



HAL
open science

Anaphore, continuité du discours et dialogue homme-machine

Laurent Romary

► **To cite this version:**

Laurent Romary. Anaphore, continuité du discours et dialogue homme-machine. *Anaphore et Deixis*, 1991, 4. hal-00521616

HAL Id: hal-00521616

<https://hal.science/hal-00521616>

Submitted on 15 Dec 2020

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.



Distributed under a Creative Commons Attribution 4.0 International License

Anaphore, continuité du discours et dialogue homme-machine.

Laurent ROMARY
Centre de Recherche en Informatique de Nancy
CRIN-INRIA Lorraine
B.P. 239, 54506 Vandœuvre Lès Nancy
romary@loria.fr
tél. : 83 91 23 88
télécopie : 83 41 30 79

0. Introduction.

La mise en œuvre d'un système de dialogue homme-machine pose un certain nombre de problèmes qui ne se réduisent pas à des considérations pratiques. En effet, la nécessité d'analyser puis de comprendre des énoncés en langage naturel font qu'il est bien souvent indispensable de se rapprocher de certaines problématiques d'autres disciplines telles que la linguistique ou la psychologie cognitive, ce qui justifie en particulier ce sentiment commun d'appartenir à la même communauté des sciences cognitives. Une des difficultés rencontrées dans ce cadre est le suivi des informations fournies dans les énoncés successifs d'un locuteur, afin de fournir à tout moment des réponses qui soient pertinentes en fonction du contexte de dialogue. Ce problème de compréhension et de construction de la continuité d'un discours s'observe plus particulièrement quand il s'agit de traiter les références - qui sont par essence principalement des co-références - ce qui nous rapproche de l'anaphore, dont des exemples apparaîtront tout au long de cette présentation. Nous allons essayer ici de dresser un bilan partiel des solutions qui peuvent s'offrir à nous pour traiter ce genre de problème, en commençant par notre propre expérience acquise dans la conception d'un certain nombre de systèmes plus ou moins opérationnels. Puis, nous essayerons de dégager quelques grandes lignes à partir de solutions proposées dans la littérature en nous limitant essentiellement aux modèles dédiés au traitement du discours en langage naturel. Enfin, nous envisagerons un cadre de représentation des connaissances qui permet de mieux appréhender les informations dont il faut tenir compte pour traiter de façon adéquate la référence.

1. présentation de quelques attitudes vis à vis de l'anaphore dans le dialogue homme-machine.

Dans le cadre des activités de notre équipe sur le dialogue homme-machine, nous développons un certain nombre de systèmes de gestion de dialogues de natures diverses. Soit le domaine visé est particulièrement réduit - l'objectif est alors de valider en situation réelle nos connaissances sur le dialogue - soit les objectifs visés sont plus ambitieux, et ce type de système est vu comme un laboratoire où de nouvelles idées sont mises en œuvre. Nous allons voir comment est traité le problème de la référence dans deux systèmes répondant respectivement aux deux catégories mentionnées ci-dessus.

1.1 Le système Partner.

Le système Partner [Morin87, Carbonell89] réalise de façon paramétrable une interface de dialogue avec des tâches simples, dans des contextes pragmatiques particulièrement bien ciblés. Il a été testé sur deux applications, à savoir :

- la gestion d'une messagerie électronique;
- la commande d'une pince simulée de robot prenant des cubes.

Le traitement de l'anaphore dans ce système repose sur une hypothèse particulièrement forte puisqu'il est admis que toute référence incomplète pourra être considérée comme une ellipse du paramètre correspondant dans la commande courante. Le fonctionnement de l'ensemble est alors relativement simple, à partir d'une grammaire sémantique de l'univers de la tâche, une vérification casuelle est effectuée pour les différents paramètres présents dans l'énoncé, et si l'un des paramètres vient à manquer, on applique une stratégie de parcours de l'historique qui consiste à conserver le candidat le plus récent satisfaisant les contraintes casuelles courantes. Cette stratégie 'du plus récent élément', pour fruste qu'elle soit, peut être retrouvée dans de nombreux systèmes ayant à traiter des problèmes de co-référence.

Nous pouvons maintenant observer un exemple de dialogue entre un usager et le système Partner où une telle stratégie est adoptée :

Sujet1 : Est-ce que j'ai reçu un message de Durand ?	Partner1 : Oui
Sujet2 : Et de Schmitt ? (<i>Ellipse</i>)	Partner2 : Non, et maintenant ?
Sujet3 : Je voudrais écouter le message de Durand	Partner3 : Voici ...
Sujet4 : Détruis-le. (<i>Anaphore</i>)	

Dans les deux cas (ellipse et anaphore) le mécanisme est le même, mais le domaine est suffisamment réduit pour qu'il n'y ait pas d'ambiguïté possible quant au choix du bon paramètre (ce paramètre pouvant être d'ailleurs une action, comme le montre le premier exemple).

Un premier bilan de cette expérience montre que le schéma simpliste de traitement de l'anaphore va de pair avec un modèle du dialogue limité à des couples d'énoncés et ne s'étendant donc pas au delà d'échanges portant sur une question et sa réponse ou un ordre et son exécution. Cette réduction de la portée du dialogue traité limite d'autant plus les risques d'ambiguïté lors du traitement des co-références. D'autre part on constate que l'on ne peut se passer d'un historique minimum du dialogue pour enregistrer une représentation - en général sémantique - des énoncés successifs du locuteur.

1.2 Le système Dial.

L'architecture du système Dial - dont le domaine d'action est celui du renseignement administratif - a été conçue au milieu des années 80 comme une plateforme de test pour l'étude de différents modèles linguistiques liés au lexique, la syntaxe, la sémantique et le dialogue. C'est ce qui explique le souci de modularité visible dans la figure 1.1 où chaque fonction du système est isolée au sein d'une unité échangeant un certain nombre d'informations avec ses voisins. Dans ce cadre, un module tel que *Synsem*, dédié au traitement syntaxico-sémantique a une portée d'analyse qui se limite à un seul énoncé et ne peut donc traiter de problème d'anaphores ou de coréférence hormis quelques cas intraphrastiques courant. Le contrôle des échanges entre l'utilisateur et la machine s'effectue donc essentiellement au niveau du module de dialogue qui prend en entrée des structures sémantiques de haut-niveau (représentation à base de grammaires de cas) pour les interpréter en fonction d'un historique propre de dialogue où sont enregistrées les différentes entités mentionnées dans les énoncés précédents.

