

UNE REPRÉSENTATION APPLICATIVE DES ÉNONCÉS ET DE LEURS DÉRIVATIONS

Ce deuxième article fait suite au premier, où j'ai essayé de montrer en quoi les étapes successives que représentent les grammaires de Harris constituent une approche constructive d'une analyse des mécanismes d'interprétation propres aux langues. Je voudrais montrer ici que cette approche doit être précisée et étendue pour décrire et analyser d'autres propriétés empiriques mais également pour garantir la cohérence de la grammaire.

L'analyse de l'insertion lexicale dans le cadre de la grammaire des opérateurs définie dans [GEMP] et [TLI] se révèle insuffisante pour décrire des propriétés empiriques comme l'existence d'énoncés qui ne peuvent être représentés par des combinaisons d'opérateurs se terminant par des arguments élémentaires et qui, de ce fait, ont une structure syntaxique qui ne peut s'exprimer dans le cadre strictement finitiste des types simples utilisés par Harris (voir les paragraphes 1 et 3). Autre exemple : il existe des énoncés mettant en jeu des relations de dépendance lexico-syntaxiques à plusieurs termes dont l'analyse applicative est conditionnée par des informations lexicales (voir le paragraphe 2). D'autre part, les opérations de réduction linguistique ne sont pas compatibles avec la façon dont est définie l'insertion lexicale *via* les opérateurs lexicaux. Les opérations de réductions linguistiques sont de trois sortes : les réductions « appropriées », les réductions anaphoriques et les réductions morpho-phonémiques qui accompagnent les deux premières. Les réductions anaphoriques jouent un rôle fondamental dans la grammaire des opérateurs pour décomposer les énoncés quelconques en combinaisons de discours élémentaires. Comme je l'ai montré dans mon premier article, ces réductions diffèrent des transformations « classiques », telles qu'elles sont définies dans [MSL], du fait qu'elles ne modifient pas les structures syntaxiques et ont pour seul effet de substituer des formes morphologiquement réduites aux formes lexicales explicites. Le caractère fini des dérivations est évidemment une condition *sine qua non* pour une grammaire et Harris fait l'hypothèse que ces réductions ne font intervenir que des relations directes entre opérateurs et arguments. Cette hypothèse ne peut être maintenue, ni pour les réductions « appropriées » ni pour les réductions anaphoriques :

— les réductions « appropriées » peuvent concerner des énoncés qui n'ont pas de représentation applicative strictement finie au sens de [GEMP] ;

— il existe des reconstructions anaphoriques qui mettent en jeu plusieurs processus d'identification emboîtés, qui doivent être hiérarchisés et contrôlés.

Enfin, l'amélioration de l'intégration de la métalangue dans la représentation de la langue, qui a conduit Harris à passer de la grammaire de [MSL] à celle

de [GEMP], implique une représentation des réductions à l'intérieur de la représentation de la langue (i.e. les dérivations doivent faire apparaître des énoncés). Pour ce faire, Harris a remplacé le système d'adressage et d'identification, qui permettait d'identifier les expressions identiques et de localiser leur position syntaxique, dans le cadre de [MSL], par un système de « interruption » décrit dans [GEMP] et précisé dans [TLI]. Ce système d'interruption tente de traiter la localisation des éléments à identifier au niveau des règles de linéarisation des structures applicatives et ne traite pas les conditions syntaxiques sur l'identification. Je montrerai que cette hypothèse ne peut être maintenue et que l'identification comme la localisation des éléments identifiés doivent être traitées au niveau des structures d'insertion lexicales, c'est-à-dire au niveau des dépendances applicatives qui définissent les structures d'interprétation et non au niveau des linéarisations qui rendent compte des contraintes sur la concaténation des expressions lexicales (voir le paragraphe 3).

Ces questions peuvent recevoir une solution en définissant l'insertion lexicale à partir d'une notion de fonction pour représenter les éléments lexicaux plus générale que celle de [GEMP]. J'en propose une définition dans le cadre technique d'un λ -calcul typé du second ordre, décrit dans : « Deriving the literal interpretation structures of a natural language in a second order λ -calculus », *Rapport Technique L.A.D.L.* 1986. Cette notion de fonction permet de représenter les dérivations à l'intérieur de la représentation des énoncés. Cet article ayant pour but de présenter quelques-unes des motivations linguistiques de cet élargissement, je ne présenterai qu'un résumé des principaux aspects techniques en annexe.

La formulation dans ce cadre technique des dérivations a pour objectif d'explicitier les fonctions grammaticales concernées à partir des mécanismes de l'insertion lexicale. La question sous-jacente à ce travail est de savoir jusqu'où cette analyse, la représentation de l'insertion lexicale et des dérivations à partir d'applications ordonnées d'éléments lexicaux, peut être poussée et à quels problèmes empiriques se heurteront cette notion de fonction d'insertion lexicale et l'unité d'analyse qui lui est associée. Je reviendrai sur la question des phrases simples mais cette fois reformulée du point de vue des dérivations : quelle unité d'analyse peut-on définir pour obtenir les dérivations les plus simples possible en conservant les hypothèses discutées dans mon premier article et quelle notion de phrase simple peut-on lui faire correspondre ? Je proposerai au paragraphe suivant une notion d'expression assertable comme unité d'analyse. Son évaluation dépend du contexte des énoncés particuliers. Elle est également beaucoup plus élémentaire que les notions de phrase simple utilisées ailleurs et aussi beaucoup plus élémentaire que la notion de discours. Plus précisément, elle fait intervenir des décompositions syntaxiques considérablement plus développées que celles qui sont généralement mises en jeu dans les composantes syntaxiques des grammaires (voir par exemple l'insertion des déterminants au paragraphe 4). L'intérêt de cette notion d'expression assertable est de simplifier la décomposition des phrases complexes et surtout de le faire en évitant de recourir à des notions sémantiques ou extra-linguistiques primitives. Ces décompositions

permettent donc d'expliciter ce qu'on pourrait appeler une sémantique des fonctions grammaticales, qu'on peut mettre en parallèle avec leur genèse en indo-européen.

1. Phrase simple, expression assertable, assertion, discours

Quelle unité d'analyse choisir, construire et comment lui faire correspondre une notion de phrase simple ? Je reprends la discussion là où je l'ai laissée à la fin de mon premier article pour aborder ces questions. Je reviendrai ici sur le caractère hétérogène des contraintes mises en jeu dans l'assertabilité et en particulier sur le fait que les contraintes sur la concaténation des éléments lexicaux n'en représentent qu'un aspect, ces dernières pouvant être décrites par des règles de linéarisation définies sur les structures applicatives qui sont à la fois les structures d'insertion lexicales et les structures d'interprétation.

Je ne tenterai pas de répondre dans ce paragraphe à la question de savoir s'il existe une notion générale de phrase dans les langues, c'est-à-dire si la convention typographique des séquences de mots séparées par des points recouvre une réalité indépendante des descriptions ou un consensus sur des conceptions grammaticales ou encore un consensus sur des conventions stylistiques.

Comme unité d'analyse, je définirai ici la notion d'expression assertable. Elle est relative aux opérations de dérivations linguistiques telles que je les définirai plus bas (réductions « appropriées » et identifications qui font partie de la réduction de l'opérateur de co-assertion décrite dans le paragraphe 3) et se construit dans des énoncés particuliers. C'est une notion qui résulte d'hypothèses et de règles pour satisfaire plusieurs sortes de contraintes empiriques. Elle a pour objet de rendre compte des aspects liés aux contraintes applicatives, c'est-à-dire liées à l'interprétation applicative de l'insertion lexicale telle que je la propose ici. Une séquence assertable est une application ordonnée d'opérateurs lexicaux (dans un sens précisé plus bas) susceptible d'être effectivement assertée moyennant qu'on lui applique une linéarisation et certains ajustements morphologiques. De plus, lorsqu'une expression assertable est assertée par un locuteur, elle devient une assertion, elle reçoit une morphologie verbale conjuguée relative au temps de l'assertion par le locuteur et subit les ajustements de concordance temps/mode à l'intérieur de ses sous-expressions assertables. Je ne définirai pas ici la notion d'assertion, qui nécessiterait un article à elle seule, mais j'utiliserai, dans les dérivations, des phrases, en supposant l'application des linéarisations réalisée et en ajoutant une morphologie verbale conjuguée aux expressions assertables¹. Ces dérivations devront donc être considérées comme des illustrations des points de grammaire dont il sera question et non comme des dérivations complètes. Il faut souligner que les opérations de réduction linguis-

1. Pour les lecteurs familiarisés avec le λ -calcul : les expressions assertables correspondent aux formes normales des termes typés qui ont un type dans le domaine des types des expressions assertables TAS (après l'application des réductions linguistiques et des réductions applicatives).

tique s'appliquent aux expressions assertables et non aux assertions. Par exemple, lorsque l'identification (analysée plus bas) concerne des arguments de statut non élémentaire, elle s'applique quand les opérateurs qui dominent ces arguments induisent des contraintes de modes différentes. On ne peut pas substituer *vienne* à *viendra* après *a annoncé* ni *viendra* à *vienna* après *ne souhaitait pas* dans (1) mais on peut avoir (2) et (4) :

- (1) *Marc a annoncé que Luc viendra ; Jean ne souhaitait pas que Luc vienne.*
- (2) *Marc a annoncé que Luc viendra, ce que Jean ne souhaitait pas.*
- (3) *Jean ne souhaitait pas que Luc vienne ; Marc a pourtant annoncé que Luc viendra.*
- (4) *Jean ne souhaitait pas que Luc vienne, ce que Marc a pourtant annoncé.*

Il faut également préciser que ce sont les expressions assertables qui sont typées. Il ne semble pas y avoir d'avantage à typer les assertions.

La notion d'expression assertable est indissociable de celle de contexte d'assertion. La structure d'interprétation d'une phrase, ou plus précisément ici, d'une expression assertable ou d'une de ses sous-expressions, peut faire intervenir le contexte où elle est assertée (l'énoncé). Par exemple, dans le contexte de (5), *espoir de Jean* doit être interprété comme *espoir de Jean que Marie reviendrait* alors que dans le contexte de (6), *espoirs de Jean* reste indéfini lexicalement, et s'interprète comme *espoirs de choses* :

- (5) *Jean a longtemps espéré que Marie reviendrait. L'espoir de Jean nous a étonnés.*
- (6) *Les espoirs de Jean nous étonnent parfois.*

Dans (5) comme dans (6), *espoir* est un opérateur à deux arguments, un argument de statut élémentaire *Jean* et un argument de statut non élémentaire, dans (5) *que Marie reviendrait*. Dans (6), on peut faire l'hypothèse que cet argument est un élément lexicalement indéfini qui a subi comme tel une réduction « appropriée ». Il reste à définir le statut syntaxique de cet élément phrastique, ce qui sera précisé plus bas.

Une différence essentielle entre la notion de phrase et celle d'expression assertable est que le pivot des premières est le verbe alors que le pivot des secondes est n'importe quel élément lexical de statut non élémentaire, par exemple les séquences assertables (7a) à (11a). De ce point de vue, la notion de discours définie dans [GEMP] a la même particularité. Ces expressions assertables ne sont pas élémentaires ; ainsi, on verra dans le paragraphe 3 que des expressions comme *une pomme* sont elles-mêmes des formes morphologiquement réduites d'expressions assertables :

- (7a) *(Un petit (crayon))*
- (8a) *(la description (de Jean, de ce paysage))*
- (9a) *(contre (attaque (notre, de la ville), Luc))*
- (10a) *(souvent (dort (Jean)))*
- (11a) *(souvent (quelque chose))*

Quand elles ne comportent pas un verbe, les séquences assertables deviennent des phrases si on leur ajoute une morphologie verbale conjuguée par l'intermédiaire d'un support de morphologie verbale, qu'on peut associer à l'opérateur de

niveau supérieur ou dans la plupart des cas à un autre opérateur (ce qui intervient sur des éléments de signification de l'assertion, par exemple (8b), (9b) (10b)). Comme support de morphologie verbale, *être* joue le rôle d'une sorte d'auxiliaire et peut être associé à un grand nombre d'opérateurs. Pour les autres, qui sont généralement des opérateurs à argument non élémentaire (adverbes, prépositions, opérateurs lexicalement indéfinis *quelque chose, chose*), on peut utiliser comme support de morphologie verbale un élément, qui porte une valeur aspectuelle en plus du temps exprimé par sa morphologie verbale, parmi la liste : *arriver, avoir lieu, intervenir, se produire, se dérouler, advenir, se faire, se passer*. Dans cet emploi de support de morphologie verbale, ces éléments n'ont pas de statut d'opérateur. Le temps qu'ils portent fait partie de la définition de l'assertion, que je ne définis pas ici. Les phrases (8b) à (11b) peuvent s'obtenir à partir des expressions assertables correspondantes :

- (7b) *Un crayon est petit*
- (8b) *La description de ce paysage est de Jean.
La description de Jean de ce paysage a eu lieu.*
- (9b) *Notre attaque de la ville a eu lieu contre Luc
Notre attaque de la ville contre Luc a eu lieu
L'attaque de la ville contre Luc est de nous.*
- (10b) *Jean dort souvent
Il arrive souvent que Jean dorme*
- (11b) *Quelque chose se produit souvent.*

Une séquence assertable élémentaire ne comporte pas de sous-expressions assertables, elle comporte donc un seul opérateur de statut non élémentaire. Les autres séquences assertables sont des combinaisons de séquences assertables. Pour des raisons sur lesquelles je reviendrai en détail par la suite, l'insertion des éléments lexicaux (de statut élémentaire ou non) dans les séquences assertables se fait en deux temps, de façon indirecte, par l'intermédiaire d'éléments lexicalement indéfinis qui sont ensuite identifiés et réduits. De ce fait, une séquence assertable élémentaire a au moins un élément lexicalement indéfini.

