'autre là* » avec « *la boîte qui se trouve ici et l'autre là* » ;
- (iv) « *la boîte ci et l'autre qui se trouve là* » avec « *la boîte qui se trouve ici et l'autre qui se trouve là* » ;
- (v) « *la boîte ci et le coffre là* » avec « *la boîte qui se trouve ici et le coffre là* » ;
- (vi) « *la boîte ci et le coffre qui se trouve là* » avec « *la boîte qui se trouve ici et le coffre qui se trouve là* ».

Pour chacune de ces paires, nous remarquons que l'acceptabilité des deux expressions est tout à fait comparable. Nous penchons donc bien pour la deuxième hypothèse. Nous en concluons que les gestes ostensifs sont associés à « *-ci* » et à « *ici* », et désignent avant tout des lieux. En prenant l'usage du déictique pur « *ici* », « *-ci* » dans « *cet objet-ci* » ne constitue plus un

indice pour la désignation d'un objet, mais devient un indice pour la désignation d'un lieu et pour un passage de ce lieu à un objet. Nous retrouvons ainsi le mécanisme d'interprétation ascendante.

Conséquences pour un traitement automatique

D'une manière générale, les quelques travaux dans les domaines de la linguistique computationnelle et du dialogue homme-machine qui s'intéressent à l'ostension indirecte ne l'intègrent qu'à l'aide d'heuristiques, c'est-à-dire en rattrapage d'un algorithme principal suite à des résultats non satisfaisants. Ces heuristiques, qui consistent à remettre en question l'ensemble des *demonstrata* identifiés pour aboutir aux référents de l'expression multimodale, sont de deux sortes. Il s'agit premièrement d'heuristiques basées sur des critères perceptifs, ayant pour but de compenser l'imprécision du geste en l'étendant à un sous-espace visuel immédiat. Wolff (1999) exploite ainsi les critères de la Théorie de la Gestalt pour structurer l'espace visuel en groupes perceptifs, et procède à l'interprétation de trajectoires gestuelles en étendant éventuellement l'ensemble des *demonstrata* identifiés au groupe perceptif qui les contient. Il s'agit deuxièmement d'heuristiques basées sur des critères linguistiques, ayant pour but de ré-interpréter le geste lorsque l'identification du référent ne peut aboutir. Ces critères linguistiques peuvent en particulier correspondre à une prise en compte du prédicat de l'énoncé et pas seulement de l'expression référentielle. C'est de cette façon qu'est interprété l'exemple de Romary (1993) « *mets de la moquette ici* ». Nous nous sommes pour notre part intéressé dans (Landragin, 2004 : 136) à l'exemple de corpus suivant : « *les formes les plus claires* » associé à un geste désignant des formes parmi d'autres, les formes désignées n'étant les plus claires que dans un sous-espace visuel et non dans la scène complète. La prise en compte du superlatif nécessite ainsi une ré-interprétation du geste, qui n'aurait pas posé problème si l'expression s'était limitée à « *les formes claires* » ou « *ces formes claires* ». D'autre part, un des aspects de l'ostension indirecte est la référence à des propriétés (ou qualités). Klipple et Gurney (2002) s'intéressent à ce problème peu étudié et montrent que tout modèle de résolution de la référence peut *a priori* être étendu pour traiter de telles références.

La présentation d'un modèle computationnel complet de la référence multimodale, qui soit basé non pas sur des heuristiques de rattrapage mais sur un algorithme principal exhaustif, ferait l'objet d'un article à part entière et n'a pas sa place ici. Nous préférons nous focaliser sur les conséquences lexicales, syntaxiques, sémantiques et pragmatiques de la prise en compte des phénomènes d'ostension indirecte. Nous voulons que ces considérations constituent des pistes concrètes pour aboutir à un modèle de compréhension automatique. Ainsi, au niveau le plus bas d'un tel système se trouve la définition du lexique. Compte tenu de la section précédente, nous considérons que, lorsqu'un geste ostensif est considéré, « *ci* » et « *ici* » constituent les deux réalisations acoustiques possibles pour le déictique pur correspondant, et que les marqueurs déictiques n'existent plus, même dans « *celui-ci* » et « *celui-là* » qui deviennent « *celui [qui se trouve] ici* » et « *celui [qui se trouve] là* ». Les conséquences syntaxiques sont alors principalement liées à la séparation entre le groupe nominal et le déictique pur. Rendre compte du [qui se trouve] implicite doit se faire à l'aide d'une dépendance ou d'un lien supplémentaire dans l'arbre syntaxique, pour que le déictique pur ne soit pas considéré comme un complément circonstanciel de lieu attaché au prédicat, mais bien comme un modifieur particulier du groupe nominal ou du pronom.