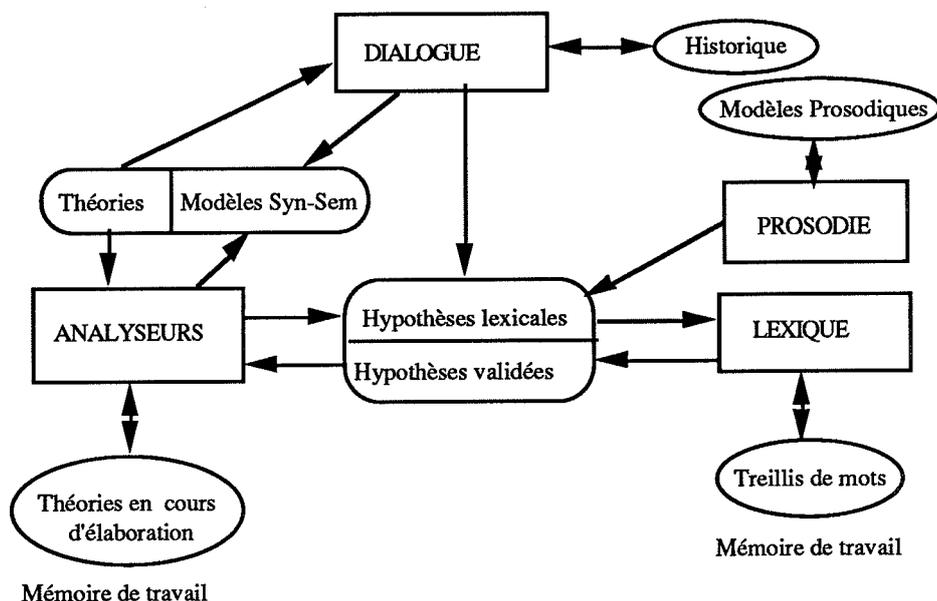


Figure 1.1 : Architecture du système Dial.

La résolution des co-références s'effectue sous la forme d'un système à base de règles établissant pour un contexte donné quel lien peut ou doit être établi avec les énoncés précédents. Ainsi une réponse à une question telle que "Où habitez vous ?" doit posséder des attributs de type lieu et quelque chose comme "Nancy" sera accepté. Ces contraintes sémantiques s'appuient sur des représentations conceptuelles (prédicats du premier ordre) qui forment la base de la structure de l'historique. Le résultat de cette analyse se présente sous la forme d'une recombinaison des objets du discours sous des étiquettes communes telle qu'on peut le voir dans la figure 1.2.

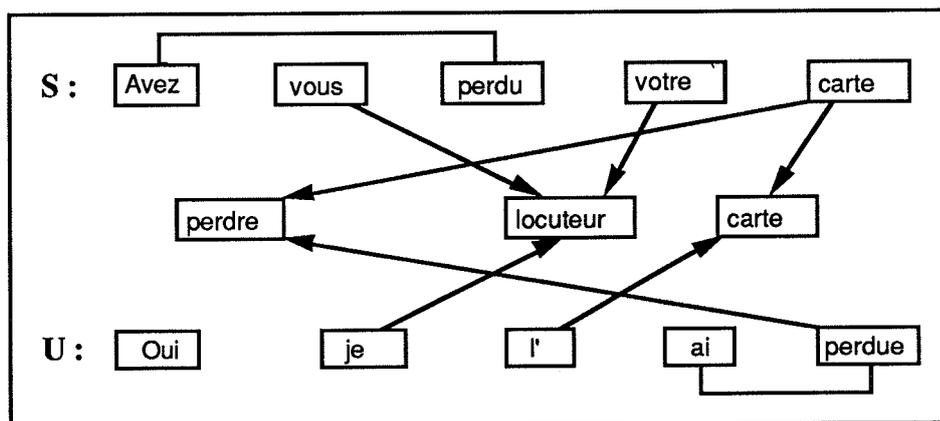


Figure 1.2 : Gestion des co-références dans le système Dial.

Ce type de traitements pose un certain nombre de problèmes qui sont liés pour la plupart au choix d'architecture effectué au départ. En effet la grande modularité et donc la séparation quasi figée des tâches correspondant à chaque module font qu'il est impossible par exemple d'effectuer des vérifications syntaxiques lors du calcul des co-références. Ainsi, comme on l'a vu, les calculs destinés à des coréférences intraphrastiques se trouvent complètement détachés des calculs nécessaires à la résolution de coréférences interphrastiques.

1.3 Un premier bilan.

Au vu des mécanismes présentés dans le cadre des deux systèmes de gestion de dialogues précédents, on observe un certain nombre de constantes dans le type de traitements réalisés à l'heure actuelle. Les deux principales techniques qui ressortent sont les suivantes :

- d'une part la nécessité d'un historique du dialogue qui mémorise les objets mentionnés dans le discours ainsi que les contraintes correspondantes;
- un algorithme élémentaire de parcours de cet historique qui se limite bien souvent à une recherche linéaire en arrière s'arrêtant au premier candidat convenable.

Cependant, un certain nombre de problèmes restent en suspens, ceux-là même qui ne sont pas pris en compte par ces techniques simples : tout d'abord, il est clair que l'on tient assez peu compte de la structure globale du dialogue pour effectuer les résolutions de co-référence ou d'anaphore, si ce n'est dans des cas tels que les couples questions/réponses que des règles peuvent prendre en compte. Par ailleurs, la résolution des ellipses/anaphores correspond à un traitement spécifique et reste ainsi séparée des autres mécanismes de compréhension relatifs par exemple au suivi des différents états ou actions par lesquels passe l'un des objets. Si donc d'aventure l'un des mécanismes a besoin de l'autre, le système n'est pas conçu pour répondre à cette attente et cela débouchera sur une possible erreur de compréhension. A ce stade, il nous apparaît nécessaire d'utiliser des modèles plus sophistiqués de représentation et de traitement des informations présentes dans un dialogue, ce qui nous conduit à étudier quelques propositions présentes dans la littérature.

2. anaphore et modélisation du discours.

Bien que notre objectif soit de parvenir à une gestion efficace des co-références dans un dialogue homme-machine, la littérature est tout aussi riche en études concernant les textes narratifs que les dialogues à proprement parler (par exemple la *Discourse Representation Theory* de Kamp). Il est intéressant de remarquer que le terme fédérateur de discours se retrouve dans l'ensemble des travaux que nous allons maintenant parcourir et recouvre relativement bien le champ d'analyse où les problèmes de co-références occupent une part importante.

2.1 Anaphore et focus.

On peut distinguer deux grandes classes d'approches correspondant à l'étude du discours dans les années 80. D'une part des modèles guidés par la tâche qui considèrent notamment que le dialogue vise seulement à servir celle-ci et d'autre part des modèles plus 'linguistiques' qui structurent un discours à partir de marqueurs présents dans la langue ou sur la base d'informations générales extraites de l'étude d'autres discours.

Dans le premier type d'approche on peut trouver les travaux de Grosz (1981) qui structure de façon hiérarchique le discours de sorte que cette structure colle littéralement à la structure logique de la tâche correspondante. Cette structure logique va correspondre soit à la description des objets du discours dont il est question (structure d'un mécanisme) soit à des actions intervenant dans la tâche (scripts de la tâche par exemple). Cette hiérarchie va alors influencer profondément la résolution des pronoms et autres anaphores puisqu'elle limite les possibilités d'accès à certaines informations présentes dans l'historique du dialogue (assimilé à l'historique de la tâche).