Certaines séquences assertables et assertions demeurent lexicalement indéfinies, par exemple (12) assertée hors contexte :

- (12) *Quelque chose est arrivé.*

Dans (12), *quelque chose* est à lui seul une expression assertable. Dans cette expression assertable, *quelque chose* est un opérateur à argument non élémentaire, c'est-à-dire a valeur dans tout le domaine des expressions assertables donc s'appliquant à une combinaison infinie d'opérateurs *quelque chose* puisque le nombre de sous-expressions assertables d'une expression assertable n'est pas borné *a priori*. On peut cependant « approximer » son expression syntaxique sans perte de valeur informative en considérant qu'il s'applique à lui-même

opérant sur son plus petit point-fixe fonctionnel ². Il suffira ici d'admettre que cette expression a une forme de clôture syntaxique qui a la valeur d'une application auto-référentielle de *quelque chose*, ce qui est une façon de dire qu'elle conserve sa valeur indéfinie lorsqu'elle est assertée hors-contexte. Certains contextes peuvent autoriser une réouverture de sa structure syntaxique et donc de sa structure d'interprétation, par exemple dans (13) :

(13) *Luc a attendu toute la journée. Quelque chose d'étrange est arrivé. Le départ de son avion a été annulé.*

Les règles et équations du système de type proposé en annexe permettent de lui assigner un type normalisé dans toutes les situations, non équivalentes linguistiquement, où elle peut être insérée. Le point important est de disposer d'une forme de clôture syntaxique compatible avec le fait que la structure d'interprétation d'un opérateur lexicalement indéfini peut être réouverte et redéfinie si cet opérateur est inséré dans un contexte qui l'autorise.

Les séquences assertables ont une définition inductive, qui correspond à la construction du domaine des types d'expressions assertables TAS proposé en annexe 2. Les expressions assertables sont les combinaisons d'opérateurs lexicaux qui ont un type dans TAS. Plus précisément, la construction de TAS fait intervenir une induction double, d'une part sur le nombre de sous-expressions assertables dont la combinaison est une expression assertable, d'autre part sur leur caractère partiellement indéfini (le nombre de sous-expressions « approximées »). Les règles de réduction applicatives pour obtenir un type dans TAS ne s'appliquent qu'après que les réductions linguistiques se soient appliquées. Un point fondamental de la notion de séquence assertable est que la construction de TAS, les règles de typage avec les équations particulières pour représenter l'approximation et les règles de réduction linguistique sont trois composantes interdépendantes.

La notion de séquence assertable se distingue de la notion de discours par plusieurs aspects dont deux principaux :

- la nature fonctionnelle des opérateurs lexicaux est différente ;
- avant leur insertion dans un énoncé particulier qu'on analysera en termes d'expressions assertables, les opérateurs lexicaux n'ont pas seulement un type syntaxique, ils ont aussi une étiquette de traits lexicaux.

De plus, la notion de discours ne permet pas de rendre compte d'inassertabilités qui résultent de violations de concordance de temps/mode, qui dépendent ici de la définition de l'assertion, distincte comme on l'a vu plus haut de la séquence assertable.

Du fait de différentes propriétés empiriques, en particulier leurs propriétés de restriction de sélection lexicale, les éléments lexicaux à arguments non élémen-

2. La notion de point-fixe fonctionnel et de famille de points-fixes a été définie par H. B. Curry dans sa logique combinatoire.

taires ont une portée qui peut varier selon le choix lexical d'autres éléments du contexte d'assertion. Avant leur insertion dans un énoncé particulier, les éléments lexicaux de statut non élémentaire doivent avoir un statut syntaxique « variable » (au sens intuitif du terme) qui tienne compte de différents paramètres permettant en général de désambiguïser leur niveau d'insertion dans un énoncé donné. Je reviendrai sur ces propriétés au paragraphe 3.

Auparavant, il faut préciser que l'idée intuitive de statut syntaxique variable, de types variables ou plus précisément de type du second ordre, permet de rendre compte du caractère variable de la portée d'un grand nombre d'éléments lexicaux mais aussi du fait que leur portée est généralement non ambiguë dans les énoncés particuliers en raison des restrictions de sélection lexicale des autres éléments lexicaux de cet énoncé (voir paragraphe 2). Ces types du second ordre permettent également de rendre compte des propriétés de réduction linguistique :

— le fait que des réductions « appropriées » s'appliquent à des éléments dont la structure syntaxique n'est pas finie et doit être fermée par l'opération d'« approximation » définie ci-dessus, parfois provisoirement, et parfois définitivement dans certaines conditions structurelles qui garantissent la conservation du contenu informatif des expressions assertables.

— le fait que les identifications doivent être contrôlées par un ordre non fini sur les types d'expressions assertables. On verra au paragraphe 3 qu'on a besoin d'une infinité dénombrable de types d'expressions assertables notamment pour spécifier la longueur d'une expression assertable (le nombre de sous-expressions assertables qu'elle comporte) dans une identification. TAS sert aussi de structure de contrôle pour l'identification. Également du fait que les identifications doivent pouvoir s'appliquer de façon imbriquée (certaines identifications nécessitent la résolution d'autres identifications).

On distinguera donc le statut syntaxique des éléments lexicaux selon qu'ils sont insérés dans un énoncé particulier ou qu'ils sont classés dans un dictionnaire. Dans le dictionnaire, ils auront un type contenant des variables qui spécifie le nombre d'arguments et leur caractère élémentaire ou à valeur dans le domaine des expressions assertables. Ils auront également une étiquette de traits lexicaux. D'un point de vue pratique, on pourrait considérer que ces étiquettes sont les colonnes des tables dans les classifications du L.A.D.L. et qu'un type du second ordre devrait être rajouté à ces éléments lexicaux. Ce statut syntaxique en partie variable autorise des relations de dépendance applicative non immédiates et non limitées *a priori*.

Dans un énoncé particulier, l'analyse de l'énoncé dans son ensemble doit permettre d'éliminer les variables et les λ -opérateurs (en les normalisant) des termes et des types et donc d'assigner un type (du premier ordre) de TAS à l'énoncé. Ce faisant, l'analyse produit la structure applicative de l'énoncé qui fournit simultanément sa structure d'interprétation et la preuve de son assertabilité. C'est l'analyse de l'énoncé qui définit l'ordre d'insertion et le type sans variable des éléments lexicaux.

Parmi les informations des étiquettes de traits lexicaux, il faut faire figurer le fait que les opérateurs à argument non élémentaire ne s'appliquent pas librement les uns sur les autres selon qu'ils ont ou non une morphologie verbale, ce dont rendent compte en partie les grammaires utilisant les parties du discours. Par exemple en disant que les adjectifs s'associent aux noms et pas aux verbes et que les adverbes s'associent aux verbes et pas aux noms. Mais du fait des réductions « appropriées », ces consignes sont très relatives comme on le verra plus bas.

Dans la plupart des théories linguistiques, les contraintes sur la formation des phrases sont décrites en termes de contraintes sur la concaténation soit des constituants syntagmatiques, soit des entités lexicales (mots ou expressions figées). Ces contraintes sont exprimées par l'intermédiaire de catégories auxiliaires ou de catégories lexicales « représentantes » d'une classe d'éléments ayant les mêmes propriétés, sous lesquelles se fait l'insertion lexicale. Les règles de formation de phrase ainsi obtenues sont ensuite associées à des transformations qui appartiennent des formes de phrases (éventuellement lexicalisées), c'est-à-dire des objets structurés par des contraintes de concaténation. Les composantes transformationnelles des différentes théories linguistiques procèdent de conceptions très différentes du fait de leurs conditions d'application. Pour certains, elles obéissent à des conditions métathéoriques supposées universelles, pour d'autres, elles obéissent à des conditions lexicales, un des enjeux de la grammaire étant la recherche d'organisations, locales ou non, internes aux lexiques des différentes langues. Les différentes notions de transformation ont tout de même un point commun : à partir de quelques opérations (permutations, déplacement de catégories syntagmatiques vides ou effacements), on accroît considérablement le nombre de phrases analysées, tout en conservant des grammaires d'insertion lexicale relativement simples. Ces composantes syntaxiques sont ensuite associées à des composantes sémantiques, elles aussi de conceptions variées.

Malgré leurs divergences, ces approches me semblent avoir un inconvénient commun, qui est de mettre en relation, à partir de composantes distinctes, des structures syntaxiques (en terme de catégories syntagmatiques ou d'expressions lexicales) qui expriment des relations de dépendance pertinentes pour rendre compte de l'ordre linéaire des mots et de certaines dépendances morphologiques qui en résultent avec les relations de dépendance pertinentes pour exprimer l'interprétation des phrases. Or il s'avère que les séquences de mots ou d'expressions lexicales qui sont intuitivement assertables, éventuellement moyennant un contexte approprié, ne sont toujours pas caractérisées et loin s'en faut.

Les règles de concaténation ne suffisent pas parce que du fait du caractère non explicite de la formulation en langue naturelle, la bonne formation met en jeu des contraintes d'interprétation. Ainsi (1), (3), (5), (7) ne sont pas assertables alors que (2), (4), (6), (8) qui ne diffèrent que par des mots appartenant aux mêmes catégories d'insertion lexicale sont assertables :

- (1) *Les murs de Jean sont choise.*
- (2) *Les murs de Jean sont prune.*
- (3) *Jean est venu mais pas bien que Luc est parti.*
- (4) *Jean est venu mais pas parce que Luc est parti.*
- (5) *How to Boticelli a Seiko-Reagan ?*
- (6) *How to Russell a Frege-Church ?*
- (7) *Les taches d'huile sont difficiles.*
- (8) *Les livres d'astro-physique sont difficiles.*

Certaines phrases ne sont reconnues comme telles par un locuteur que dans un contexte qui permet la reconstruction d'une ellipse anaphorique. Ainsi (7) devient assertable dans le contexte de (9) :

- (9) *Jean a passé la matinée à enlever les taches d'huile du canapé.*
Les taches d'huile sont difficiles Ø.

La bonne formation met donc en jeu différentes formes d'ellipse, locales à la phrase ou anaphoriques dans le cadre de l'énoncé. Une unité d'analyse contextuelle est nécessaire pour rendre compte de relations anaphoriques mais aussi de relations de relations comme on le verra ci-dessous. Les différentes notions existantes de phrase et la notion de discours ne sont pas suffisantes pour rendre compte d'autres problèmes empiriques qu'on examinera ensuite.

La notion de discours, qui tient compte de certaines des dépendances mettant en jeu l'énoncé, est proposée dans [GEMP] mais les exemples (7) et (9) ne peuvent être convenablement décrits en faisant intervenir cette notion et la notion d'opérateur qu'elle comporte. *Difficile* est un opérateur qui requiert un argument de statut élémentaire et un argument phrastique. Dans (7), il n'est pas possible de reconstruire l'argument phrastique manquant à partir d'une identification avec une combinaison de mots de statut phrastique. Bien qu'un tel répertoire n'existe pas actuellement, on peut supposer qu'il serait possible d'établir un répertoire de réductions « appropriées » où figureraient des opérateurs avec des listes d'arguments implicites pour des sélections données d'arguments explicites. Par exemple, *difficile* s'il a *livre* comme premier argument admet la réduction appropriée de *à lire*. Si un tel répertoire existait, on n'y trouverait pas de réduction appropriée pour (7) et on ne pourrait pas lui assigner un type de discours. En revanche dans (9), il est possible d'induire automatiquement l'hypothèse d'une réduction sous identification dans le contexte. L'analyse syntaxique de ce contexte selon la méthode de [GEMP] permettrait de constater l'existence de deux candidats pour cette identification parce que ce contexte contient deux expressions pour lesquelles on déduirait un type de discours : *enlever les taches du canapé* et *Jean a passé la matinée à enlever les taches du canapé*. La méthode de [GEMP] ne permettrait pas de choisir entre ces deux candidats et fournirait deux structures de reconstruction à (9) pour lesquelles on pourrait déduire un type de discours. En l'absence d'informations lexicales associées aux types, concernant à la fois les traits de sélection et les contraintes de « projection », les deux structures reconstruites correspondraient à des séquences de mots inassertables :

Jean a passé la matinée à enlever les taches du canapé. Les taches d'huile sont difficiles à enlever les taches du canapé.

Jean a passé la matinée à enlever les taches du canapé. Les taches d'huile sont difficiles à Jean passer la matinée à enlever les taches du canapé.

Pour rendre compte de l'assertabilité de (9) en lui associant sa structure d'interprétation correcte, il faut disposer de l'information concernant le trait de projection lexicale de *difficile* qui permet d'identifier son argument élémentaire *tache* avec l'argument élémentaire de son argument phrastique. Il faut également être capable de rendre compte du fait que la reconstruction de l'argument non élémentaire de *difficile* fait intervenir une identification des deux occurrences de *tache*, ce qui permet ensuite d'identifier l'argument phrastique de *difficile* à : \emptyset *enlever du canapé*. Dans cet exemple, la reconstruction d'un argument fait intervenir trois opérations d'identification qui doivent être hiérarchisées entre elles. Cet exemple montre donc qu'il existe des réductions qui mettent en jeu d'autres réductions et que les éléments lexicaux concernés par ces réductions peuvent occuper des positions d'insertion arbitrairement distantes les unes des autres.