En ce qui concerne l'analyse sémantique, les *demonstrata* et les référents doivent être distingués par des variables différentes. Les liens entre ces variables sont réalisés grâce à des primitives particulières, rendant compte des différents types de passage des premiers aux seconds. Quel que soit le formalisme adopté pour l'analyse, qu'il soit à base de structures de représentation du discours, de formes logiques ou de simples structures de traits, il est possible de matérialiser ces primitives en question. Dans le cas de formes logiques, on peut imaginer que l'expression « *cette couleur* » avec un geste désignant un objet peut, d'une manière très simplifiée car tous les aspects du décalage sémantique ne sont pas intégrés ici, se représenter par : « cherchez x et y tels que (*demonstrata* (x) et référent (y) et couleur (y) et accessible_à_partir_de (y, x)) ». Dans le cas de structures de traits, un trait supplémentaire correspondant à « accessible_à_partir_de (y, x) » peut être prévu dans le même but.

Un mécanisme pragmatique s'appuie sur ces formules ou structures pour y intégrer des informations contextuelles et aboutir à une interprétation de l'énoncé qui soit pertinente par rapport au dialogue en cours. Selon notre approche, les informations pragmatiques doivent être prises en compte de manière ascendante, en commençant par les plus concrètes (celles liées au lieu de désignation), et en intégrant petit à petit les informations les plus abstraites telles que des connaissances encyclopédiques. Le principe de l'algorithme est de tester successivement les six types de transfert sémantique que nous avons évoqués précédemment, chaque interprétation qui aboutit étant retenue. Cette méthode permet d'identifier des ambiguïtés qu'un système classique ignorerait, par exemple dans le cas de « *ces fauteuils* » avec un geste désignant un seul fauteuil, qui, dans un système de dialogue homme-machine, serait traité soit comme une erreur de l'utilisateur, soit comme une erreur du système (l'erreur de reconnaissance de la parole étant souvent avancée dans ce type de situation). Un autre avantage de cette méthode consistant à gérer les interprétations selon une échelle graduée est la possibilité de remettre en cause une interprétation lorsque celle-ci s'avère inadéquate, par exemple lorsque l'utilisateur fait comprendre au système que celui-ci s'est trompé. Même l'extension du lieu vers l'objet peut être remise en cause, ce qui permet de reconsidérer *a posteriori* un geste très imprécis. Notre méthode présente cependant quelques inconvénients. Il peut en particulier s'avérer dangereux d'envisager systématiquement une interprétation incluant une ostension indirecte. Ce phénomène n'est pas si fréquent et peut être supposé à tort dans des cas d'ostension directe. Nous considérons ainsi que lorsqu'une interprétation directe et une interprétation indirecte de l'ostension sont simultanément possibles, nous privilégions l'interprétation directe. C'est le cas de l'exemple le plus simple qui soit d'expression référentielle multimodale, à savoir l'association de « *cet objet* » avec un geste désignant un objet : l'interprétation directe est bien entendu la référence directe à l'objet désigné, et l'interprétation indirecte correspond à la référence générique « *ce type d'objet* ». Dans le cadre d'un dialogue portant sur des objets concrets, il est clair que la première interprétation doit être favorisée.