Sidner se situe dans une perspective opposée (bien qu'ils se retrouvent dans leur article de 1986) puisqu'il propose une analyse plus discursive des informations présentes dans un dialogue. Il introduit en particulier la notion de *focus* suite à la remarque simple

qu'un discours à un moment donné parle de "quelque chose" et que c'est ce quelque chose qu'il faut enregistrer pour mieux comprendre les énoncés à venir. En fait Sidner établit différents types de focus suivant le type des traitements à réaliser ultérieurement :

- focus des acteurs
- focus du discours
- focus potentiel des acteurs
- focus potentiel du discours

Nous verrons dans les exemples qui suivent le rôle exact de ces différents focus. Pour l'heure nous pouvons déjà présenter le cycle élémentaire de traitement nécessaire pour la gestion de ces focus :

- | |
|---|
| <ol style="list-style-type: none">1) choisir un focus2) résoudre les anaphores3) mettre à jour la liste des focus |
|---|

• **première étape : choisir un focus.**

Le cycle présenté ci-dessus est une procédure itérative qui est mise en œuvre successivement pour chaque énoncé à traiter par un système de gestion de dialogues. Ainsi, à chaque étape, la procédure spécifique de choix d'un focus utilise les informations mémorisées lors des étapes précédentes. Cependant, lors du traitement du premier énoncé, il n'existe pas encore d'élément dans l'un quelconque des focus mentionnés plus haut et il est donc nécessaire d'établir certains critères permettant d'isoler des focus potentiels pour la suite des opérations. Un certain nombre d'éléments sont en fait classiquement considérés comme pertinents dans un discours d'un point de vue thématique et sont à ce titre introduits par Sidner pour initialiser les listes de focus¹. En premier lieu - bien que ce ne soit pas nécessairement le cas le plus courant - on peut mentionner les "there-insertion", qui sont des mises en évidence d'un item particulier par une expression du type "il y a x" (*there is x*) permettant de poser x comme thème du discours à venir (on trouvera dans les "there-insertion" d'autres expressions telles que "il était une fois", qui ne sont pas de notre ressort pour les types de dialogues que nous considérons).

Une autre catégorie - plus importante pour nous - d'éléments susceptibles d'apparaître dans les listes de focus se présente sous la forme de groupes nominaux en situation particulière dans l'énoncé soit pour des raisons syntaxico-sémantiques soit parce qu'ils occupent un rôle thématique prépondérant. Ainsi, on considère que du point de vue d'une pronominalisation possible dans un deuxième énoncé, l'entité ayant occupé le rôle d'agent dans le premier énoncé est prioritaire (au sens des probabilités), suivi éventuellement des autres cas de cet énoncé tel que le patient par exemple. Cependant, Sidner ne considère pas exactement la situation sous cet angle et préfère introduire - comme on l'a déjà vu - un focus spécifique d'acteur dans lequel vont se trouver réunis les différents éléments en position d'agent comme le montre l'exemple suivant. Soit la suite d'énoncés :

*Mary prit une pièce de sa tirelire
Elle la plaça sur la table près de Pierre.*

A l'issue de l'analyse du premier de ceux-ci, 'Mary' est mis dans le focus d'acteurs et 'une pièce' dans le focus du discours. Il s'en suit des opérations spécifiques d'une part pour résoudre le pronom 'elle' en position d'agent dans le deuxième énoncé et d'autre part pour 'la' en position d'objet. On peut remarquer que cette division en deux classes de la notion de focus est intéressante en soi, bien qu'elle puisse être vue comme un

¹ La notion de focus recouvre partiellement le *thème*, tel qu'il est défini en linguistique. Cependant, sa fonction spécifique d'outil de calcul de l'anaphore fait que son rôle thématique n'intervient que partiellement dans son établissement et n'est d'ailleurs pas particulièrement mis en avant par Sidner.

moyen ad hoc de régler certains problèmes de traitement. C'est en fait une tentative de structurer de manière élémentaire le focus par rapport à la structure de l'énoncé qu'il est censé traduire. L'intuition que l'on peut avoir ici est de pousser encore plus loin cette notion en considérant le focus non plus comme une entité homogène, mais plutôt comme un objet complexe qui s'apparenterait à un modèle de mémoire plutôt qu'à une structure informatique (tableau ou pile...).

Enfin, il ne faut pas oublier un dernier élément : le syntagme verbal, qui va nous permettre de résoudre des ellipses ou même certaines anaphores 'verbales'.

Maintenant que nous avons en main l'essentiel des éléments utiles pour une analyse complète d'un énoncé, nous pouvons effectuer une synthèse dans le cadre de l'exemple suivant :

La semaine dernière, j'ai acheté des fraises pour Pierre au supermarché.

Comme il s'agit d'un premier énoncé, les listes calculées ne sont que *prévisionnelles*, puisque dans l'optique de Sidner, ne seront considérés comme focus que les éléments sur lesquels une coréférence aura été faite. La liste prévisionnelle des focus du discours est alors la suivante :

- 1) "Les fraises" (le thème de cet énoncé)
- 2) Pierre (le patient), le supermarché (déterminant le contexte de lieu)
- 3) "Je" (l'agent, rejeté en position secondaire puisqu'acteur de surcroît)
- 4) "avoir acheté des fraises" (le syntagme verbal qu'il ne faut pas oublier)

Par ailleurs on peut calculer la liste prévisionnelle des focus des acteurs qui ne doit contenir à ce titre que des objets animés selon Sidner (bien que cette hypothèse soit un peu trop restrictive dans certaines applications) :

- 1) "Je" (l'agent de notre énoncé)
- 2) Pierre (le patient est mis en arrière dans le focus des acteurs, mais on peut remarquer que l'ordre est inversé dans le focus du discours).

C'est dans ces deux listes que nous allons pouvoir choisir les candidats pour la résolution des différentes anaphores de l'énoncé à venir.

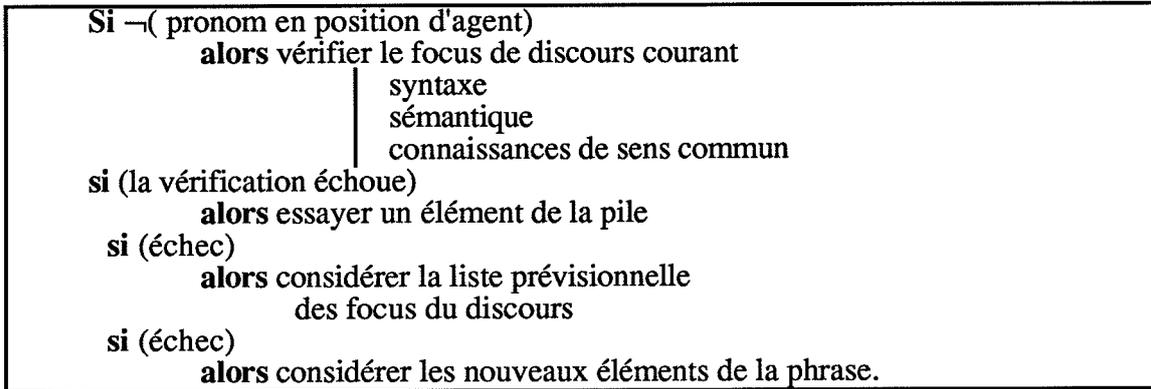
• deuxième étape : résoudre les anaphores.