L'intérêt de la notion d'opérateur lexical telle qu'elle est mise en jeu dans les expressions assertables pour les reconstructions anaphoriques apparaît également dans des exemples comme :

(10) *Jean a parlé à Marc. Cette erreur nous a étonnés.*

On peut étendre la méthode syntaxique proposée dans [GEMP] pour inférer directement à partir de la formulation de (10) que le fait que Jean a parlé à Marc était une erreur. *Erreur* est un opérateur à argument phrastique. Pour assigner un type de discours à (10), la méthode de [GEMP] consiste à faire l'hypothèse que l'argument manquant a subi une réduction. La présence de l'élément anaphorique *cette* et le fait qu'on peut assigner un type de discours à : *Jean a parlé à Marc* permet d'induire automatiquement que l'argument phrastique manquant provient d'une identification dans le contexte de la co-assertion. La co-assertion ne contient qu'un élément ayant un type de discours, on peut donc identifier ce discours (*Jean a parlé à Marc*) à l'argument réduit à une forme zéro de *erreur*. On peut également appliquer cette analyse pour rendre compte de l'ambiguïté de (11) ou de (12) (en dehors d'un contexte qui permette une désambiguïsation éventuelle) :

(11) *Jean a parlé à Luc et est rentré chez lui. Cette erreur nous a étonnés.*

(12) *Paul a déclaré que Jean a parlé à Luc. Cette erreur nous a étonnés.*

Mais les types simples qui sont définis dans [GEMP] ne permettent pas de rendre compte du fait que la combinaison de discours qui joue le rôle d'antécédent dans des exemples analogues est désambiguïsée par l'existence d'autres identifications. Ces identifications doivent être appliquées à un niveau plus interne :

Paul a déclaré que Jean a parlé à Luc. L'erreur de Luc nous a étonnés.

Marc a nié que Paul avait déclaré que Jean avait parlé à Luc. L'erreur de Marc nous a étonnés.

Comme le nombre d'expressions assertables qui peuvent entrer dans la composition d'un énoncé n'est pas limité *a priori*, on a besoin d'une infinité

dénombrable de types pour distinguer la longueur d'un antécédent phrastique. J'emploie le terme « longueur d'une expression assertable » pour désigner le nombre de sous-expressions assertables qu'elle contient.

2. Relations de dépendances lexico-syntaxiques à trois termes ou plus

Aucune grammaire de catégories auxiliaires ne peut rendre compte du fait que certaines entités lexicales ont une portée, ou plus précisément ici une structure d'interprétation, qui dépend non seulement de relations de dépendance lexicales mais aussi de relations de dépendances lexicales et/ou de relations anaphoriques dans l'énoncé.

J'indiquerai dans ce paragraphe comment ces relations peuvent faire intervenir un nombre *a priori* non limité de termes qui ne peuvent être décrites dans aucune grammaire existante. Le principal intérêt de la représentation applicative de l'insertion lexicale que je propose est de rendre compte de cette propriété sous ses multiples aspects empiriques, en ramenant ces relations de dépendance à des relations applicatives simples. Ceci n'est possible que pour des combinaisons lexicales concrètes. Aucune définition de la bonne formation en termes de phrase ou d'unité d'analyse abstraite ne peut être proposée pour une langue dans son ensemble, *a fortiori* pour l'analyse des langues en général. On verra dans les paragraphes suivants comment cette représentation applicative de l'insertion lexicale s'avère renforcer l'intégration de la métalangue dans la langue.

Même les ellipses les plus simples, les réductions « appropriées », mettent en défaut les grammaires de constituants, aussi bien pour décrire la bonne formation que pour décrire les structures d'interprétation (les relations de portée). Des substantifs qui n'ont même pas un statut prédicatif peuvent occuper en position adverbiale du fait de réductions « appropriées » de prépositions ou de locutions prépositionnelles et du choix lexicalement approprié du verbe dans (2), (4), (6), (8), (10) mais (1), (3), (5), (7), (9) ne sont pas assertables :

- (1) *Jean a commandé un crayon table pinceau.*
- (2) *Jean a commandé un lapin sauce madère.*
- (3) *J'ai dormi toute une montre.*
- (4) *J'ai dormi tout un cadran d'horloge.*
- (5) *J'ai couru tout un ascenseur.*
- (6) *J'ai couru toute la rue de Seine.*
- (7) *Marie parle oncle.*
- (8) *Marie parle bébé.*
- (9) *Peter flies banana.*
- (10) *Peter flies Pan Am.*

Ce qui est vrai pour les noms l'est également pour les autres parties du discours, dont on peut « détourner » les fonctions grammaticales ou non selon le choix lexical des autres éléments de la phrase. Par exemple, *large* peut être employé adverbialement dans (11) mais pas dans (12) qui, de ce fait, n'est pas assertable.

Pour la même raison, les phrases (13) sont ambiguës mais pas les phrases (14) (la notation (a/b) signifie qu'on peut substituer *b* à *a* dans l'énoncé où se trouve la parenthèse et aussi dans la structure applicative sous-jacente) :

- (11) *Jean a compté large.*
- (12) *Jean a dormi large.*
- (13) *Jean (tricote / aime / achète / commande) ses chandails larges.*
- (14) *Jean (lave / range / use / salit) ses chandails larges.*

En position adverbiale, les adjectifs n'en conservent pas moins leurs restrictions de sélection lexicale, leur « portée adverbiale » dépend alors du choix lexical des autres éléments de la phrase, par exemple (15) et (16) ou (17) et (18). Mais les portées adverbiales des adjectifs sont incompatibles avec une coordination et rendent (19) et (20) inassertables sous les interprétations correspondantes :

- (15) *Jean a bu son café chaud.*
- (16) *Jean a bu son café heureux.*
- (17) *Jean a quitté Paris heureux.*
- (18) *Jean a quitté Paris polluée.*
- (19) *Jean a bu son café chaud et heureux.*
- (20) *Jean a quitté Paris heureux et polluée.*

Les prépositions ont également des propriétés de portée variable qui dépendent de leurs restrictions de sélection lexicale ; (21) ou (23) est ambiguë mais pas (22) ni (24) :

- (21) *Jean a quitté Luc en bonne santé.*
- (22) *Jean a quitté Paris en bonne santé.*
- (23) *Jean aime ses chandails en laine.*
- (24) *Jean lave ses chandails en laine.*

Les prépositions qui ont un statut d'opérateur à argument non élémentaire ont des propriétés de portée variable particulièrement courantes dans les phrases où figurent également une nominalisation dont un argument n'est pas explicite. Ce dernier peut souvent être identifié à l'argument élémentaire de la préposition du fait des propriétés de « projection lexicale » et du fait des propriétés de restrictions de sélection de celle-ci. Par exemple :

- (25) *Jean a mentionné nos attaques de la ville contre Marc.*
- (26) *Jean a oublié nos attaques contre Marc.*
- (27) *Jean a mentionné nos attaques contre Marc.*

Dans (25), l'opérateur *attaques* est instancié par ses deux arguments élémentaires *nos* et *ville*. *Nos attaques de la ville* a le statut d'une séquence assertable sur laquelle *contre* peut opérer. Sous cette structure applicative, (25) s'interprète comme : *Jean a mentionné quelque chose. Cette chose est nos attaques de la ville contre Marc.* Mais *contre* peut également opérer sur *a mentionné* lui-même opérant sur *Jean* et *nos attaques de la ville* comme argument élémentaire et non élémentaire respectivement. Sous cette structure applicative, (25) s'interprète comme : *Jean a mentionné quelque chose contre Marc. Cette chose est nos attaques de la ville.*

Par opposition à *mentionner*, *oublier* a une très faible vraisemblance d'occurrence dans le domaine de sélection lexical de *contre* et bien que ces deux verbes fassent partie du domaine de sélection syntaxique de *contre*, (27) est ambiguë alors que (26) ne l'est pas. Dans (26), *contre* n'opère pas sur *oublier* (on pourrait cependant construire un contexte qui force *contre* à opérer sur *oublier* du fait qu'il est compatible syntaxiquement) ; par contre il est lexicalement très vraisemblable sur *attaques*. Pour cette raison, le second argument de *attaques* a subi une réduction à une forme zéro après identification avec l'argument élémentaire *Marc de contre*.

Les verbes peuvent avoir des arguments de statut élémentaire ou non, les adjectifs ont également des statuts syntaxiques différents et ont de ce fait des propriétés de portée différentes, par exemple *foncé* et *échancré* :

Jean a un chandail rose foncé.

Jean a un chandail rose échancré.

On produit ses séquences inassertables en forçant un adjectif à argument élémentaire comme *échancré* dans la position de linéarisation d'un opérateur à argument non élémentaire :

Jean a un chandail d'un rose échancré clair.

Certaines prépositions peuvent être employées dans une fonction adjectivale en étant « déterminées » par des adverbes, qui obéissent à des restrictions de sélection lexicale mettant en jeu le choix du verbe, de l'adverbe et de la préposition ; le dernier exemple est inassertable :

Jean est (très / de plus en plus / souvent / depuis toujours) contre la démagogie.

Jean lutte (de plus en plus / souvent / depuis toujours) contre la démagogie.

Jean lutte très contre la démagogie.

On a vu que des éléments lexicaux appartenant à des parties du discours différentes peuvent jouer le rôle d'une partie du discours encore différente. Par exemple, les noms et les adjectifs peuvent être employés adverbialement et les prépositions adjectivement. Les mots appartenant à une même partie du discours ont des propriétés syntaxiques différentes, par exemple les prépositions et les adjectifs. Il semble donc que, pas plus que les catégories syntagmatiques, les parties du discours ne soient les plus pertinentes pour rendre compte des contraintes d'assertabilité et d'interprétation du français.

Certaines grammaires de constituants font dépendre la génération de la structure syntaxique d'un membre de phrase d'une relation de dépendance lexicale entre deux éléments de cette phrase. Il s'agit par exemple de la grammaire dite « lexicale fonctionnelle » ou des Tree Adjunct Grammars de A. Joshi ou encore de la grammaire logique de J. Y. Morin. La génération d'un sous-arbre syntagmatique dépend d'un « pattern matching » de traits booléens. Ces grammaires permettent donc de subordonner l'insertion d'un constituant syntagmatique, et son sous-arbre, à un agrément lexical entre cet élément et un élément d'une catégorie syntagmatique donnée, qui n'est pas nécessairement

contiguë à cet élément, du point de vue des relations de dépendance syntagmatique. Les conditions lexicales sur la structuration des sous-arbres sont limitées par la nature des relations de dépendance context-free ou context-sensitive des catégories syntagmatiques. En plus des inconvénients descriptifs des catégories syntagmatiques comme catégories d'analyse, déjà mentionnés, le type de récursion qu'autorise ces grammaires est insuffisant. J'essaierai d'en donner l'idée intuitive sans entrer dans les détails techniques. D'un point de vue purement formel, ces grammaires ne peuvent rendre compte de situations où l'insertion d'une sous-structure dépend d'une relation de relation(s) de dépendance lexicale (c'est-à-dire les relations de dépendance lexicale imbriquées, à trois termes ou plus, mentionnés dans ce paragraphe). A moins, bien sûr, de faire intervenir un dispositif de contrôle³ s'appliquant aux « pattern matching » notamment pour les hiérarchiser, qui aurait un statut métalinguistique dans la grammaire. Il serait bien sûr possible de lui supposer un caractère cognitif universel pour le justifier, des situations similaires se sont déjà produites en grammaire générative. Mais cette situation peut paraître tout aussi arbitraire que de prétendre qu'on peut programmer une grammaire algébrique avec un algorithme rationnel du moment qu'on programme également un dispositif de contrôle du nombre de lettres engendrée dans chaque position syntaxique de la grammaire rationnelle.

Il me semble plus intéressant de préciser la nature formelle des relations de dépendance mises en jeu par les mécanismes de formation et d'interprétation propres aux langues naturelles en construisant le statut syntaxique des éléments lexicaux dans les énoncés particuliers, ce qui suppose de pouvoir représenter les dépendances lexicales à l'intérieur des dépendances syntaxiques. Ces dernières doivent rendre compte simultanément de la bonne formation et de la (ou les) structure d'interprétation sous laquelle un énoncé est assertable. Un des problèmes est le caractère variable de la portée de certains éléments lexicaux selon le choix lexical d'autres éléments lexicaux de l'énoncé. Les propriétés de portée des mots ne dépendent pas de leur appartenance à une catégorie syntagmatique ou à une partie du discours mais d'un statut applicatif initial et de relations de dépendance lexicales, qu'on peut exprimer à l'intérieur de ce statut applicatif. Un autre problème est le caractère elliptique de la formulation en langue naturelle, ce qui rend les règles de bonne formation syntaxiques tributaires des règles de réductions linguistiques. Or ces mécanismes de réduction linguistiques font intervenir des contraintes d'interprétation, notamment dans le cadre des relations anaphoriques, qui interfèrent avec des dépendances

3. On peut imaginer d'utiliser différents procédés de programmation pour réaliser un tel dispositif de contrôle. Par exemple, pour celles de ces grammaires qui sont implémentées en Prolog, on pourrait utiliser une fonction « frozen » pour retarder l'unification de traits booléens jusqu'à ce que la construction de l'arbre de dérivation ait atteint un troisième terme à unifier avec les deux précédents. Mais il resterait à décrire les conditions dans lesquelles cette fonction « frozen » pourrait s'appliquer.

lexicales. Je donnerai ici quelques exemples de dépendances syntaxiques mettant en jeu des dépendances anaphoriques et lexicales à plusieurs termes pour illustrer cette question difficile.

