

Anne Daladier
LACITO, CNRS
a.daladier@wanadoo.fr

L'inconsistance de la théorie des opérateurs de Harris : de la fin des grammaires finitistes à une typologie inductive

Résumé

Zellig Harris a mis en évidence le rôle essentiel de l'ellipse grammaticale sur des éléments lexicaux indéfinis ou sur des éléments lexicaux redondants dans une construction donnée et l'inutilité des catégories de parties du discours. Ces hypothèses novatrices ont cependant été desservies par une croyance en une structure mathématique du langage limitée à des dépendances strictement linéaires et finies sur des éléments lexicaux ; sa théorie des opérateurs s'est précocement révélée inconsistante (Daladier 1990). Par ailleurs sa théorie a méconnu la sémantique grammaticale. Une grammaire basée sur des dépendances d'éléments grammaticaux et lexicaux multilinéaire et non finie permet de construire une typologie inductive sans catégories prédéfinies. Elle permet par exemple de décrire des langues austroasiatiques conservatrices, principalement isolantes, à grammaticalisations non conventionnelles, dont la complexité équivaut à celles des langues indo-européennes modernes. Propriétés de portée variable et ellipses dépendant des constructions fondent la complexité et confirment l'hypothèse de Drift de Sapir (1921) : l'évolution avec ses morphologies plus ou moins compliquées et règles d'accord ne crée pas de complexité.

Mots clefs: contrainte distributionnelle, grammaticalisation, opérateur lexical, opérateur grammatical, argument, argument non fini, ellipse grammaticale, portée variable d'éléments lexicaux selon la construction, restrictions de sélection non linéaires, non iconicité, mathématiques ou philosophie intuitioniste, typologie inductive, austroasiatique, évolution interne, complexité, morphologies: isolante, dérivationnelle, flexionnelle et mixtes.

Abstract

Zellig Harris has shown that Grammatical Categories are not necessary to define interpretation structures but contrary to what he thought these structures can neither be finite nor restricted to simple linear dependencies. Harris intuitions on grammatical ellipsis being a core property and the superficial character of part of speech categories still are novel and promising but his belief in simple lexical dependency relationships to ground interpretation structures are inconsistent, (Daladier 1990). Moreover, his theory did not account for grammatical semantics. A grammar based on non finite multilinear dependency relationships between both grammatical and lexical elements makes it possible to account without predefined categories for unusual grammaticalisations of austroasiatic conservative languages having mostly isolating morphologies and which are not less complex than modern Indo-European languages. Variable scope properties of many lexical operators, grammatical ellipsis and multi-linear dependencies between both grammatical and lexical elements ground complexity in languages of different families at different stages of evolution and confirm the Drift hypothesis of Sapir (1921): evolution with its more or less intricate morphologies and agreement rules does not create complexity.

Key words: distributional constraint, grammaticalization, lexical operator, grammatical operator, argument, non finite argument, grammatical ellipsis, variable scope properties of lexical elements according to constructions, non linear selection restrictions, non iconicity, intuitionist mathematics or philosophy, inductive typology, austroasiatic, internal evolution, complexity, morphologies: isolating, derivational, flexional and mixed.

1. Zellig Harris, chercheur non conventionnel, quelques souvenirs, collaborations

Zellig Harris est venu à Paris en 1978, à l'invitation de Maurice Gross, faire les conférences qui ont donné lieu aux *Notes du cours de syntaxe*. C'est là que je l'ai rencontré, quand je finissais mon doctorat. J'ai été immédiatement fascinée par son audace intellectuelle alors qu'il semblait si réservé, si retranché, hors de portée de tout sentiment tout en montrant beaucoup d'humour. A la fin de son cycle de conférences, je lui ai demandé s'il accepterait de répondre à mes questions, ce qui prendrait sans doute quelques jours, si je venais le voir à New York l'année suivante. Il m'a répondu avec un grand sourire que ça dépendrait entièrement des questions et qu'en tout cas je ne les lui envoie pas avant parce qu'il ne lisait jamais ce genre de choses. Quelques mois plus tard, je suis partie faire un post-doc au département d'informatique de l'université de Pennsylvanie pour tenter ma chance de discuter avec lui quand ça lui conviendrait.