Il ne suffit pas de disposer des éléments pertinents pour la résolution des anaphores, il faut aussi établir la ou les stratégies à utiliser pour choisir le candidat convenable parmi tous les items potentiels. Notre propre expérience en la matière (cf les présentations des systèmes Partner et Dial) nous a montré que les stratégies disponibles sont assez faibles et il n'est pas étonnant que la première règle énoncée par Sidner soit la fameuse *règle de proximité* qui conseille d'essayer en priorité la plus récente des entités apparues dans le contexte discursif. De fait, cette règle est généralement la plus efficace du point de vue des calculs et donne en général d'excellents résultats, comme le montre Sidner dans l'exemple suivant où "il" va bien reprendre le dernier cerf-volant introduit et non pas celui que Janet envisage d'acheter :

C'était l'anniversaire de Leon. Pierre et Jeanne allèrent au grand magasin. Ils voulaient acheter des cadeaux. Janet décida d'acheter un cerf-volant. "Ne fait pas ça", lui dit Pierre, "Leon a déjà un cerf-volant. Il est vraiment très bien".

Le défaut principal des règles de ce type réside dans leurs manque d'un soutien théorique qui permettrait de les justifier et de les insérer dans un cadre élargi de l'étude du discours. Dans la plupart des systèmes, elles résultent d'ailleurs de l'observation expérimentale de discours pris un peu au hasard ou bien pire, de l'intuition initiale des concepteurs de systèmes de dialogue. De plus, la règle de proximité est en elle-même critiquable, car elle ne tient pas du tout compte de la structure effective du discours sur

lequel elle agit. Ainsi, les systèmes de gestion de dialogue doivent souvent la contrecarrer par d'autres règles établissant les cas pour lesquels elle n'est pas applicable. On retrouve ceci chez Sidner, puisqu'il introduit une *règle élémentaire de résolution* établissant une stratégie adaptée à la structure de données (les focus et les listes prévisionnelles de focus) qu'il a introduite. Cette règle peut être présentée sous forme algorithmique de la façon suivante :



Dans cet algorithme, on voit apparaître un nouvel élément que nous n'avons pas encore mentionné, il s'agit de la *pile* des focus, comprenant les anciens focus relatifs aux énoncés précédant le dernier énoncé et organisés sous la forme d'une pile, c'est à dire une structure où l'objet le plus accessible est l'objet le plus récemment introduit ("last in-first out").

Cette règle ne fait qu'établir un ordre de priorité entre les différentes structures de mémorisation introduites par Sidner. En aucun cas elle ne prend en compte l'interaction éventuelle des différents candidats au fil des énoncés, par exemple dans le cadre de différents événements d'une histoire, ou de liens pragmatiques qui pourraient être connus du locuteur. On peut ainsi réfléchir à la réelle validité théorique d'une telle règle - tout à fait utile par ailleurs au niveau pratique, dans un premier temps - dans le sens où elle ne fournit pas d'explication concernant le fonctionnement de l'activité discursive. En effet, elle tend à séparer les indices utilisables pour la résolution des anaphores, et plus généralement des co-références, de la source première de ceux-ci à savoir l'organisation sous-jacente du discours (ce qui correspond au niveau de la *narration* par opposition aux *fonctions* et aux *actions* dans le texte déjà ancien de R.Barthes, 1966). D'une manière générale, cette remarque remet en question la notion de règle, telle qu'elle est couramment utilisée en informatique pour le traitement du langage naturel et des dialogues. Elle sert souvent à masquer notre manque de connaissance des structures plus fines de notre matière d'étude et nous empêche bien souvent de concevoir effectivement des modèles plus fins de cette même matière.

Considérons maintenant une résolution dans le cadre de l'exemple suivant :

*La semaine dernière, j'ai acheté des fraises pour Pierre au supermarché.
 Elles étaient particulièrement fraîches.*

Si nous nous attachons au problème de la résolution de l'anaphore "elles" dans le deuxième énoncé, on s'aperçoit que le focus et la pile des focus sont vides respectivement (puisque nous n'en sommes qu'au deuxième énoncé); il n'est donc possible que d'explorer la liste prévisionnelle des focus du discours ("elles" n'étant pas en position d'agent). Cette liste avait été obtenue précédemment :

- 1) Les fraises (le thème)
- 2) Pierre (patient), le supermarché (lieu)
- 3) "Je" (l'agent)
- 4) "avoir acheté des fraises" (le syntagme verbal)

Le premier candidat ("les fraises") convenant parfaitement, on arrête là la recherche et l'anaphore est supposée résolue. A ce stade, il faut rappeler qu'un certain nombre de vérifications sont à faire. Pour ce qui est des vérifications syntaxiques, aucun problème ne se pose ici. Cependant, il faut s'assurer que "fraises" est un sujet acceptable pour "être fraîches". Où doivent donc se situer ces vérifications ? Est-ce au niveau sémantique, à l'aide de traits ad hoc ? Cette solution, la moins élégante, est en général la plus facile à réaliser. L'autre possibilité est l'introduction de prototype - on se situe alors à un niveau plus pragmatique - mais cette solution, pour autant qu'elle paraisse belle, n'en est pas moins la plus lourde à gérer. Nous ne discuterons pas plus avant ce problème, tout en gardant à l'esprit qu'il restera d'importance lors de toute mise en œuvre de système de gestion de dialogues.

• troisième étape : mise à jour de la liste des focus

La mise à jour de la liste des focus repose sur les mêmes principes qui nous avaient permis de construire initialement la liste prévisionnelle des focus. La seule différence provient qu'il est possible, à partir du deuxième énoncé, d'établir le *focus du discours*. Celui-ci est positionné sur l'élément qui a été activé lors des résolutions des co-références. Ce focus du discours disparaît éventuellement des deux listes prévisionnelles des focus du discours et des focus d'acteur ainsi que le montre l'exemple complet suivant :

- a. *La semaine dernière, j'ai acheté des fraises pour Pierre au supermarché.*
- b. *Elles étaient particulièrement fraîches.*
- c. *Je voulais les utiliser pour le dessert,*
- d. *mais quelqu'un les avait toutes mangées.*
- e. *J'ai découvert que c'était Marc qui les avait mangées.*
- f. *Il mange beaucoup.*

Nous avons déjà traité le cas de l'énoncé a et les conséquences de cette analyse sur la résolution de "elles" dans b. Nous pouvons maintenant - en reprenant les mêmes principes de construction - établir les différentes structures de Sidner à l'issue de b :

- b : *Focus du discours : les fraises*
Liste prévisionnelle des focus du discours : Pierre, le supermarché, "Je"
Liste prévisionnelle des focus d'acteurs : "Je", Pierre

puis à l'issue de c, d et e :