Avant d'en venir au vif du sujet, je donnerai un exemple relativement plus simple faisant intervenir : assertabilité, relation anaphorique et relation lexicale entre l'aspect de l'opérateur de niveau supérieur et le choix du déterminant d'un argument de son argument.

Les verbes de la table 1 de *Méthodes en syntaxe* de Maurice Gross sont des verbes à argument phrastique dont le sujet est obligatoirement celui de leur argument. Ils sont donc analysés ici comme opérateurs à un argument de statut non élémentaire. Ceci rend compte de l'inassertabilité d'exemples comme (1) et (2) hors contexte et de l'assertabilité de (3) où *Jean* linéarisé en tête de phrase est le premier argument de *écrire* :

- (1) *Jean (achève | cesse | doit | a failli | tente | peut).*
- (2) *Jean (achève | cesse | doit | a failli | tente | peut) que Luc écrive un livre.*
- (3) *Jean achève d'écrire un livre.*

Jean cesse peut devenir une phrase si cette séquence est insérée dans un contexte d'assertion qui permet l'identification de l'argument phrastique de *cesse* :

- (4) *Jean a écrit des livres pendant trente ans. Il a cessé aujourd'hui Ø.*
Ø = *d'écrire des livres*

On remarque que cette propriété de réduction anaphorique n'est pas générale pour les verbes de cette table. D'une part, elle ne s'applique pas à tous les verbes et d'autre part, lorsqu'elle s'applique, elle dépend d'une valeur aspectuelle subtile, qui résulte par exemple de la relation entre la valeur aspectuelle exprimée par les verbes de cette table et le choix du déterminant de l'argument de son argument ; ainsi cette réduction n'est pas possible dans (5) alors qu'elle l'est dans (4), (6) :

- (5) *Jean a écrit des livres pendant trente ans. Il a achevé aujourd'hui.*
- (6) *Jean a écrit un livre pendant trente ans. Il a achevé aujourd'hui.*

Il existe une forme de relativation restrictive qui induit le subjonctif dans la relative, avec une interprétation différente de la relative à l'indicatif si cette dernière est possible, qu'elle soit appositive ou restrictive. On observe une relation de dépendance entre le mode de la principale, le mode de la relative et le choix du déterminant de l'antécédent de la relative. On peut asserter (7) et avec une interprétation différente (9), (10) mais pas (8) (on remarquera que l'interprétation de *un* n'est pas la même dans (9) et dans (10) :

- (7) *Jean ne m'a pas apporté de livre qui soit dans ma bibliothèque.*
- (8) *Jean ne m'a pas apporté de livre qui est dans ma bibliothèque.*
- (9) *Jean ne m'a pas apporté un livre qui est dans ma bibliothèque.*
- (10) *Jean ne m'a pas apporté un livre qui soit dans ma bibliothèque.*

Il existe des relations de dépendance qui mettent en jeu une interférence entre les propriétés de « projections lexicales » de deux opérateurs à argument non élémentaire pour la réduction d'un argument élémentaire d'un de ces deux

opérateurs ou figurant dans leur descendance. Dans (11), *Jean* argument élémentaire de *a étonné* n'est pas identifié à l'argument élémentaire de *décision* (l'argument non élémentaire de *a étonné*). En l'absence d'une identification anaphorique dans un contexte de niveau supérieur, l'argument élémentaire de *décision* provient d'une réduction « appropriée » de l'indéfini *quelqu'un*. Dans (12), la propriété de projection lexicale de *a angoissé* permet l'identification de son argument élémentaire *Jean* avec l'argument élémentaire de *décision* mais cette identification n'est pas obligatoire, contrairement à l'insertion directe d'une infinitive sous un verbe à complétive, et l'argument élémentaire de *décision* peut également provenir de la réduction d'un indéfini *quelqu'un*, ce qui rend (12) ambiguë. Pour les mêmes raisons, (13) est également ambiguë :

- (11) *La Ø décision de Ø partir a étonné Jean.*
- (12) *La Ø décision de Ø partir a angoissé Jean.*
- (13) *Luc a annoncé la Ø décision de Ø partir.*

On remarque que l'argument élémentaire de l'infinitive s'identifie à l'argument élémentaire de son opérateur immédiat (verbe ou nom). Pour des raisons que je ne peux développer ici, il semble que cela soit dû à ce qu'on pourrait appeler le « mode d'assertion » des infinitives.

L'interférence des contraintes de « projection » peut mettre en jeu l'application de trois opérateurs à argument non élémentaire, par exemple une préposition, un verbe et un nom dans (15), (16) et (17) :

- (14) *La décision de Jean de Ø venir nous a étonnés.*
- (15) *Pour Jean, la décision de Marc de Ø venir a été facilitée par cet événement.*
- (16) *Pour Jean, la Ø décision de Ø venir a été facilitée par cet événement.*
- (17) *Pour Jean, la Ø décision de Ø venir nous a angoissés.*

L'argument élémentaire de *venir* est identifié à celui de *décision*, *Jean* dans (14) et *Marc* dans (15). Dans (16), l'argument élémentaire de *décision* peut être identifié à celui de *pour*, *Jean*, mais cette identification n'est pas obligatoire et un contexte de niveau supérieur pourrait en favoriser une autre. Dans (17), l'argument élémentaire de *décision* ne peut être identifié à celui de *pour* mais l'est avec celui de *angoissés*, *nous*.

Il semble exister des symétries entre l'interférence des contraintes de « projection lexicale » et l'interférence des inductions de modes (indicatif ou subjonctif) pour l'application de deux et trois opérateurs à argument phrastique mais les données concernées sont encore insuffisamment étudiées.

D'autres formes de dépendances syntaxiques mettant en jeu des relations de dépendance lexico-syntaxiques à plus de trois termes sont analysées par Danielle Leeman dans : « Les compléments adverbiaux de phrase de type *en toute N* : *en toute objectivité*, *Paul est un crétin* » (*Mélanges Lerond, Linx* 1990), où elle montre, entre autres choses, que des aspects très fins d'interprétation de *toute* procèdent des contraintes lexico-syntaxiques qui régissent l'existence de ce type de séquence adverbiale.

J'espère avoir indiqué l'intérêt d'avoir une représentation homogène de l'insertion lexicale et des règles de dérivation linguistique, indépendamment de

toute hypothèse sur l'origine cognitive ou les mécanismes neurologiques du langage.

Le choix de la représentation applicative de l'insertion lexicale et des règles de réductions linguistiques, qui doivent rendre compte notamment des ellipses anaphoriques, dans un λ -calcul du second ordre tel qu'il est proposé ici, a pour objectif de répondre à la question de savoir quelle est la nature des mécanismes de formation et d'interprétation propres aux langues naturelles.

Il semble inévitable d'admettre que tous les éléments lexicaux de statut non élémentaire doivent avoir une description de sous-catégorisation lexicale si on veut rendre compte de toutes les structures d'interprétation et caractériser l'assertabilité d'une langue dans son ensemble. Savoir si cette tâche est faisable pratiquement est une autre question.

3. Interprétation lexicale et applicative de l'opérateur de co-assertion et interprétation anaphorique de la réduction de cet opérateur

Pour Zellig Harris, le processus même d'identification des éléments coréférents est interne aux langues. Cette hypothèse est fondamentale, et contraste avec les autres conceptions de la coréférence selon lesquelles des éléments lexicaux sont coréférents et ont les propriétés linguistiques propres aux différentes langues si les référents que ces éléments dénotent, référents dans un univers réel ou objets mentaux, selon les théories, sont identiques. La connaissance de leur identité serait donc antérieure à la formulation linguistique ou à son interprétation. Ainsi en grammaire générative, on analyse les propriétés anaphoriques (certaines sont syntaxiques) mais la coréférence elle-même, et en particulier la localisation dans les énoncés des éléments coréférents, n'est pas produite par la grammaire. Ces analyses utilisent une co-indexation des éléments coréférents et tentent de décrire les propriétés des phrases où des éléments sont co-indexés mais la façon dont cette co-indexation serait produite n'appartient pas à la grammaire. La référence est parfois décrite comme une « aptitude » qu'aurait le locuteur à identifier une expression lexicale à « un fragment de réalité », son analyse est renvoyée à l'analyse des mécanismes cognitifs. Cette conception de la référence est issue de celle qui a été longuement développée par les philosophes du langage. L'hypothèse de Harris exclut elle la notion de référence au sens de la philosophie du langage, qui fait appel à une ou des réalités extérieures au langage. La coréférence doit être analysée à partir d'identifications d'éléments lexicaux, ces identifications et leur localisation devant être produites par la grammaire.

Je reprends cette hypothèse mais le traitement grammatical qu'en propose Harris ne me semble pas pouvoir être maintenu pour différentes raisons empiriques et techniques que j'exposerai dans ce paragraphe et qui se résument à deux aspects non indépendants :

— L'identification et la localisation seront produites au niveau applicatif et non au niveau des linéarisations.

— L'identification s'applique à des éléments lexicaux de statut élémentaire ou non et notamment ici à des éléments qui ne sont que partiellement instanciés ou qui n'ont pas le même type (i.e. le même statut syntaxique) mais dont les types sont ordonnés dans le domaine des types d'expressions assertables (TAS, voir annexe).

L'opérateur de co-assertion sera toujours noté ici « ; » bien que son interprétation et ses propriétés soient différentes de celle de l'opérateur qu'a introduit Zellig Harris (j'utiliserai aussi « . » comme variante typographique mieux adaptée aux usages contemporains, dans des exemples de dérivation). Ici « ; » sera considéré comme un opérateur lexical ayant comme tel certaines propriétés particulières, comme le fait d'autoriser l'identification et la réduction d'arguments de ses arguments. Cette propriété appartient aussi à d'autres opérateurs à argument non élémentaire, on l'a vu par exemple pour les propriétés de « projection lexicale » des verbes à complétive. Dans le cas de « ; », cette propriété prend des formes et une ampleur particulières : « ; » autorise l'identification de deux expressions lexicales identiques ou partiellement identiques qui co-occurrent dans la descendance de ses deux arguments. Cette identification autorise la réduction morphologique, sous forme zéro mais aussi sous différentes formes pronominales de l'élément identifié dans son deuxième argument. Selon la position syntaxique de cet élément, la forme réduite pouvant lui être substituée peut prendre des formes pronominales différentes. Il existe des contraintes syntaxiques sur l'identification et d'autres sur l'application des réductions c'est-à-dire ici sur la substitution des formes réduites dans la structure applicative d'un « ; ». Ces contraintes sont formulées et hiérarchisées sous forme d'expressions conditionnelles emboîtées dans une λ -expression qui exprime l'élément lexical « ; » et qui s'élimine au cours de son évaluation (réduction du λ -calcul) lorsqu'elle est insérée dans un contexte particulier comme dans le cas de tous les autres opérateurs lexicaux (annexe). L'expression de l'opérateur lexical « ; » est donc une fonction composée avec des sous-fonctions exprimant les mécanismes d'identification et leur contrôle, la localisation ainsi que différentes contraintes sur les relations anaphoriques.

Je considère donc ici que l'interprétation de coréférence résulte du fait d'asserter dans un même énoncé deux expressions d'une certaine forme (expressions assertables selon la définition de L.) présentant des répétitions d'éléments lexicaux, eux aussi d'une certaine forme syntaxique. Par exemple, c'est la répétition du mot *Jean* ou de la séquence *une pomme* qui crée la possibilité pour le locuteur d'identifier ces deux occurrences et ce faisant de les interpréter comme faisant référence à un même individu :

(1) *Jean est venu hier. Jean a téléphoné.*

(2) *Marc a mangé une pomme. J'avais acheté une pomme.*

« ; » crée un contexte, il s'agit du contexte d'assertion dont il a été question à propos de la définition des expressions assertables, qui autorise l'interprétation de coréférence si les conditions sont réunies. Comme les réductions ne sont pas obligatoires de façon générale, l'identification qui produit l'interprétation de coréférence ne l'est pas non plus. L'opérateur « ; » permet de lier (au sens intuitif

et au sens du λ -calcul) l'interprétation des énoncés, c'est-à-dire d'interpréter une combinaison d'expressions assertables comme une seule expression assertable. C'est le fait de co-asserter deux expressions assertables qui introduit une certaine fonction autorisant l'identification d'occurrences de sous-expressions. Cette fonction produit l'interprétation de coréférence en « liant » (au sens du λ -calcul) les variables correspondant aux deux instanciations identifiées. Il est très important dans la méthode proposée ici que les identifications concernent des instanciations et non les variables. Cette identification donne lieu à différentes formes de réductions morphologiques, dont différents pronoms, qui sont substitués à l'expression lexicale (l'instanciation) identifiée. Par exemple, (3) et (4) sont des formes réduites de (2), elles ont toutes les trois la même structure applicative, syntaxique et d'interprétation. Elles ne diffèrent que par des formes de linéarisation et des morphologies différentes de formes réduites :

(3) *Marc a mangé une pomme, je l'avais achetée.*

(4) *Marc a mangé une pomme, que j'avais achetée.*

Je montrerai au paragraphe 4 comment dériver les relatives appositives et restrictives à partir du même mécanisme d'identification. Cette analyse fait intervenir celle des déterminants, qui feront également l'objet de ce paragraphe suivant, ce qui me permettra de donner la structure applicative des exemples précédents. Auparavant je montrerai pourquoi le traitement de la coréférence ne peut faire intervenir les linéarisations (c'est-à-dire des opérations sur la concaténation des éléments lexicaux).