Conclusion

Nous avons avancé un certain nombre d'arguments en faveur d'un mécanisme ascendant pour l'interprétation de la référence ostensive indirecte : premièrement avec la constatation que la production d'un geste ostensif ne dépend pas de la nature du *demonstratum* et que sa compréhension commence par l'identification d'un lieu ; deuxièmement en montrant la possibilité d'effectuer une caractérisation ascendante systématique des décalages sémantiques intervenant pour chacune des formes possibles de l'ostension indirecte, y compris quelques

cas extrêmes ; et troisièmement en montrant que certains mécanismes de décalage peuvent être rendus flagrants en testant la possibilité de reprises linguistiques ultérieures à l'ostension. Nous avons proposé quelques pistes pour un traitement automatique fin de l'ostension indirecte, et montré comment leur prise en compte dans tout modèle computationnel de la référence s'avère essentielle.

Il nous semble d'autre part intéressant de remarquer à quel point un phénomène souvent considéré comme marginal illustre en fait de nombreux aspects de la référence, et met l'accent sur des problèmes fondamentaux. Qu'elle s'effectue entre humains, ou entre un humain et une machine, la communication spontanée est par essence multimodale. Il nous semble par conséquent important d'analyser les phénomènes liés à l'association du geste et du langage. La rareté des corpus multimodaux ainsi que les difficultés techniques liées à la constitution et à l'exploitation de ces corpus, de même que celles liées à la réalisation d'un système de compréhension multimodale, ne devraient pas entraver ces analyses.

Enfin, nous noterons que dans les exemples donnés et d'une manière générale nous nous sommes focalisé sur les expressions référentielles et nous avons quelque peu minimisé le rôle du prédicat. Or les différences entre les exemples « *ce livre est abîmé* », « *ne lis pas ce livre* », « *ce livre est sur ma table de chevet* » ou encore « *j'ai déjà ce livre* » se clarifient surtout dès que l'on prend en compte à la suite de Nunberg (1995) les transferts sémantiques qui opèrent au niveau du prédicat. Dans la dernière expression, le prédicat permet par exemple d'associer un livre-exemplaire au fait de posséder, et permet ainsi de faire le dernier passage du livre-contenu au livre-exemplaire. La prise en compte des caractéristiques sémantiques du prédicat dans le processus d'interprétation ascendante constitue ainsi la principale perspective envisagée pour ce travail.

Bibliographie

- Caplan, B. (2002), "Quotation and Demonstration", *Philosophical Studies*, 111(1): 69-80.
- Corazza, E. (1995), *Référence, contexte et attitudes*, Montréal/Paris : Bellarmin/Vrin.
- Corblin, F. (1987), *Indéfini, défini et démonstratif*, Genève : Droz.
- Grevisse, M. (1997), *Le bon usage*, édition refondue par André Goosse, Paris : Duculot.
- Kaplan, D. (1989), "Demonstratives", in Almog, J., Perry, J. et Wettstein, H. (éd.) (1989), *Themes from Kaplan*, New York : Oxford University Press, 482-563.
- Kleiber, G. (1993), "L'espace d'ici : sur la pragma-sémantique des adverbes spatiaux. Le cas d'il fait chaud ici", *Cahiers de linguistique française*, 14: 85-104.
- Kleiber, G. (1994), "Y a-t-il un IL ostensif ?", in *Anaphores et pronoms*, Louvain-La-Neuve : Duculot, 125-141.
- Klippel, E. et Gurney, J. (2002), "Some Observations on Deixis to Properties", in Van Deemter, K. et Kibble, R. (éd.) (2002), *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, Stanford : CSLI Publications, 355-390.