- c : *Focus du discours : les fraises*
Liste prévisionnelle des focus du discours : le dessert
Focus d'acteur : "Je"
- d : *Focus du discours : les fraises*
Liste prévisionnelle des focus du discours : quelqu'un
Focus d'acteur : "Je"
Liste prévisionnelle des focus d'acteurs : quelqu'un
- e : *Focus du discours : les fraises*
Liste prévisionnelle des focus du discours : Marc
Focus d'acteur : "Je"
Liste prévisionnelle des focus d'acteurs : Marc

A ce stade, et au vu de l'analyse précédente, on peut faire une dernière remarque concernant les éléments introduits dans les focus et listes de focus. Il apparaît que l'on retrouve d'une certaine manière la vision commune à de nombreux systèmes de gestion de dialogue, à savoir l'entité du discours vue comme une étiquette transmise d'énoncé à énoncé. Ces étiquettes, que l'on retrouve aussi dans la DRT de Kamp sous la forme de variable du discours, assurent la continuité référentielle le long du discours. Cependant, comme elles sont définies *a priori* et en dehors de toute structure du référent, elles ne

permettent pas aux entités d'évoluer au cours du discours (changement de propriétés, participation à des actions). Ce point est crucial pour la mise en place d'un véritable système de dialogue "intelligent", car, en dehors de discours très courts, il est difficile de se limiter à des univers purement synchroniques au niveau factuel. Bien plus, prendre en compte ces phénomènes permet de considérer dans le même cadre le locuteur lui-même, dont les connaissances (ou *l'état de connaissances*) évolue au fil d'un dialogue (si le système lui apporte des données par exemple). On voit donc là encore qu'un important travail de modélisation reste à faire, ce qui est le sens de la démarche que nous présenterons dans la dernière partie.

2.2 Prise en compte de la structure du discours (Grosz et Sidner).

Dans un article de 1986, Grosz et Sidner combinent leurs approches respectives du dialogue pour proposer une structure plus complète de celui-ci s'appuyant d'une part sur une représentation de la tâche (préoccupation antérieure de B.Grosz) et d'autre part sur une utilisation des différents éléments introduits par Sidner et que nous avons présentés dans la section précédente. Nous n'allons décrire ici que les points les plus importants pour notre analyse, sachant que la mise en correspondance de deux modèles d'origine différente est une tâche délicate et en général ponctuelle.

Le principe général régissant le modèle proposé par Sidner et Grosz est qu'un discours peut être décomposé suivant trois axes :

- **la structure linguistique.**

Tout discours est formé d'une suite de segments emboîtés les uns dans les autres, formant des contextes où les co-références s'effectuent de manière privilégiées (on peut rapprocher cette idée des travaux de M.Charolles sur les chaînes de co-référence). Comme dans le cas de nombreux travaux sur le dialogue ou le discours en général, cette notion de structure linguistique pose un certain nombre de problèmes définitoires. Ceci est d'autant plus clair chez Grosz et Sidner qu'ils indiquent que les anaphores se résolvent dans le cadre de cette structure, elle-même définie par les co-références apparaissant dans un discours. Cette définition récursive ne se conçoit qu'à l'aide des outils supplémentaires qui sont introduits.

- **la structure intentionnelle.**

La structure intentionnelle représente la "logique" du discours puisqu'elle indique les liens de sens entre les différents segments de la structure linguistique (pour se fonder avec celle-ci dans certains cas). C'est une structure entièrement guidée par la tâche et qui s'appuie sur deux relations entre les segments linguistiques : d'une part la dominance, indiquant qu'un segment ne peut se résoudre tant que le segment dominé n'est pas encore résolu, et la précédence qui indique au contraire qu'un segment particulier doit être clos avant que le suivant ne commence. Ces relations font fortement penser à des relations entre événements (cf le modèle des Zones Temporelles introduit plus loin)

- **la structure attentionnelle.**

Cette dernière structure reprend dans ses grandes lignes les constructions à base de focus déjà proposées par Sidner. Cette notion d'attentionnalité reflète l'état de connaissance de l'interlocuteur après la compréhension d'un énoncé puisqu'il s'agit des éléments mémorisés, et donc saillants, qui sont susceptibles d'être réutilisés dans le cadre de co-références. La pile que Sidner introduit pour enregistrer les différents contextes propres à chaque énoncé est en fait filtrée par la structure intentionnelle qui définit ainsi des conditions d'accessibilité d'un contexte à un autre.

3. Et l'anaphore temporelle ?

3.1 Temps et objets du discours.

Nous nous sommes limités jusqu'ici à l'étude des anaphores et plus généralement des co-références à des objets explicites du discours participant aux actions, états ou événements de la tâche. Il est cependant possible de rapprocher cette notion d'anaphore de celle introduite dans le cadre de nombreux travaux sur le temps et l'aspect en linguistique. Bien que ce rapprochement, comme le fait remarquer G.Kleiber, est assez flou au niveau définitoire, il montre cependant que d'une certaine manière, on ne peut isoler, dans un énoncé, les références à des objets et les références à des actions. Bien plus, toute co-référence, ou supposée telle, provient en fait de la mise en correspondance de *deux* groupes référentiels intervenant dans des contextes prédicatifs différents, c'est à dire sur des plages temporelles généralement disjointes. Il résulte donc de ceci que deux références ne s'effectuent que très rarement sur le même "objet" (au sens d'une étiquette unique...), mais bien plus sur deux états différents de quelque chose dont le locuteur et l'auditeur perçoivent la continuité. Ainsi, la notion d'anaphore peut dès le départ s'articuler avec la notion de temps sans qu'il n'y ait - justement - de co-référence temporelle (ou d'anaphore temporelle). Par ailleurs, il est vrai qu'il existe malgré tout de véritables anaphores temporelles exprimées en général par des groupements adverbiaux ("la veille"). De plus, ceux-ci s'opposent à ce qu'on appelle des déictiques temporelles ("hier") formant un ensemble d'expressions permettant de situer un énoncé dans le temps. La difficulté qui apparaît ici est que l'on aboutit nécessairement au problème des temps grammaticaux que nous avons rapidement mentionnés au début de ce paragraphe. Nous pouvons rapidement illustrer ce phénomène par différents exemples proposés par Hinrichs (1986) dans lesquels les adverbiaux et les temps grammaticaux ne peuvent être interprétés (positionnement du contexte) qu'en combinant les informations apportées par chacun.

(1) *Nous sommes sortis la semaine dernière.
Nous nous sommes bien amusés.*

(2) *Quand le train arriva, Jean s'avança.*

(3) *A cet instant, la bombe explosa. Trois heures plus tard, je me réveillai dans un étrange hôtel.*

Dans 1, la continuité du discours au niveau temporel permet de situer le deuxième passé composé à l'intérieur du contexte établi par le premier ; dans 2, "quand" établit une "simultanéité" entre deux événements et dans 3, la succession entre les deux événements est l'archétype de l'anaphore temporelle (même si effectivement il n'y a pas réellement *anaphore*).

Il semble donc qu'il y ait des rapprochements étroits à effectuer entre la notion de co-référence classique à des objets du discours et des problèmes plus temporels, qui correspondent d'ailleurs à des champs d'étude souvent séparés. Ce rapprochement apparaît de façon encore plus troublante lorsqu'on observe les traitements proposés dans chaque cas, pour lesquels on retrouve des mécanismes très similaires pour ne pas dire équivalents dans le cadre de certains modèles. Nous allons présenter rapidement deux aspects du traitement temporel, à savoir la représentation en trois points de Reichenbach et la DRT, en nous limitant aux seules données utiles pour notre exposé.