Dans [GEMP] et [TLI] 5., Harris analyse la coréférence comme résultant de trois propriétés : a) la répétition de mots ; b) l'existence de mots qui forment la localisation dans les formes linéarisées d'autres mots comme (en français) : *le premier, le suivant, le précédent, ce dernier*, etc. ; c) l'existence de mots qui forment l'identité d'autres mots ; cette identité impliquerait pour Harris l'identité de leurs référents : « *same means 'has the same individual referent'* » ([TLI] 5.3). On verra plus bas qu'en fait, l'opérateur lexical *same* ou des opérateurs synonymes comme *est identique à* ne réalisent pas plus l'identification sous-jacente à la relativation ou d'autres opérations anaphoriques que la formulation en langue naturelle de la multiplication ne permet de réaliser une multiplication dans l'exemple de Péano que j'ai mentionné dans mon premier article.

Pour Harris, la relativation serait produite par une propriété particulière des linéarisations, l'interruption, ce mécanisme permettant à l'auditeur de localiser l'antécédent d'un pronom par sa contiguïté avec ce dernier dans la linéarisation de surface. Par exemple, en transposant sa dérivation au français, on aurait :

Un garçon a disparu ; un garçon est petit ; le premier garçon est le même que le garçon suivant

Un garçon — le premier est le même que le suivant — un garçon est petit — a disparu
Un garçon, qui est petit, a disparu

La relation anaphorique entre un pronom relatif et son antécédent résulte de la situation où deux mots identiques se retrouvent contigus par ce procédé d'interruption. J'énumérerai quelques-uns des inconvénients de cette analyse :

● *Le premier garçon est le même que le suivant* ne peut avoir un statut de discours de base, ne peut être asserté hors-contexte et sa dérivation ferait elle-même intervenir une dérivation anaphorique.

● Les propriétés de restriction de sélection des mots du vocabulaire qui expriment l'identité (en français *égale à*, *identique à*, *même que*, etc.) ne rendent pas compte des conditions dans lesquelles se réalise une relation anaphorique. Ceci se manifeste de plusieurs façons ; par exemple, on peut très bien asserter la phrase ambiguë :

(5) *Luc a disparu, Jean a fait son apparition, le premier est le même que le second.*

qui peut être interprétée soit comme un changement d'identité pour un même individu, soit comme une contre vérité (ce qui ne l'empêche pas d'avoir un sens). Il est bien évident que même sous l'interprétation où *Jean* et *Luc* ont le même référent, aucune relativation n'est possible à partir de (5). Un autre aspect du fait que les mots utilisés par Harris ne réalisent pas l'identification sous-jacente aux relations anaphoriques est qu'un même élément lexical qui co-occure en ayant deux statuts applicatifs (types) différents ne peut être source non seulement de relativation mais plus généralement de relation anaphorique que dans des conditions bien particulières (précisées par la relation d'ordre sur les types de TAS). Par exemple, les deux occurrences de *quelque chose* ne peuvent pas être identifiées, au sens d'une identification anaphorique, dans (6) et bien que rien dans les grammaires de Harris n'empêche de les dériver, (7) et (8) sont inassertables :

(6) *Jean a mangé quelque chose ; que Luc est venu est quelque chose.*

(7) *Jean a mangé quelque chose qui est que Luc est venu.*

(8) *Luc est venu, chose que Jean a mangée.*

● La propriété d'interruption s'avère résulter elle-même de l'anaphore (que cette dernière se manifeste sous la forme d'une relativation ou sous une autre forme de pronominalisation) ; elle ne s'applique pas sinon. Ainsi, on peut asserter (10) mais difficilement (11) :

(9) *Jean a décidé de revenir avec nous ; (Jean | Marc) partira la semaine prochaine*

(10) *Jean — il partira la semaine prochaine — a décidé de revenir avec nous.*

(11) *Jean — Marc partira la semaine prochaine — a décidé de revenir avec nous.*

● L'interruption ne peut rendre compte de la différence entre relativation restrictive et relativation appositive. Il faudrait supposer qu'elles correspondent à deux phénomènes syntaxiques différents.

● Seuls les opérateurs lexicaux à argument non élémentaire induisent des contraintes de mode qui subordonnent de façon particulière le « mode d'assertion » de leur argument, or une relative peut recevoir un tel mode (subjonctif) subordonné à celui de la principale, comme dans :

Apporte-moi un livre qui soit sur les Papous.

On ne peut pas en rendre compte si c'est une propriété de linéarisation qui est responsable de la formation de la relative. En revanche, on peut l'analyser en considérant que l'opérateur de co-assertion ne porte pas sur des discours mais

sur des expressions assertables et que l'identification qui fait partie de sa structure d'interprétation peut s'appliquer sous des modes d'assertion différents, ces modes étant induits par des contraintes lexico-syntaxiques comme dans le cas général.

• Dans [TLI] 5.3, Harris dit que, en dehors des relations anaphoriques dans le cadre des relatives, la localisation de l'antécédent doit être explicitée par des pronoms qui font référence à l'ordre des mots ou par une dépendance morphologique genre / nombre, sinon l'auditeur ne peut pas localiser l'antécédent et la phrase résultante est ambiguë. Pour lui, la reconnaissance de l'antécédent d'un pronom serait donc due à des phénomènes superficiels par rapport à la reconnaissance de la structure applicative. S'il est indéniable que d'une façon très générale, qui englobe le problème des relatives, des propriétés de dépendance morphologique comme les accords genre / nombre / personne et l'ordre des mots interviennent pour aider le locuteur à interpréter une phrase et notamment éviter des ambiguïtés (sinon pourquoi de nombreuses langues les auraient-elles mises en place ?), l'hypothèse de Harris est contredite par de nombreux faits. Les pronoms, relatifs ou non, peuvent référer de manière non ambiguë à leur antécédent du fait des contraintes qui définissent les structures applicatives : les restrictions de sélection lexicales et syntaxiques (types). Ceci est illustré par des exemples aussi banals que les suivants où les différentes occurrences de *il* et l'occurrence de *en* ne sont pas ambiguës :

Jean a pris l'autobus. (Il était bondé / il est mauvais marcheur / il en a résulté une certaine confusion.)

La localisation de l'antécédent d'un pronom grâce à l'ordre des mots n'intervient en fait que dans des situations très limitées de relativation et devient souvent inopérante quand l'antécédent fait partie d'un groupe nominal comportant d'autres noms. Par exemple, la désambiguïsation est lexicale dans les exemples (1), elle est syntaxique dans (2), (3) (4), (5) et (6). Dans (2) *qui*, et donc son antécédent, doit être de type élémentaire parce que *repartir* requiert un argument élémentaire. Dans (3), *qui* doit être de type non élémentaire donc son antécédent également, parce que *être acceptée* requiert un argument non élémentaire. Il n'y a qu'une expression assertable sans morphologie verbale associée dans ce contexte d'assertion qui soit candidat comme antécédent, donc (3) n'est pas ambiguë. En revanche, (4) est ambiguë mais (5) et (6) ne le sont plus. On en rend compte en imbriquant les reconstructions anaphoriques à l'intérieur de celle qui définit la « longueur » de l'antécédent de l'ellipse anaphorique (l'argument non élémentaire de *action* dont la « longueur » varie selon le contexte et qui pourrait dans d'autres contextes avoir une ou plusieurs sous-expressions assertables « approximées » selon la définition donnée au paragraphe 1) :

- (1) *Marc aime beaucoup le chapeau de Paul (qu'il a rencontré dans la rue / qu'il lui a offert la semaine dernière).*
- (2) *Jean a oublié la démonstration de Luc, qui vient de repartir.*
- (3) *Jean a oublié la démonstration de Luc, qui vient d'être acceptée.*
- (4) *Luc a révélé que Jean a escaladé la tour Montparnasse. Cette action nous a étonnés.*

(5) *Luc a révélé que Jean a escaladé la tour Montparnasse. L'action de Jean nous a étonnés.*

(6) *Luc a révélé que Jean a escaladé la tour Montparnasse. L'action de Luc nous a étonnés.*

Ce sont alors les restrictions de sélection syntaxiques et/ou lexicales qui permettent à l'auditeur de localiser l'antécédent, comme dans le cas général des relations anaphoriques, qu'elles mettent en jeu un pronom ou une ellipse anaphorique. Comme on l'a vu au paragraphe précédent, cette désambiguïsation lexico-syntaxique peut opérer « à distance » dans la structure applicative du contexte d'assertion d'une combinaison arbitrairement longue d'expressions assertables.

Toutes ces objections, ainsi d'ailleurs que le matériel contenu dans cet article, a été présenté à Harris peu de temps avant la fin de la rédaction de [GEMP] et a été ensuite l'objet de longues discussions au cours des années qui ont suivi. Je ne sais si on doit voir là le signe d'une modification qui intervient dans les derniers chapitres de [TLI] concernant le système d'interruption. Ce système fait intervenir une transformation d'une sorte nouvelle qui modifie le statut syntaxique des opérateurs en introduisant un argument phrastique dans une structure applicative qui était déjà complètement instanciée. Ce processus est décrit au paragraphe 8.3 à propos de la réduction de (1) en (2) en deux étapes :

(1) *a man asked for you ; the man spoke strangely*

(2) *a man, who spoke strangely, asked for you*

a) « ; » se réduit morphologiquement en *wh-* et la seconde occurrence de *man* se réduit en *-o* du fait de sa position sujet.

b) L'interruption transforme la « forme » de S_1 ; S_2 . S_1 reçoit un point d'insertion phrastique auquel est attachée la relative. La première occurrence de *a man* est redéfinie comme point d'insertion phrastique de S_1 .

D'une façon qui pourra paraître curieuse, cette nouvelle transformation ramène l'analyse des relatives à la description désormais classique qu'en avait proposée Kuroda en 1968 dans « English relativization and certain related problems », *Language* 44. Cette transformation était définie sur des structures de concaténation et avait pour résultat de transformer une phrase (S_2 dans l'exemple (1)) en modifieur de nom, grâce à la catégorie d'insertion auxiliaire DET. Cette transformation est parfaitement cohérente avec les structures syntagmatiques d'insertion lexicale et l'hypothèse d'autonomie de la syntaxe et de la sémantique de la grammaire de Chomsky. Cette transformation n'est malheureusement pas compatible avec la description des structures syntaxiques comme structures d'insertion lexicale et comme structures d'interprétation proposée dans [GEMP] et dans le reste de [TLI]. Pour ne donner qu'un exemple, une fois « ; » réduit à *wh-*, il ne serait plus un opérateur de co-assertion sur S_1 et S_2 mais un opérateur opérant sur la seule phrase S_2 . L'insertion elle-même de S_2 dans S_1 ne peut être définie dans la grammaire des opérateurs.

La relativation ne peut donc pas s'analyser à partir du mécanisme d'interruption, ce qui constitue un résultat négatif intéressant dans la mesure où

l'interruption faisait réintervenir les contraintes de concaténation à un niveau important de la grammaire, étant donné le rôle de la relativation dans les décompositions des contraintes linguistiques. Peut-être peut-on en conclure que la relation entre les contraintes sur l'ordre linéaire des mots, c'est-à-dire les contraintes de concaténation, et les contraintes applicatives que font intervenir les mécanismes d'interprétation est plutôt une relation de superposition entre deux systèmes hétérogènes qu'une relation d'interaction. Ceci n'est peut-être pas très étonnant si l'on songe à l'histoire des langues indo-européennes. Les grandes fonctions grammaticales comme les formations de relatives et de complétives, l'emploi de déterminants, l'emploi de prépositions, de conjonctions de subordination et de suffixations casuelles, de particules aspectuelles infixées et de suffixes de conjugaison sont des fixations morphologiques relativement tardives qui résultent de la spécialisation de mécanismes agglutinants mettant en jeu plusieurs thèmes pronominaux disposant chacun de formes atones et toniques, dans les langues indo-européennes archaïques. Ces mécanismes agglutinants ont été décrits par les grands spécialistes de ce domaine à partir d'appositions hiérarchisées. Il faut mentionner les travaux plus récents, très importants, de Françoise Bader, qui reprennent de façon détaillée la genèse des fonctions grammaticales à partir des thèmes pronominaux, pour l'ensemble des langues indo-européennes. Son analyse précise, complète et étend les hypothèses formulées par Whitney, Meillet, Vendryes, Wackernagel et Benveniste, pour ne citer qu'eux. On pourra trouver quelques-unes de ses vues et une bibliographie partielle de ses travaux dans : « Structure de l'énoncé indo-européen » et « La particule Hittite *san* ». Je pense pouvoir donner à ces mécanismes d'apposition l'interprétation applicative de l'insertion lexicale telle que je l'ai décrite mais je ne pourrai évidemment pas développer ce point important ici et donnerai simplement un exemple à propos de la complétivation, plus bas. Je n'ai pris connaissance de la genèse historique des fonctions grammaticales qu'après avoir défini les dérivations du français dont je tente de préciser le cadre ici. Rétrospectivement, l'analogie entre ces dérivations me paraît frappante.

D'après A. Meillet et J. Vendryes dans : *Traité de grammaire comparée des langues classiques*, les règles morphologiques d'accord et les contraintes sur l'ordre des mots seraient plus récentes que la genèse des grandes fonctions grammaticales précédemment mentionnées.

4. Insertion des déterminants et analyse des déterminations appositives et restrictives

Deux hypothèses liées seront importantes dans ce paragraphe :

— L'opérateur de co-assertion aura, non pas le caractère paratactique des grammairiens traditionnels ou de Harris, mais le caractère « corrélatif » des structures indo-européennes ayant donné naissance à la relativation, et à la complétivation (ainsi qu'aux autres subordinations conjonctives et prépositionnelles). En particulier, les pronoms de base de statut non élémentaire que j'utilise sous la co-assertion peuvent avoir des types différents.