Je l'ai donc revu à l'automne 1978, chez lui à New York. Il habitait le rez-de-chaussée d'une petite maison dans le Village acquise en communauté avec des amis à lui, anarchistes et juifs, musiciens pour certains. Lui-même jouait du violon et le piano Steinway de concert de sa femme trônait dans leur salon. Il m'a raconté qu'il passait un peu plus de la moitié de l'année dans son kibboutz pro palestinien en Israël, où il était conducteur du car qui emmenait les gens aux champs d'orangers, ce qui lui permettait de réfléchir et d'écrire en paix toute la journée en attendant de les ramener. Il m'a précisé qu'ainsi il n'était jamais dérangé par des questions de linguistes qui auraient pu lui faire perdre le fil de son propre débat intérieur. Il venait quelques mois à New York, avec sa femme, qui avait conservé une importante activité professionnelle en physique théorique, après avoir été la légataire universelle d'Einstein et après avoir hésité entre une carrière de pianiste et une carrière de physicienne. Il en profitait pour voir des amis et collègues et s'occuper de ses publications.

A ma première visite, je lui ai posé des questions très iconoclastes sur sa théorie transformationnelle, remettant également en question tout le cadre de mon travail de thèse, ce qui l'a fait rire. Il a appelé sa femme pour venir écouter ça. Elle a ri aussi et lui a dit devant moi qu'elle n'avait pas le temps d'écouter ni de formaliser sa nouvelle théorie des opérateurs mais qu'il avait peut-être trouvé l'oiseau pour faire ça. Lui-même avait abandonné la notion de transformation et avait imaginé une représentation grammaticale directe des structures lexicales et de leurs réductions avec comme seules catégories : les notions d'élément lexical élémentaire et d'élément lexical opérateur (avec élément grammatical marqueur d'argument intégré à l'opérateur lorsqu'un marqueur d'argument est obligatoire). Par exemple, *l'aptitude de Jean au dessin* et *Jean est apte au dessin* auraient des structures voisines : *aptitude-de-à (Jean, dessin)* et *apte-à (Jean, dessin)*.

Harris me donnait rendez-vous environ une heure par quinzaine pour me présenter un par un les chapitres de sa nouvelle théorie des opérateurs, qui allait devenir Harris (1982). Il me donnait généralement rendez-vous dans des halls de grands hôtels, au milieu du passage et du bruit d'un luxe feutré ou parfois dans un petit restaurant vietnamien où il emmenait aussi sa fille déjeuner. Il m'a expliqué un jour où je riais de ne pas l'entendre, que c'était pour mieux nous concentrer sur le fond.

Quand j'ai eu une vue d'ensemble de sa nouvelle théorie, je me suis mise au travail de formalisation, dans une théorie des types simples c'est-à-dire une théorie où les seules catégories sont celles d'opérateur et d'argument et sont des constantes (i.e. ne varient pas selon le contexte). Conformément à ses hypothèses, la règle de réduction syntaxique n'est plus une opération de transformation sur la structure d'interprétation mais une simple réduction de forme de surface des opérateurs ou arguments lexicaux (voir la section suivante). La structure des énoncés devient une structure de l'information littérale, la modifier implique de changer la valeur de l'énoncé. Cette règle de « réduction de redondances » telles que Harris (1982 et 1990) l'a précisée, concerne principalement des éléments référentiels, c'est-à-dire des éléments déjà connus dans le contexte, des éléments indéfinis simples ou opérateurs (infra) qui n'apportent pas d'information nouvelle ou encore des réductions métonymiques où les mots opérateurs désignant une contenance matérielle, temporelle ou encore spatiale sont si attendus qu'ils peuvent être omis. Cette opération de réduction pose de multiples questions, en particulier des questions de « complexité » au sens technique du terme, sur lesquelles je reviens plus bas.

Après quelques semaines de formalisation qui m'ont amenée à préciser les hypothèses de la première théorie linguistique sans parties du discours, jamais imaginée, je me suis aperçue que cette

théorie était inconsistante. Je reviens dans la section suivante sur les raisons de cette inconsistance. Je n'avais pas à ce moment de solution de rechange. Nous en avons parlé, le point l'a préoccupé, il en a parlé à ses proches, Henri et Danuta Hiz ainsi qu'à son frère et à sa belle-soeur, biologistes. Harris a alors décidé de faire l'hypothèse qu'il pourrait contourner ce problème en codant un corpus de textes appartenant à un univers de discours spécifique et en montrant que leurs structures informatives dans cet univers de discours étaient strictement finies. Il a alors lancé un projet sur l'analyse d'un gros corpus de textes scientifiques, avec l'hypothèse qu'il prouverait la possibilité de coder de façon finie l'information de ces textes à partir de leur formulation linguistique. Ce corpus concernait une controverse en biologie sur la réaction antigénique à laquelle son frère et sa belle-sœur avaient participé.