3.2 La représentation de Reichenbach.

Reichenbach (1947) a proposé une étude des temps grammaticaux de l'anglais sur la base de trois points temporels E, R, S, où :

- E (Event time) est le temps de l'événement décrit dans le discours ;

- S (Speech time) est le moment où le discours est exprimé², temps supposé en général ponctuel ;

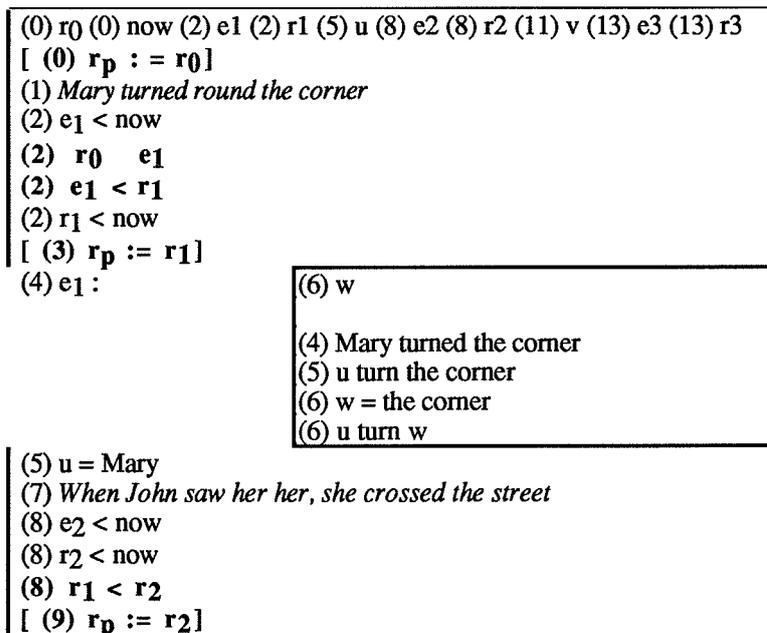
- R (Reference time) est un point de référence permettant d'une part d'articuler E et S l'un par rapport à l'autre et d'autre part de tenir compte du contexte temporel défini par des adverbiaux ou des énoncés ayant précédé l'énoncé courant.

L'aspect intéressant de cette représentation est pour nous le point de référence, qui permet finalement de tenir compte de la progression du discours, c'est-à-dire - en simplifiant - de la partie anaphorique de son analyse temporelle. Servant de mémoire pour l'énoncé courant, il agit plus ou moins comme un *focus temporel* sur lequel vont s'appuyer toutes les nouvelles références temporelles. Ainsi, dans l'exemple suivant, de même qu'une mémoire est nécessaire pour suivre les co-références au même individu "Pierre", une mémoire doit enregistrer chaque événement de l'énoncé, pour suivre l'évolution temporelle de l'histoire. Ce parallèle est d'autant plus facile à faire, dès lors qu'on se souvient de la remarque que les objets référés sont eux-mêmes susceptibles d'évoluer :

Pierre ouvrit la bouteille. Il se servit un verre. Il admira le bouquet. Les yeux fermés, il but le vin.

3.3 Exemple d'utilisation dans le cadre de la DRT.

Nous mentionnons ici les travaux de Barbara Partee (1984) qui a intégré une représentation des événements du discours dans le cadre de la DRT (Discourse Representation Theory) de Kamp. Le parallèle entre le traitement de la référence à des objets du discours et des événements à caractère temporel passe ici par la notion de variable qui, dans une DRS (Discourse Representation Structure), assure la continuité référentielle à un même objet. B.Partee a représenté le même type de mécanisme, en le combinant avec les trois points de Reichenbach, pour fournir une représentation uniforme où des relations entre variables temporelles ($e_1 < r_1$) apparaissent au même titre que des relations entre des variables standard de la DRT (u turn w) comme on peut le voir dans la figure 3.1.



²Ce moment, facile à définir dans le cas d'un discours oral, peut être considéré comme une abstraction dans le cadre de textes écrits. Suivant les cas on considérera le temps de l'écriture (compte rendu), le temps du narrateur (cf le jeu des temps dans le Nom de la rose d'U.Eco) ou le temps de lecture (instructions à exécuter).

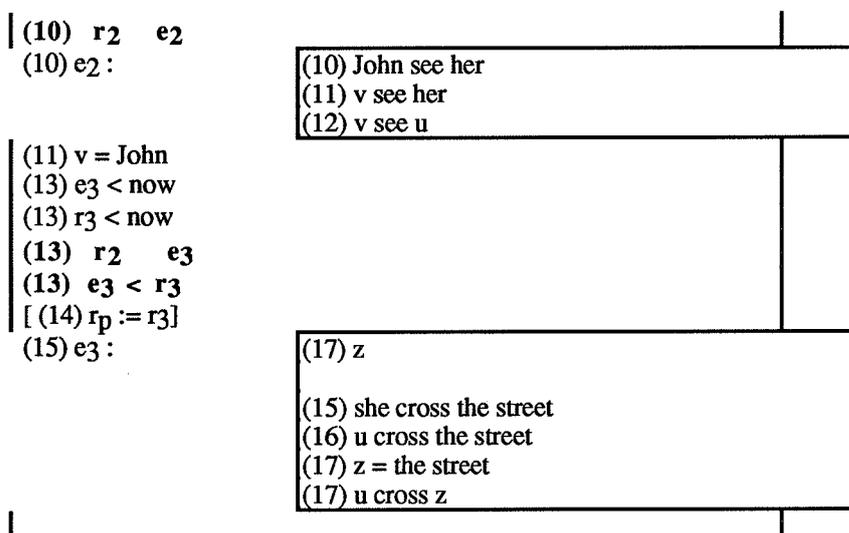


Figure 3.1 : représentation d'un discours(d'après Partee, 1984)

3.4 Un bilan partiel.

L'examen de différentes pistes de recherche concernant les problèmes du temps et leur lien très étroit avec les mécanismes de l'anaphore montre que nous sommes en face de deux domaines qu'il est difficile d'étudier de façon totalement séparée. Cependant il est clair que bon nombre d'études sur l'anaphore négligent les phénomènes d'évolution des référents par souci de délimitation du champ de recherche, et par ailleurs, les études sur le temps correspondent souvent à des besoins spécifiques de l'étude du verbe et des modalités où le problème de la référence n'a que peu de place. Même la DRT, qui pourrait apparaître en première analyse comme un cadre fédérateur pour ces problèmes, est limitée par son propre cadre de représentation qui ne s'attache pas à structurer l'articulation entre les référents ou les événements, en limitant la référence à l'établissement d'une étiquette unique pour un objet dont les propriétés doivent être stables durant un discours particulier. Pourtant l'articulation (et non pas simplement la juxtaposition) entre les deux notions est fondamentale et peut même guider la compréhension d'un discours comme le montre l'exemple tiré de Webber (1988) :

*John went to visit Mary. He had bought her some flowers.
He picked 5 red roses and 4 white ones. Unfortunately,
they failed to cheer her up.*

Dans cet exemple Webber propose une simple analyse temporelle (c'est à dire s'appuyant sur des indices temporels) pour situer les différents événements apparaissant dans ce bout de texte. Ainsi, la dernière phrase est incluse au niveau temporel dans le temps de la première, malgré la digression intermédiaire. En fait, cette analyse peut être renforcée si l'on considère les chaînes de co-références et la ré-apparition d'une référence à Mary dans le dernier énoncé. Cette information tend à resituer la dernière phrase dans le contexte où Mary avait été introduite, ce qui facilite du coup l'interprétation temporelle.