— La plupart des éléments lexicaux non indéfinis, qu'ils soient de statut élémentaire ou non, seront insérés indirectement sous leurs opérateurs, par l'intermédiaire de pronoms de base ensuite identifiés puis réduits. Ceci a pour but de permettre d'analyser les grandes fonctions grammaticales du français contemporain, dont les fonctions anaphoriques, à partir des mécanismes ordinaires de l'insertion lexicale.

Comme exemples, je proposerai une analyse des déterminants à partir de relations entre phrases, une analyse de l'interprétation restrictive des relatives à partir d'applications hiérarchisées de relatives appositives et une décomposition de la complétivation à partir de la relativation. Cette analyse de l'insertion lexicale à différents niveaux d'une structure applicative par l'intermédiaire de relativations s'applique également aux adverbes et aux prépositions.

L'idée de ramener les relatives restrictives aux relatives appositives revient à Harris. Cependant, l'analyse qu'il propose ne peut être appliquée de façon générale. Je la présenterai brièvement parce qu'elle révèle ce qui me semble être une insuffisance dans l'analyse des déterminants. Or l'analyse des déterminants est ici une pierre de touche des autres décompositions.

Pour Harris, le déterminant indéfini n'a pas de statut applicatif, c'est une simple marque associée morphologiquement à un groupe nominal. Le déterminant défini est, lui, considéré comme une forme réduite de pronom. Dans [GEMP], la dérivation suivante, transposée au français, est proposée pour l'interprétation non générique du défini :

- celui-là est venu ; celui-là est un homme.*
- *celui-là, qui est un homme, est venu.*
- *celui-là, un homme, est venu.*
- (1) *L'homme est venu.*

Cette dérivation a le défaut de dériver (1) hors contexte alors que (1) n'est assertable et interprétable que dans un contexte qui explicite l'antécédent du défini. Il ne semble donc pas satisfaisant d'autoriser une dérivation qui permet d'insérer un déterminant défini indépendamment de la relation anaphorique qui lui permet d'exister.

Toujours dans [GEMP], les dérivations suivantes sont proposées (transposées au français) pour dériver la relativation restrictive de (2) et la relativation appositive de (3) :

- quelqu'un a téléphoné ; quelqu'un est un homme ; tu connais un homme.*
- *quelqu'un a téléphoné ; quelqu'un est un homme, que tu connais.*
- *quelqu'un, qui est un homme que tu connais, a téléphoné.*

(où la première relative perdrait sa virgule et deviendrait restrictive au moment où elle est associée à la deuxième relative appositive) :

- *quelqu'un, un homme que tu connais, a téléphoné.*
- (2) *un homme que tu connais a téléphoné.*
- un homme a téléphoné ; je connais un homme.*
- (3) *un homme, que je connais, a téléphoné.*

La dérivation de (2) n'est pas satisfaisante parce que cette association des relatives appositives n'impose pas que leur combinaison soit restrictive. En français comme en anglais, on peut itérer autant de relativations appositives qu'on le souhaite sans qu'aucune ne devienne restrictive à un moment donné, à partir de l'antécédent de cette dérivation. Par exemple : *Quelqu'un, qui est un homme, que tu connais, que tu as revu récemment, et qui est de passage à Paris, a téléphoné*. De plus, on peut appliquer la dérivation de Harris à un antécédent qui n'a jamais de relative restrictive, comme un nom propre : *Max, qui est un homme, que je connais, a téléphoné*.

La cause de la défaillance de ces dérivations me semble résider dans l'insuffisance de l'analyse des déterminants. En fait, dans les relatives restrictives, l'interprétation du déterminant doit être associée à l'ensemble de la séquence nom-relative et ceci vaut aussi avec un déterminant défini. Par exemple : *l'homme que tu connais a téléphoné* peut être paraphrasé par : *celui que tu connais, d'homme, a téléphoné*. Dans le cas des relatives appositives, ce sont les interprétations des deux phrases qui sont associées, chacun des deux substantifs restant déterminé par son propre déterminant.

Il existe des arguments diachroniques en faveur de l'analyse du déterminant *un* à partir d'un pronom indéfini *un* (par exemple en ancien français : *s'il dist un, il pense el* au sens de : *s'il dit quelque chose, il le pense*. Les emplois de *un* comme pronom sont encore fréquents au 18^e siècle et il en existe aussi pour l'anglais *a, an*. En français moderne, on emploie *quelqu'un* mais il reste quelques traces de l'emploi de *un* comme pronom comme dans *l'un et l'autre*). Plus précisément, il faut distinguer plusieurs interprétations du déterminant indéfini, qui viennent de l'extension de l'emploi du pronom dont il est issu. En anglais et en français, mais plus généralement dans les langues indo-européennes anciennes (sanskrit, grec, latin), le numéral a subi une extension de son emploi qui l'a amené à concurrencer la forme pronominale indéfinie dérivée d'un thème démonstratif atone. *Un* vient du numéral *unus* et *a, an* du numéral et pronom indéfini (toujours utilisé comme tel) *one*, d'où les interprétations spécifique ou non spécifique des pronoms et déterminants indéfinis, qui ont cumulé plusieurs emplois. Malheureusement, je ne peux pas m'arrêter ici sur l'histoire de *un, quelqu'un, et a, one*.

Les exemples qui fournissent des arguments en faveur de l'analyse des déterminants définis et indéfinis à partir de pronoms font apparaître une relation d'interprétation inclusive entre ce pronom et le substantif qu'il détermine. Des données existent en français en faveur de l'analyse de *le* à partir de pronoms *celui-là, celui-ci* (*cist, icil* en ancien français). En plus des arguments diachroniques, les phrases clivées du français actuel font apparaître un *de* dont il faut rendre compte : *Max en a mangé un — de gâteau*. Maurice Gross a utilisé cet argument dans : *Grammaire transformationnelle du français, syntaxe du nom* pour dériver *le N* de *celui de N* dans des dérivations hors contexte comme : *Max a mangé celui-là de gâteau* → *Max a mangé le gâteau*.

Je ferai l'hypothèse que les déterminants *un* et *le* sont des formes réduites de pronoms qui sont introduits dans une structure d'insertion lexicale comme

argument d'une préposition *de*, qui est elle-même une forme réduite d'une locution qui explicite l'interprétation inclusive mentionnée ci-dessus. L'interprétation anaphorique de *le* sera dérivée à partir d'un contexte qui la produit. La conséquence de cette hypothèse est de considérer une phrase apparemment simple telle (4) ci-dessous comme une forme réduite de co-assertion où *un homme* provient d'une expression assertable : *de sorte (quelqu'un, homme)* qui se réduit à : \emptyset (\emptyset , homme) dans ce contexte ⁴.

- quelqu'un a disparu ; quelqu'un est de sorte homme.*
- *quelqu'un a disparu ; quelqu'un est \emptyset homme.*
- *quelqu'un, qui est \emptyset homme a disparu.*
- *quelqu'un \emptyset , \emptyset homme a disparu.*
- (4) *un \emptyset \emptyset homme a disparu.*

La séquence assertable correspondant à (4) est :

\emptyset ((a disparu (un)), (\emptyset (\emptyset , homme))

On peut revenir maintenant à la dérivation des relatives restrictives et appositives, que j'illustrerai à partir des exemples suivants où les identifications que les relatives appositives et restrictives mettent en jeu ne sont pas les mêmes :

- (5) *un homme qui est venu hier a disparu.*
- (6) *un homme, qui est venu hier, a disparu.*

Dans (5), la relative est restrictive du fait qu'on applique deux identifications au pronom indéfini, qui devient anaphorique lors de la deuxième identification :

- quelqu'un a disparu ; quelqu'un est venu hier ; quelqu'un est \emptyset homme*
- *quelqu'un, qui est venu hier, a disparu ; quelqu'un est homme*
- *quelqu'un, qui est homme qui est venu hier, a disparu*
- *un homme qui est venu hier, a disparu*

Dans (6) la relative est appositive et l'interprétation de *un* est non anaphorique, les deux appositions de *un homme* se réduisent indépendamment l'une de l'autre :

- quelqu'un a disparu ; quelqu'un est \emptyset homme ; quelqu'un est venu hier ; quelqu'un est \emptyset homme*
- *quelqu'un, qui est \emptyset homme, a disparu ; quelqu'un, qui est \emptyset homme, est venu hier*
- *quelqu'un, homme, a disparu ; quelqu'un, homme, est venu hier.*
- *un homme, qui est venu hier, a disparu.*

4. Je rappelle que, pour les raisons indiquées au paragraphe 1, les dérivations proposées dans cet article ont valeur d'illustration de la discussion. Particulièrement dans ce paragraphe, elles devraient être précisées de plusieurs points de vue pour rendre compte notamment des interprétations génériques de *un* et *le* et des interprétations spécifique ou non spécifique de *un* mais surtout pour rendre compte des relations entre aspect et détermination. Sans entrer dans cette question, le choix des sortes d'expressions assertables que je propose ici doit permettre d'analyser et de hiérarchiser les paramètres qui interviennent dans ces relations. Pour ne donner qu'un exemple, l'aspect perfectif induit par le verbe et/ou un adverbe peut sélectionner l'interprétation spécifique d'un pronom indéfini ou d'un déterminant indéfini (*une pomme tombe toute les cinq minutes vs. une pomme est tombée à 9 h 47*) ; à l'inverse, un contexte qui impose l'interprétation indéfinie non spécifique à des pronoms qui instancient, sans identifications ultérieures, un verbe, peuvent imposer un aspect imperfectif à l'assertion correspondante.

Dans (7), la relative est restrictive du fait de l'interprétation anaphorique de *le*. La dérivation s'applique à des formes morphologiques de base différentes de celles de (5) mais la structure de dérivation serait la même :

(7) *L'homme qui est venu hier, a disparu.*

Dans les exemples donnés ci-dessus, les antécédents de relatives sont de statut élémentaire, les formes morphologiques de base qui se réduisent en articles aussi. A la différence des autres éléments lexicaux, les pronoms et les éléments lexicalement indéfinis n'ont pas de type initial, ils peuvent être de statut élémentaire ou non selon le contexte. Par exemple, les nominalisations peuvent avoir des déterminants définis, provenant de la formation de la nominalisation ou d'une relation anaphorique faisant intervenir un contexte de niveau supérieur, ou des déterminants indéfinis. Elles peuvent aussi être les antécédents de relatives avec des contraintes supplémentaires sur l'insertion de leur déterminant.

En plus des formes pronominales de base qui se réduisent en articles *un, des* et *le, les*, j'ai utilisé dans les paragraphes précédents les éléments lexicalement indéfinis *quelque chose* ou *chose*, comme éléments anaphoriques ou non anaphoriques. On peut également supposer qu'ils interviennent dans la dérivation des complétives, la particule complétive étant considérée alors comme une sorte de variante de ce pronom de statut non élémentaire associé au pronom relatif. Peut-être peut-on considérer *chose* comme une forme de base apparentée au pronom démonstratif, sinon il faudrait choisir une autre forme de base ; le point que je voudrais souligner est relatif à la structure applicative de la complétivation. On aura par exemple la dérivation schématique suivante (toujours sans tenir compte de la distinction entre assertion et expression assertable), avec la structure applicative correspondante :

(8) *Jean sait quelque chose ; Jean est parti est quelque chose*
Jean sait quelque chose qui est Luc est parti
Jean sait que Luc est parti

Dans (8), le statut syntaxique de la première occurrence de *quelque chose* est « approximé » selon la définition introduite au paragraphe 1, mais son type lui permet d'être identifié à la seconde occurrence, complètement instanciée de *quelque chose* (voir annexe).

La façon dont les articles sont insérés par rapport aux noms qu'ils déterminent, dans l'interprétation applicative qui est proposée ici, présente des analogies structurelles avec d'autres formes de « détermination » comme la complétivation. Des analogies apparaissent aussi pour l'insertion des adverbes et des prépositions.

RÉFÉRENCES

BADER F., 1987. « Structure de l'énoncé indo-européen », *Papers from the 7th international conference on historical linguistics*, John Benjamins.

- BADER F., 1988. « La particule Hittite *san* », *A linguistic Happening in memory of Ben Schwartz*, Peeters, Louvain-la-neuve.
- BENVENISTE E., 1935. *Origines de la formation des noms indo-européens*, Paris, Adrien Maisonneuve.
- CHURCH A., 1941. *The Calculi of Lambda Conversion*, Princeton U. Press.
- DALADIER A., 1984. « Traitement de coréférences dans une représentation applicative des textes », *Actes du 4^e congrès Reconnaissance des Formes et Intelligence artificielle*.
- DALADIER A., 1985. « Using types to define assertability in natural languages », Stanford Conférence, résumé dans *Journal of Symbolic Logic*, vol. 51.
- DALADIER A., 1986. « Deriving the Literal Interpretation Structures of a Natural Language in a Second Order λ -Calculus », mimeo U. Paris 7.
- GIRARD J. Y., 1987. « The System F of variable Types Fifteen Years Later », *Theoretical Computer Science* 50.
- GROSS M., 1986. *Grammaire transformationnelle du français, syntaxe du nom*, Cantilène.
- HINDLEY R., SELDIN J., 1986. *Introduction to Combinators and Lambda-Calculus*, Cambridge University Press.
- HINDLEY R., SELDIN J., eds. 1981. *To H.B. Curry, Essays on Combinatory Logic, Lambda-Calculus and Formalism*, Academic Press.
- KLEENE S., 1952. *Introduction to Metamathematics*, Van Nostrand.
- KRIVINE J. L., 1985. *Programmation en Arithmétique Fonctionnelle du Second Ordre*, mimeo. Equipe de logique mathématique, Université Paris 7.
- KURODA S. Y., 1968. « English relativization and certain related problems », *Language* 44.
- LEEMAN D., 1990. « Les compléments adverbiaux de phrase de type *en toute N : en toute objectivité, Paul est un crétin* », *Mélanges Lerond, Linx n° 23*, Paris X Nanterre.
- MEILLET A. et VENDRYES J., 1924. *Traité de grammaire comparée des langues classiques*, Champion (première édition).
- SCOTT D., 1976. « Data Types as Lattices », *STAM Journal of Comput.*, vol. 5, n° 3.
- SCOTT D., 1980. « Lambda Calculus : Some Models, Some Philosophy », *The Kleene Symposium*, Reidel.
- TARSKI A., 1955. « A Lattice-Theoretical fixpoint Theorem and its Applications », *Pacific J. of Math.*, 55.
- WHITNEY W. D., 1875. *La Vie du langage*, Paris, Félix Alcan.