- Lakoff, G. et Johnson, M. (1980), "Metaphors we Live by", Chicago : University of Chicago Press.
- Landragin, F. (2004), Dialogue homme-machine multimodal, Paris : Hermès.
- Nunberg, G. (1979), "The Non-Uniqueness of Semantic Solutions: Polysemy", *Linguistics and Philosophy*, 3(2): 143-184.
- Nunberg, G. (1995), "Transfers of Meaning", *Journal of Semantics*, 12:109-132.
- Nunberg, G. (2002), "The Pragmatics of Deferred Interpretation", in Horn, L. et Ward, G. (éd.) (2002), *Handbook of Pragmatics*, Oxford : Blackwell.
- Powell, G. (1998), "The Deferred Interpretation of Indexicals and Proper Names", *UCL Working Papers in Linguistics*, 10:143-172.
- Quine, W.v.O. (1960), *Word and Object*, Cambridge : MIT Press.
- Quine, W.v.O. (1968), "Ontological Relativity", *Journal of Philosophy*, 65(7): 185-212.
- Quine, W.v.O. (1971), "The Inscrutability of Reference", in Steinberg, D. et Jakobovits, L. (éd.) (1971), *Semantics*, Cambridge : Cambridge University Press, 142-154.
- Récanati, F. (1993), *Direct Reference. From Language to Thought*, Oxford : Blackwell.
- Reimer, M. (1991), "Demonstratives, Demonstrations, and Demonstrata", *Philosophical Studies*, 63(2): 187-202.
- Reimer, M. (1996), "Quotation Marks: Demonstratives or Demonstrations?", *Analysis*, 56(3): 131-142.
- Riegel, M., Pellat, J.-C. et Rioul, R. (1994), *Grammaire méthodique du français*, Paris : Presses Universitaires de France.
- Roberts, C. (2002), "Demonstratives as Definites", in Van Deemter, K. et Kibble, R. (éd.) (2002), *Information Sharing: Reference and Presupposition in Language Generation and Interpretation*, Stanford : CSLI Publications, 89-136.
- Romary, L. (1993), "L'interprétation de ici dans des énoncés de positionnement", in Vivier, J. (éd.) (1993), *Le dialogue homme-robot en langage naturel : problèmes psychologiques*, Caen : Presses Universitaires de Caen.
- Webber, B. (1991), "Structure and Ostension in the Interpretation of Discourse Deixis", *Natural Language and Cognitive Processes*, 6(2): 107-135.
- Wolff, F. (1999), *Analyse contextuelle des gestes de désignation en dialogue homme-machine*, Thèse de doctorat, Université Henri Poincaré de Nancy.

Tableau 1 : Synthèse des différentes formes de l'ostension indirecte

réfèrent	demonstratum	exemple	mécanisme
lieu	lieu	« mets une prise ici » + pointage vers un mur	lieu → lieu (vers la bonne hauteur)
	lieu	« mets de la moquette ici » + pointage vers le sol	lieu → lieu (vers un lieu plus large)
	lieu	« plante un clou ici » + pointage vers un mur	lieu → lieu (vers un lieu plus précis)
	objet	« assieds-toi ici » + pointage vers une chaise	lieu → objet → lieu
propriété	lieu	« j'aime cette couleur » + pointage vers le ciel « cette odeur m'incommode » + salle fermée	lieu → propriété
	objet(s)	« j'aime cette couleur » + objet(s) « j'aime ce confort » + fauteuil(s)	lieu → objet(s) → propriété
	personne(s)	« je ne veux pas faire cette taille » + personne(s)	lieu → personne(s) → propriété
personne	lieu	« le chef » + un bureau vide « Dieu » + le ciel	lieu → personne
	personne	« cet homme » + un homme	lieu → personne
	personne	« il a de la chance » + une jolie fille	lieu → personne → personne
	objet(s)	« cet homme a une grosse tête » + un chapeau	lieu → objet(s) → personne
objet	lieu	« où est la fenêtre ? » + bureau encombré	lieu → objet
	objet	« cet objet » + un objet	lieu → objet
	objet(s)	« cette eau » + bouteille(s) remplie(s) d'eau	lieu → objet(s) → objet
	objets	« le livre » + amas de papiers avec un livre	lieu → objets → objet
objets	objets	« ces objets » + un amas d'objets	lieu → objets
	objet	« ces 2 livres » + un livre (avec un autre pas loin)	lieu → objet → objets
	objets	« ces 3 livres » + 2 livres (avec un autre pas loin)	lieu → objets → objets
	objet	« enlève ces fauteuils » + un fauteuil	lieu → objet → propriété(s) → objets
espèce	objet	« j'aime bien ces fauteuils » + un fauteuil	lieu → objet → espèce
	objets	« j'aime bien ces fauteuils » + fauteuils identiques	lieu → objets → espèce
	objets	« j'aime bien ces fauteuils » + fauteuils différents	lieu → objets → propriété(s) → espèce