L'interaction entre les phénomènes référentiels et temporels a aussi une conséquence sur la forme que doivent prendre les modèles utilisés pour la conception de systèmes intelligents de dialogue homme machine. Il apparaît que non seulement il est nécessaire de posséder un modèle du discours lui-même tels que ceux de Sidner ou de Kamp par exemple, mais que cela était insuffisant dès lors que l'on considère qu'un référent est en fait un objet complexe au niveau structure (l'ensemble des propriétés qu'il possède) mais aussi au niveau temporel. Il est donc indispensable de pousser plus avant des recherches sur des modèles de représentation des connaissances adaptés aux exigences du traitement de la référence, notamment au niveau temporel. Un tel modèle est relativement complexe à construire et nous ne proposons dans nos travaux actuels que des éléments de réponses sur des points très précis [Romary91]. L'ensemble de ces

travaux repose néanmoins sur une hypothèse particulière qui est que toute entité (concept) et ses propriétés ont une existence susceptible d'être limitée dans le temps.

4. proposition d'un cadre de représentation et de réflexion.

Nous travaillons actuellement sur un modèle de représentation des informations présentes dans un discours dont nous n'allons donner que les lignes principales. Notre objectif est de montrer qu'il est possible de proposer un cadre de représentation des référents où les informations synchroniques et diachroniques apparaissent de façon homogène. Ce type de modèle semble pour nous une des voies propices à un traitement des phénomènes de co-référence incluant les différents aspects dont nous avons parlé jusqu'ici.

4.1 Définition d'un modèle de représentation.

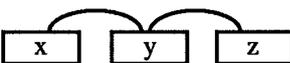
La représentation s'appuie sur un ensemble d'objets temporels élémentaires : les Zones Temporelles (ZTs) entre lesquelles sont définies deux relations élémentaires, l'adjacence qui marque que deux ZTs se succèdent dans le temps (pas nécessairement strictement) et l'inclusion qui indique qu'une Zone en contient une autre d'un point de vue toujours temporel. Ces définitions se traduisent de la façon suivante :

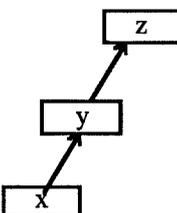
Z = ensemble d'objets structurés

$RZ = \{in, prec\}$ deux relations binaires sur Z

Afin de permettre à ces relations de fonctionner à l'intérieur d'un ensemble de ZTs, il est nécessaire de leur adjoindre un certain nombre de règles : d'une part des règles de transition permettant de déduire une troisième relation entre deux zones x et z , connaissant certaines relations entre x et y , et y et z , et d'autre part des règles d'incompatibilité limitant les relations pouvant apparaître entre deux zones données (une représentation graphique est associée à ces règles pour visualiser leur sémantique) :

• règles de transition

$$\frac{prec(x,y), prec(y,z)}{prec(x,z)}$$


$$\frac{in(x,y), in(y,z)}{in(x,z)}$$


• règles d'incompatibilité

$$\frac{in(x,y), in(y,x)}{\textcircled{R}} \qquad \frac{prec(x,y), prec(y,x)}{\textcircled{R}}$$

$$\frac{in(x,y), prec(x,y)}{\textcircled{R}} \qquad \frac{in(x,y), prec(y,x)}{\textcircled{R}}$$

Ces objets temporels vont maintenant former l'ossature d'une représentation plus complète des informations exprimées dans un discours. En effet, chaque zone temporelle va définir un espace dans lequel un certain nombre de faits relatifs aux objets du discours resteront valides. Dans le cadre de cet exposé, nous nous limitons aux seuls faits décrivant des relations spatiales entre les objets du discours. Ainsi, une zone temporelle

contiendra un certain nombre de faits spatiaux qui sont valides à l'intérieur même de ces zones. La définition des ces faits spatiaux est analogue à celle des expressions temporelles : on introduit un ensemble de variables spatiales accompagnées de deux relations simplifiées exprimant les liens géographiques pouvant apparaître entre ces variables. On a ainsi les définitions suivantes :

S = ensemble de symboles représentant des objets concrets (dans le cadre réduit qui nous intéresse ici.

$RS = \{ \text{dans}, \text{a_cote_de} \}$, deux relations binaires sur S

Là encore il est nécessaire d'ajouter un certain nombre de règles à ces relations. On notera une seule différence de forme avec les règles précédentes, différence qui exprime le caractère non-orienté de la relation 'a_cote_de'.

Règles de transition :

$$\frac{\text{dans}(x,y), \text{dans}(y,z)}{\text{dans}(x,z)}$$

$$\frac{\text{dans}(x,y), \text{a_cote_de}(y,z)}{\text{a_cote_de}(x,z)}$$

$$\frac{\text{a_cote_de}(x,y)}{\text{a_cote_de}(y,x)}$$

Règles d'incompatibilité :

$$\frac{\text{dans}(x,y), \text{a_cote_de}(x,y)}{\textcircled{R}}$$

$$\frac{\text{dans}(x,y), \text{dans}(y,x)}{\textcircled{R}}$$

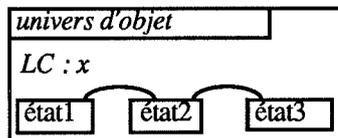
Les deux systèmes de représentation/raisonnement temporel et spatial que nous venons de présenter doivent maintenant être combinés pour former un modèle homogène de représentation des informations spatio-temporelles exprimées dans un discours. Une de nos hypothèses essentielles étant que tout objet référé dans un énoncé, ainsi que ses différentes propriétés ont une existence limitée dans le temps, il est immédiat d'utiliser les zones temporelles pour décrire ces objets. Le principe mis en œuvre ici repose sur l'idée que certaines zones temporelles peuvent être spécifiquement affectées à la représentation du cadre d'existence d'un objet du discours. Nous appellerons ces zones des *Univers d'Objet*. De manière plus complète, nous pouvons présenter la structure de tels univers d'objet ainsi :

Zone temporelle structurée :