ANNEXE

Cette annexe présente un caractère technique que je ne pourrai expliciter ici et s'adresse à des lecteurs déjà familiarisés avec le λ -calcul et les systèmes de types.

1. Construction du domaine des types d'expressions assertables, TAS

TAS est un domaine de types sans variables, qui sont obtenus quand les structures d'interprétation des énoncés sont évaluées.

Avant leur insertion dans un énoncé particulier, les expressions lexicales à arguments phrastiques ont un type initial du second ordre (i.e. des variables de type sur lesquelles opèrent l'abstraction et les réductions). Ces variables indiquent le nombre d'arguments et selon leur rang, leur statut élémentaire (o-aire) ou à valeur dans TAS. L'opérateur lexical « ; » s'évalue en faisant intervenir une fonction d'identification qui sera décrite à la fin de cette annexe. Les autres expressions lexicales sont soit élémentaires, par exemple *Jean* ou *pomme*, soit des opérateurs à arguments élémentaires ou non élémentaires, soit des pronoms ou des éléments lexicalement indéfinis qui n'ont pas de type initial mais seulement un type du premier ordre dans un énoncé particulier. Un opérateur lexical a au plus trois arguments, quel que soit leur statut applicatif.

Le système de type a trois types de base (i.e. sans variables) : n , o , μ

— n est le type des expressions élémentaires

— o est un type tel que : $\Lambda\alpha. \{o\} = o$ pour toute expression de type

— μ est un type tel que : $\Lambda\alpha. \{\mu\} = \mu$ pour toute expression qui ne type pas un élément lexicalement indéfini ou un couple d'éléments lexicalement indéfinis.

TAS_i est le sous-domaine des types de séquences assertables qui ont i opérateurs de statut non élémentaires, c'est-à-dire i sous-expressions assertables. Ainsi, *être la conséquence de* requiert deux arguments de statut non élémentaires, par exemple *le départ de Jean* et *l'arrivée de Marc*. En fait, les opérateurs ne sont pas insérés directement les uns sur les autres mais insérés par l'intermédiaire d'éléments lexicalement indéfinis. En se reportant au paragraphe 3 et après avoir vu les règles de typage, le lecteur pourra vérifier que :

un homme est venu a le type o dans TAS₃

Jean sait qu'un homme est venu a le type o dans TAS₆

quelque chose est arrivé a le type $\mu \rightarrow o$ dans TAS₁

TAS est construit à partir d'une induction double :

— sur la longueur des expressions assertables (leur nombre de sous-expressions assertables)

— sur le nombre d'expressions « approximée » en position terminale.

Les types de chaque sous-domaine de TAS (chaque TAS_i) sont ordonnés selon la relation induite par l'« approximation ». D'une façon générale, l'ordre sur les types de TAS correspond intuitivement à une notion linguistique particulière de capacité informative sur laquelle je reviens à la fin de cette section.

TAS est un treillis complet tel que :

$$\text{TAS} = \langle \overline{\text{TAS}}, \leq \rangle$$

$$\overline{\text{TAS}} = \bigcup_{i \in \mathbb{N}} \overline{\text{TAS}}_i$$

$$\overline{\text{TAS}}_i = \{\mu, o\} \quad \mu < o$$

Induction des éléments de $\overline{\text{TAS}}_i$ à partir de $\overline{\text{TAS}}_1$: deux situations empiriques doivent être distinguées pour obtenir les types de TAS :

1) l'insertion lexicale d'une expression assertable ayant un type dans TAS_{i-1} sous un opérateur lexical qui requiert 1 argument de statut non élémentaire.

2) L'insertion lexicale d'un couple d'expressions assertables ayant un type respectivement dans TAS_j et TAS_k , où $j + k = i - 1$, sous un opérateur lexical qui requiert un couple d'arguments non élémentaires.

Dans ces deux situations, les éléments de TAS_i sont obtenus en utilisant le type initial de l'opérateur à insérer sur une expression assertable et les règles de typage décrite plus bas.

— Le g.l.b. et le l.u.b. de TAS_{i-1} , μ et o , sont préservés dans TAS_i .

— si $\mu < a < o$, $a \in TAS_{i-1}$ alors $(a \rightarrow o) \in TAS_i$

— si $\mu < a < o$ et $\mu < c < o$

ou $b \in TAS_{i-2}$, $c \in TAS_2$ et $(b \rightarrow (c \rightarrow o)) \in TAS_i$

ou $b \in TAS_{i-3}$, $c \in TAS_3$ et $(b \rightarrow (c \rightarrow o)) \in TAS_i$

etc.

ou $b \in TAS_2$, $c \in TAS_{i-2}$ et $(b \rightarrow (c \rightarrow o)) \in TAS_i$

Ordre des éléments de \overline{TAS}_i :

(i) $a \in TAS_i$, $b \in TAS_i$, $\mu < b < o$

alors $a \leq b$ si b peut être obtenu en substituant des occurrences de o à $(\mu \rightarrow o)$ dans a

Construction de TAS :

Les éléments de TAS sont ceux de l'union des TAS_i . Cette union n'est pas finie parce que la longueur des séquences assertable n'est pas limitée *a priori*.

Ordre des éléments de TAS :

Les éléments de TAS sont hiérarchisés de deux façons, premièrement en conservant l'ordre de chaque TAS_i , c'est-à-dire selon (i) et deuxièmement selon (ii) :

(ii) $a \in TAS$, $b \in TAS$, $a \leq b$ si $b = a \rightarrow o$

La construction de TAS met en jeu les deux propriétés suivantes :

• pour tout terme U produit d'opérateurs lexicaux ayant un type dans TAS, il existe un opérateur lexical X requérant un argument non élémentaire, tel que X appliqué à U est un terme dont le type est dans TAS

$\forall U, U : a, a \in TAS_k, \exists X, X : \Lambda \alpha. (\alpha) \rightarrow (\alpha \rightarrow o),$

$XU : b, b \in TAS_{k+1}$

• Pour tout couple de termes (U, V) , chaque terme ayant un type dans TAS, il existe un terme Z requérant deux arguments non élémentaires tels que Z appliqué à (U, V) est un autre terme dont le type est dans TAS. Après avoir lu la fin de ce paragraphe et les règles de typage, en particulier la règle (5) et les équations (6), le lecteur pourra vérifier que :

$\forall (U, V), U : a, V : b, a \in TAS_i, b \in TAS_j,$

$\exists Z, Z : \Lambda \alpha \beta. (\alpha \rightarrow (\beta \rightarrow (\alpha \rightarrow (\beta \rightarrow o))))), Z(U, V) : c$

si $\mu < a \leq 0$ et $\mu < b \leq 0$, alors $c \in \text{TAS}_{i+j+i}$
 si $\mu < a < 0$ et $b = 0$, alors $c \in \text{TAS}_{i+1}$
 si $a = 0$ et $\mu < b \leq 0$, alors $c \in \text{TAS}_{j+1}$

L'interprétation linguistique des valeurs de TAS est la suivante : les types de TAS ne définissent pas des valeurs d'assertabilité, puisque toutes les expressions qui ont un type dans TAS sont également susceptibles de devenir des assertions moyennant l'application des règles de linéarisation, les ajustements morphologiques et les conditions lexicales sur les réductions « appropriées » mentionnées au cours de cet article. L'ordre des types de TAS correspond à un ordre sur une notion de valeur informative que les séquences assertables peuvent prendre. Une expression assertable contenant un nombre donné d'opérateurs peut être plus ou moins informative selon qu'elle est plus ou moins finement instanciée. Plus un type contient d'occurrences de μ plus l'expression assertable correspondante est non complètement instanciée.

2. Règles de types

L'ensemble des types est le plus petit ensemble T tel que :

- T contient trois types de base :
 - n : pour les opérateurs lexicaux élémentaires
 - o : point fixe maximal d'un opérateur lexical requérant un argument élémentaire
 - μ : point fixe minimal d'un opérateur lexical requérant un argument non élémentaire
- T contient une liste infinie de variables de types :
 - $\alpha_1, \dots, \alpha_k, \dots, \beta_1, \dots \in V$
- si $\alpha, \beta \in T$, βn , alors $\alpha \rightarrow \beta \in T$
- si $\alpha \in V$, $\beta \in T$, βn , alors $\Lambda \alpha. \beta \in T$

Un terme typé est défini inductivement de la manière suivante :

(1) une variable de type ou une constante de type est un terme

$$(2) \frac{x : \alpha \quad t}{\tau} \quad \tau \neq n$$

$$(3) \frac{t : \gamma \rightarrow \tau \quad u : \gamma}{t(u) : \tau} \quad \gamma = n \text{ ou } \gamma \in \text{TAS}$$

$$(4) \frac{t : \gamma}{\Lambda \alpha. t : \Lambda \alpha. \gamma} \quad \gamma \neq n \text{ et } \alpha \text{ n'est pas libre dans } \gamma$$

$$(5) \frac{t : \Lambda \alpha. \gamma}{\{ \tau \} : \gamma[\tau/\alpha]} \quad \tau \in \text{TAS} - \{ \mu \} \text{ ou } \tau = \mu \text{ et } = \lambda x. \text{quelque chose } (x) \text{ sinon } t\{\mu\} = \mu$$

(6) On a les équations suivantes :

$$\forall \alpha \in T \quad \alpha \rightarrow \mu = \mu$$

$$o \rightarrow o = o$$

$$\tau \rightarrow (\tau \rightarrow o) = \tau \rightarrow o \text{ si } \mu < \tau < o$$

$$\gamma \rightarrow (\tau \rightarrow (\gamma \rightarrow (\tau \rightarrow o))) = (\tau \rightarrow (\gamma \rightarrow o)) \text{ si } \mu < \tau < o \text{ et } \mu < \gamma < o$$

$$\gamma \rightarrow (\tau \rightarrow (\gamma \rightarrow (\tau \rightarrow o))) = (\tau \rightarrow o) \text{ si } \gamma = o \text{ et } \mu < \tau < o$$

$$\gamma \rightarrow (\tau \rightarrow (\gamma \rightarrow (\tau \rightarrow o))) = (\gamma \rightarrow o) \text{ si } \tau = o \text{ et } \mu < \gamma < o$$

Les équations et les règles de réduction de type décrivent la normalisation des termes typés pour les différents cas de construction des expressions assertables, après qu'on ait appliqué les reconstructions linguistiques. Elles décrivent donc les différents cas d'insertion lexicale. Les contraintes sur la normalisation des types, c'est-à-dire les contraintes indiquées à droite de la règle (5) et les équations (6) garantissent l'accroissement monotone de l'information (au sens décrit précédemment) sous l'insertion lexicale. Les équations (6) permettent d'éliminer les types des sous-expressions assertables incomplètes lorsque les conditions d'une reconstruction linguistique peuvent s'appliquer. On a vu que ces conditions sont l'insertion d'une expression contenant une ou plusieurs sous-expressions « approximées » dans un contexte d'assertion qui autorise des identifications permettant de réinstancier une ou plusieurs des sous-expressions incomplètes. Linguistiquement parlant, il s'agit des ellipses anaphoriques (voir les règles de réduction linguistiques plus bas). Les quatre dernières équations de (6) décrivent les différentes situations où le type d'une expression incomplète ne doit pas être éliminé, c'est-à-dire les différentes situations où une expression assertable demeure partiellement incomplète, en l'absence d'un contexte autorisant une reconstruction anaphorique, tout en préservant l'accroissement monotone de l'information sous d'éventuelles insertions de niveau supérieur.

3. Inférence de reconstructions linguistiques

X est une séquence de mots qui a un type dans TAS s'il est possible de reconstruire une source Z d'une dérivation linguistique telle que Z est un produit d'expressions assertables, c'est-à-dire : $Z \stackrel{*}{\bar{L}} X$ (où $Z \stackrel{*}{\bar{L}} X$ est la fermeture réflexive et transitive de \bar{L}).

La règle de dérivation linguistique $\vec{\bar{L}}$ comporte deux sous-règles de dérivation, notées $lid^{\vec{\bar{L}}}$ et $loc^{\vec{\bar{L}}}$, qui sont hiérarchisées. Elles correspondent aux réductions linguistiques sous les mécanismes d'identification de l'opérateur de co-assertion « ; » et les réductions « appropriées », qui sont des réductions locales.

On se demande si une réduction linguistique a eu lieu quand on ne peut pas dériver un type de TAS à partir des opérateurs lexicaux explicites d'un énoncé à analyser. Quand on infère l'hypothèse d'une reconstruction linguistique, on commence par tenter une reconstruction sous $Lid^{\vec{\bar{L}}}$.

Une des hypothèses fondamentale de cette approche est que les dérivations linguistiques ne modifient pas les structures applicatives qui représentent

simultanément les structures d'insertion lexicale et les structures d'interprétation des énoncés. Autrement dit, $\overset{*}{\Gamma}$ préserve les types des séquences assertables et ne produit que des liaisons de variables qui ont pour effet de permettre des substitutions d'instanciations, qui ont soit une forme normale, soit une forme normale « approximée ».

On peut définir les réductions linguistiques « appropriées », qui s'appliquent à des sous-expressions qui demeurent incomplètement instanciées grâce à la notion d'approximation définie dans loc^{\rightarrow} et grâce à la règle de type (5) associées aux équations déjà décrites.

Id^{\rightarrow} est une règle d'identification (ou plutôt, comme on le verra plus bas, une famille de formes conditionnelles d'identification) dont la principale caractéristique est d'avoir TAS comme structure de contrôle. Elle permet de rendre compte de formes d'ellipses anaphoriques, qui n'avaient pas été décrites jusqu'à présent.

Finalement, $X : \Lambda\alpha\beta.(\alpha \rightarrow (\beta \rightarrow (\alpha \rightarrow (\beta \rightarrow o))))$ (où X est un opérateur lexical) a l'interprétation suivante :

X est un terme typé qui peut devenir une expression assertable s'il est appliqué dans un contexte où il peut être instancié par des termes qui ont des types dans TAS.

3.1. La réduction loc^{\rightarrow}

Je donnerai un exemple avant de préciser la règle.

Les opérateurs *espérer* ou *espérer* ont un argument non élémentaire, qui est explicite dans (d) mais pas dans (e) ni dans (f) :

(d) *Jean espère que Marie viendra.*

(e) *Jean a été informé que Marie viendra et il a bon espoir pour Pierre.*

(f) *Les gens espèrent.*

(f) est représentée par (g), où Y représente le plus petit point fixe fonctionnel d'un opérateur (en souvenir de l'opérateur paradoxal de Curry). (g) est dérivée de (h) par Loc^{\rightarrow} . (h) est dérivée de (k) par μ -réduction, telle qu'on la définit après cet exemple :

(k) $\lambda xy. \text{espèrent}(x, y) : n \rightarrow \Lambda\alpha. \alpha \rightarrow (\alpha \rightarrow o)$ (gens : n)
 $(\lambda z. \text{choses}(z) : \Lambda\beta. \beta \rightarrow (\beta \rightarrow o))$

$\overset{\mu}{\rightarrow}$ (h) $\lambda xy. \text{espèrent}(x, y) : n \rightarrow \Lambda\alpha. \alpha \rightarrow (\alpha \rightarrow o)$ (gens : n)
 $(\lambda z. \text{choses}(z) : \Lambda\beta. \beta \rightarrow (\beta \rightarrow o)) (Y : \mu)$

\rightarrow $\lambda xy. \text{espèrent}(gens, y) : \Lambda\alpha. \alpha \rightarrow (\alpha \rightarrow o)$ (choses(Y) : $\mu \rightarrow o$)

\rightarrow $\text{espèrent}(gens, Y\text{choses}) : \mu \rightarrow o \rightarrow o$

Loc^{\rightarrow} (g) $\text{espèrent}(gens, YO) : \mu \rightarrow o \rightarrow o$

Les opérateurs lexicaux à argument non élémentaire peuvent s'appliquer à leur plus petit point fixe fonctionnel mais de façon non significative s'ils ne sont pas des éléments lexicalement indéfinis comme *chose*, *quelque chose*, ils ont alors le type μ ($Y : \mu$). Par contre, (i) $Y\text{chose} : \mu \rightarrow o$ est introduit directement comme règle (règle (5) dans 2), et je considère ce terme comme l'approximation d'un produit infini d'opérateurs *chose* demeurant typés au second ordre. Une

expression qui a une ou plusieurs sous-expressions (i) a une forme normale plus ou moins approximée selon le nombre de sous-expressions (i). Les termes qui ont une forme normale approximée et qui sont interprétables comme expressions assertables sont définis inductivement selon la construction de TAS.

La μ -réduction est définie en relation avec l'approximation :

- Elimination : règle (5) et les contraintes équationnelles données dans la section 2.
- Introduction : s'il n'existe pas de terme typé $A : \tau$, $\tau \in \text{TAS}\{-\mu\}$, sur lequel peut s'appliquer λx . quelque chose $(x) : \Lambda \alpha. \alpha \rightarrow (\alpha \rightarrow o)$ dans son contexte d'assertion, alors Y chose $:\mu \rightarrow o$ est introduit comme forme normale approximée de cette sous-expression.

3.2. La règle de réduction Lid^{\rightarrow}

Lid^{\rightarrow} s'applique sous l'opérateur de co-assertion « ; » et ses conditions d'application sont définies par une propriété qui joue un rôle fondamental ici. Quand deux expressions sont co-assertées, i.e. assertées dans le même contexte d'assertion, leurs interprétations sont liées. Les arguments qui ont des formes normales ou approximées sont identifiées. Mais on identifie également des sous-expressions si elles ont des types ordonnés dans TAS et que les éléments lexicaux de l'une sont compris dans les éléments lexicaux de celle qui a le plus grand type. Celle qui est partiellement moins instanciée que l'autre est réévaluée comme étant égale à la mieux instanciée.

Je donnerai une description intuitive de l'évaluation de « ; », puis un exemple et à la fin du paragraphe une partie de sa description formelle.

Soient A et B deux expressions assertables, qui ont comme telles des types dans $\text{TAS}\{-\mu\}$. Id représente une expression qui détermine s'il y a deux sous-expressions X et X' pour lesquelles l'une des conditions (1), (2), (3), (4) décrites plus bas sont réunies. Si c'est le cas, Id abstrait une des deux variables de type correspondant à celle des deux instanciations (X ou X') à laquelle sera substituée une forme réduite (pronom ou forme zéro) dans l'expression de tête de l'opérateur « ; ». Comme les expressions identifiables peuvent contenir des sous-expressions, Id cherche les plus longues expressions identifiables dans le contexte. Id fournit aussi les localisations syntaxiques respectives de X et X' si elles peuvent être identifiées. Ceci se fait à partir de projections successives dans des expressions conditionnelles emboîtées. Comme on le sait, les expressions conditionnelles emboîtées sont exprimées comme des applications ordinaires sous les interprétations usuelles du vrai et du faux comme première et seconde projection respectivement.

Le processus d'identification fait partie de l'interprétation lexicale de « ; » tout comme des verbes tels *demander* et *promettre* ont des traits de « projection lexicales » qui autorisent des identifications anaphoriques particulières dans leur descendance.

La lambda expression qui représente Id s'élimine au cours de son évaluation et les structures d'interprétation des expressions assertables se ramènent à des applications ordonnées d'opérateurs lexicaux.

Les principaux aspects de l'interprétation de « ; » peuvent être résumés de la façon suivante :

$\lambda xys. ; (x,y,Id(s)) (A :a) (B :b)$
 s est un n -uple de variables, $a, b \in \text{TAS}-\{\mu\}$,

l'égalité entre les termes est la relation d'équivalence induite par la β -réduction,

$A = [...X...], B = [...X'...]$,

L'une des cinq situations suivantes se produit :

- (1) $a = b = n$ et $X = X'$
- (2) $a = b$, $a, b \in \text{TAS}-\{\mu\}$ et $X = X'$
- (3) $a < b$, $a, b \in \text{TAS}-\{\mu\}$ et X peut être obtenu de X' en substituant un opérateur lexicalement indéfini ou un couple d'opérateurs lexicalement indéfinis opérant sur leur plus petit point fixe en position terminale, ou à un couple de produits d'opérateurs lexicaux
- (4) $a > b$, $a, b \in \text{TAS}-\{\mu\}$ et X' peut être obtenu de X comme X peut l'être de X' dans (3).
- (5) a et b ne sont pas ordonnés l'un par rapport à l'autre (le type n ne peut être comparé à un type de TAS ou X et X' ne comportent pas les mêmes opérateurs lexicaux.

Exemple

J'ai proposé la décomposition de la complétivation à partir de la relativation. J'indique ci-dessous la dérivation simplifiée correspondante qui suffira à illustrer mon propos. (c) est dérivé de (a) :

- (a) *Jean espère quelque chose ; Pierre est venu est quelque chose.*
- (b) *Jean espère quelque chose qui est Pierre est venu*
- (c) *Jean espère que Pierre est venu.*

(a) $Lid \vec{}$ (b) $Loc \vec{}$ (c), où la particule complétive est une forme morphologique réduite de *quelque chose qui est*.

Analyse de la séquence B de (a), *Pierre est venu est quelque chose*

On trouve dans le dictionnaire le type initial des entités lexicales :

Pierre : n , $\lambda x. \text{est venu } (x) : n \rightarrow o$, $\lambda y. \text{quelque chose } (y) : \Lambda \alpha. \alpha \rightarrow (\alpha \rightarrow o)$

Comme α est à valeur dans TAS, la seule façon d'ordonner l'application des entités lexicales de B est la suivante :

$\lambda x \text{ quelque chose } (y) : \Lambda \alpha. \alpha \rightarrow (\alpha \rightarrow o) ((\lambda x. \text{est venu } (x) : n \rightarrow o) (\text{Pierre} : n))$
 $\vec{\beta} \text{ quelque chose } (\text{est venu } (\text{Pierre})) : o.$

Analyse de la séquence A : *Jean espère quelque chose.*

On obtient dans le dictionnaire les types initiaux :

Jean : n , $\lambda xy. \text{espère } (x,y) : n \rightarrow \Lambda \alpha. \alpha \rightarrow (\alpha \rightarrow o)$,

$\lambda z. \text{quelque chose } (z) : \Lambda \beta. \beta \rightarrow (\beta \rightarrow o)$

Ici, les traits lexicaux mentionnés dans l'article à propos des règles de linéarisation doivent être utilisés pour reconstruire l'ordre applicatif des éléments lexicaux :

$$\lambda xy . \text{ espère } (x,y) : n \rightarrow \Lambda \alpha . \alpha \rightarrow (\alpha \rightarrow o) \quad (\text{Jean} : n) \quad (\lambda z . \text{ quelque chose } (z) : \Lambda \beta . \beta \rightarrow (\beta \rightarrow o))$$

Si on essaie d'évaluer la structure de A avant d'évaluer la structure de son contexte d'assertion, on ne peut obtenir qu'une forme approximée c'est-à-dire :

$$(\lambda z . \text{ quelque chose } (z) : \Lambda \beta . \beta \rightarrow (\beta \rightarrow o)) (Y : \mu) : \mu \vec{\mu} \text{ Y quelque chose} : \mu \rightarrow o$$

alors A devient :

$$\text{ espère } (\text{Jean}) (\text{ Y quelque chose }) : \mu \rightarrow o \rightarrow o,$$

ce qui est incorrect. Il faut évaluer le contexte ; (A,B) où Y quelque chose : $\mu \rightarrow o$ est identifiable à :

$$(\text{ quelque chose } (\text{ est venu } (\text{ Pierre}))) : o, \text{ où } \mu \rightarrow o < o.$$

$$\lambda \text{trs.}; (t,r, \text{Id}(s)) (A : \mu \rightarrow o \rightarrow o) (B : o)$$

$$= \Lambda \alpha \lambda x.; (A(x : \alpha) B(x : \alpha)) (\text{ quelque chose } (\text{ est venu } (\text{ Pierre}))) : o : o$$

On peut revenir maintenant au cas général :

$$\lambda xys. ; (x,y, \text{Id}(s)) \text{US}$$

s'évalue selon que U et V ont deux sous-expressions identifiables et selon leur type. On doit considérer trois cas :

(i) $\Lambda \alpha \lambda x. ; (U(x)V(x)) (A : c)$, où $c \in \text{TAS-}\{\mu\}$

(ii) $\lambda x : n. (U(x)V(x)) (A : n)$

(iii) $\lambda xy. ; (x,y) (U) (V)$

Le cas (i) correspond aux cas (2), (3), (4) d'identification, le cas (ii) correspond au cas (1) d'identification et le cas (iii) correspond au cas (4), où aucun argument identifiable n'occure dans les arguments de « ; ». De plus, dans le cas (i), si $c < o$, on doit essayer de réévaluer U et V sous un opérateur de co-assertion de niveau supérieur si ; (U,V) n'est pas asserté dans le contexte vide. Finalement, le type de :

$$\lambda xys. ; (xy(\text{Id}(s))) (U : a) (V : b)$$

s'évalue selon deux cas, de la manière suivante :

(j) $(a \rightarrow (b \rightarrow (a \rightarrow (b \rightarrow o))))$

si aucune identification des instanciations n'est possible, ou si l'identification met en jeu deux arguments ayant le même type (ce type peut être n ou un élément de $\text{TAS-}\{\mu\}$).

(jj) $(\Lambda \gamma . (a[\gamma/c] \rightarrow (b[\gamma/d] \rightarrow (a[\gamma/c] \rightarrow (b[\gamma/d] \rightarrow o)))))) \{e\}$

où $c, d, e \in \text{TAS-}\{\mu\}$, et soit $c < d$ ou $d < c$, $c = \sup(c, d)$

si l'identification met en jeu deux sous-expressions de types non égaux.