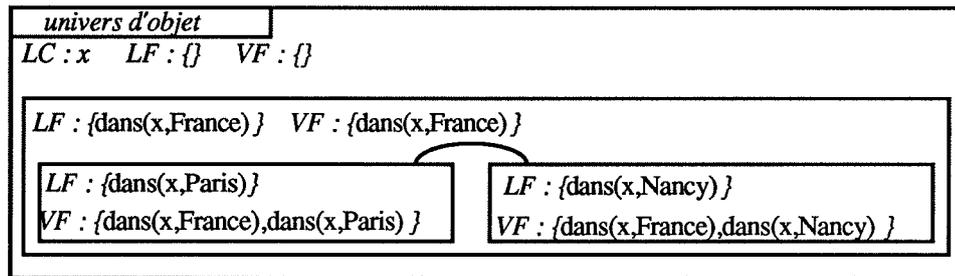
univers d'objet =
 Constante locale (LC)
 Faits locaux (LF)
 Faits Valides (VF)

où la constante locale est éventuellement un symbole sur lequel va porter toute prédication d'ordre spatial, les faits locaux, un ensemble de faits définis directement au niveau de la zone temporelle courante, et les faits valides, tous les faits qui peuvent être déduits, soit des faits locaux, soit de tout fait hérité de zones contenant la zone temporelle courante (au sens de la relation *in*). En fait, nous pouvons étendre cette structure à toute zone temporelle, aucune constante locale n'étant définie pour des zones ne pouvant pas être assimilées à des univers d'objet.

Dans ce cadre, un objet décrit par une Constante Locale x , peut apparaître dans différents états ou événements :



Bien plus, ces différents états peuvent être eux-mêmes structurés, ce qui permet d'exprimer la continuité de certains faits, alors que d'autres évoluent au cours d'un discours. Ainsi, si nous apprenons pour un référent qu'il passe successivement de Paris à Nancy tout en restant en France, la représentation obtenue sera la suivante :



Ainsi, toute structure temporelle associée à un objet du discours, aussi complexe qu'elle puisse être, peut être représentée à l'intérieur d'un univers d'objet. Ceci permet de prendre en compte un certain nombre de phénomènes liés à la référence. Tout d'abord, nous avons mentionné l'importance d'une représentation qui suive les différents états d'un objet tels qu'ils peuvent être décrits dans une suite d'énoncés. Il est aussi possible d'envisager des références complexes faisant intervenir d'anciennes informations sur un objet (e.g. : "le livre qui était sur la table"). Enfin, il est possible de tenir compte du résultat d'une action pour établir les informations valides concernant un objet au moment du calcul des références. Ceci est particulièrement important lorsqu'on travaille dans des univers factuels où l'on agit directement sur les objets d'un univers.

A titre d'exemple, nous pouvons reprendre le discours culinaire mentionné par G.Kleiber : "Tuez un poulet bien vif et bien gras. Préparez le pour le four, coupez le en quatre et rôtissez-le pendant une heure". Les différentes actions (tuer, préparer, couper, rôtir) modifient successivement les propriétés physiques de l'objet 'poulet vif et bien gras' du départ. Cependant, afin de reconnaître une continuité à ce discours, il est nécessaire de l'interpréter de sorte que les références successives soient associées à un même concept. Dans le cadre que nous avons défini, cela revient à ne considérer qu'un seul univers d'objet (et donc une seule constante locale) contenant les différents états par lesquels passent le poulet. De plus, si l'on dispose de scripts généraux concernant l'art culinaire, il est possible d'intercaler entre ces états un certain nombre de schémas explicatifs qui renforcent le caractère cohérent de ce discours.

5 Conclusion : discours et connaissances du monde.

Dans cet exposé, nous nous sommes essentiellement attachés à présenter un certain nombre de difficultés et de problèmes qui apparaissent lorsque l'on désire mettre en œuvre un calcul de référence dans un système de dialogue homme-machine. Nous avons essentiellement porté notre attention sur deux aspects :

- la description des problèmes de coréférence au niveau du discours, ce qui nous a permis de mettre en évidence la nécessité de prendre en compte la dimension temporelle des objets du discours;
- la proposition d'un cadre de représentation de certaines informations liées aux référents afin de mettre en place les premiers éléments d'un modèle complet des informations discursives.

Un de nos objectifs en définissant un tel cadre est ne plus être obligé de traiter l'anaphore isolément par des mécanismes ad hoc (focus etc...), mais d'avoir à notre disposition une représentation unique, commune à tous les types possibles de référence (et donc de co-référence). Ceci permet de plus de ne plus être obligé de fonctionner par l'intermédiaire de règles liées à un formalisme spécifique, mais plutôt de définir des possibles en ce qui concerne les accès à cette représentation. Ces possibles, par opposition à un jeu de règles, vont nécessiter une plus étroite collaboration entre

linguistes et informaticiens, puisque le transfert d'expertise risque de passer par une étude s'appuyant sur des cadres formels communs.

6. Bibliographie

- Barthes R., 1966, "Introduction à l'analyse structurale des récits", *Communications*, 8, Points Seuil, 1981.
- Carbonell N. et Pierrel J.M., 1989, "Vers un dialogue naturel homme-machine : apport des études sur les interfaces orales en langue naturelle", *Actes du colloque sur l'ingénierie des interfaces homme-machine*, Cargèse.
- Grosz B., 1981, "Focusing and description in natural language dialogues", In Joshi A., Webber B. and Sag. I. (eds), *Elements of discourse understanding*, pp.48-105, Cambridge University Press.
- Grosz B. and Sidner C., 1986, "Attention, Intentions and the structure of discourse", *Computational Linguistics*, 12, pp.175-204.
- Hinrichs, E., 1986, "Temporal anaphora in discourses of English", *Linguistics and Philosophy*, 9, pp.63-82.
- Kamp H. & Rohrer C. (1983) "Tense in texts", in *Proceedings of the 1981 Linguistics Conference at Konstantz*, Germany, pp.250-269.
- Morin P. et Pierrel J.M., "Partner : un système de dialogue oral homme-machine", *Actes du colloque Cognitiva 87*, Paris 18-22 mai 1987.
- Partee B., 1984, "Nominal and temporal anaphora", *Linguistics and Philosophy*, 7, pp.243-286.
- Pierrel J.M., 1990, "Vers une meilleure intégration de la parole dans des systèmes de communication homme-machine", *Traitement du signal*.
- Reichenbach H., 1947, *Elements of Symbolic Logic*, London, Macmillan.
- Romary L., 1990, "Perception, langage, raisonnement : une même représentation temporelle", *Actes du troisième colloque de l'Arc*, Paris, mars 1990.
- Romary L., 1991, "Integration of spatial and temporal information produced by a natural language discourse", in *proc. Kmet 91*, Sophia-Antipolis, 22-24 avril 1991.
- Roussanaly A., *DIAL : la composante dialogue d'un système de communication orale homme-machine finalisée en langage naturel*, Thèse de doctorat de l'université de Nancy I, 1988.
- Sidner C., 1981, "Focusing for interpretation of pronouns", *American Journal of Computational Linguistics*, 4, pp.217-231.
- Sidner C., 1983, "Focusing and discourse", *Discourse Processes*, 6, pp.105-142.
- Webber B., 1988, "Tense as discourse anaphor", *Computational Linguistics*, 14, pp.61-73.