Ce projet a mobilisé sept personnes pendant quatre ans, de 1980 à 1984, il était basé à l'université de Columbia. Il a donné lieu à un cycle de conférences données par Harris à Columbia en 1986, publiées dans Harris (1988), sur ses idées générales. Les résultats du projet sur la structure de l'information dans ce corpus ont été réunis sous forme d'un livre collectif en 1984 et finalement publiés dans Harris et al. (1989). À l'issue de ce projet, le problème linguistique lié à l'inconsistance de sa théorie s'est avéré non contournable. J'ai montré dans mon chapitre de Harris et al. (1989) que la structure informative en effet finie de ce corpus, était en réalité indépendante de sa formulation linguistique et même de la langue utilisée, comme je le résume plus bas.

Les amitiés et l'estime que Harris a suscitées ont joué un grand rôle pour lui dans différentes occasions. Par exemple, Anna Morigio-Davis grande spécialiste de l'indo-européen l'a aidé à publier chez Oxford University Press sa théorie des opérateurs, Marcel-Paul Schutzenberger mathématicien français l'a aidé à publier chez Kluwer le volume sur l'analyse des textes scientifiques, Peter Matthews a fait financer par Columbia et la NSF ce projet. Ce projet a aussi financé mes voyages et séjours biannuels pendant quatre ans pour parler en fait surtout de grammaire générale et de formalisation et a renforcé encore notre amitié. Grâce au rapport de recherche que Harris m'avait fait en 1979, j'avais eu sans difficulté un poste au CNRS en informatique théorique, ce qui m'a donné tout loisir de participer aux séminaires, colloques et discussions de logiciens intuitionnistes déjà bien convaincus des limitations des calculs de prédicats ou des théories finitistes des opérateurs, ou plus précisément des lambda calculs typés au premier ordre, pour construire des langages de programmation performants. Ces langages de programmation permettaient d'optimiser le déroulement en parallèle des calculs. J'ai proposé dans Daladier (1986) une solution théorique permettant de conserver l'hypothèse d'une grammaire sans catégories linguistiques prédéfinies mais à condition d'étendre les opérateurs représentant les éléments lexicaux à des variables d'opérateurs et non pas à des constantes d'opérateurs comme dans Harris (1982), théorie dont j'avais montré l'inconsistance dès 1979. Les variables d'opérateur ont généralement des structures finies dans le contexte des constructions qui définissent leur structure d'interprétation. Comme la théorie de l'information dans les textes scientifiques ne permettait pas de résoudre la question de la structure d'interprétation non finie des énoncés dans des langues comme le français ou l'anglais (voir plus bas), j'ai donc finalement proposé une grammaire constructiviste au sens des mathématiques intuitionnistes de Brouwer et de Heyting, dans Daladier (1990b). Privilégiant notre amitié, je n'ai pas cherché à publier ce résultat dans une revue scientifique mais j'ai rédigé un texte présentant le problème et sa solution pour des chercheurs en informatique théorique dans un rapport technique du LADL de 1986, après l'avoir présenté en 1985 au séminaire de Jean-Louis Krivine à Paris 7. Après la publication du livre sur la structure informative de textes scientifiques, j'avais proposé à Harris de faire une synthèse de ce que lui et ses principaux adeptes proposaient, dans le numéro de 1990 de la revue *Langage*. J'ai donc publié dans ce numéro deux articles présentant le résultat d'inconsistance de sa théorie avec ma solution consistant à élargir sa théorie en considérant tous les éléments lexicaux à arguments phrastiques comme des opérateurs dont la portée varie selon l'énoncé dans lequel ils sont utilisés. Cette solution implique donc une importante modification de la notion d'opérateur lexical et une très significative généralisation de la règle de réduction de manière à ce que les énoncés de base ne comportent pas plus d'un opérateur à argument phrastique. L'un des résultats, pour moi inattendu à cette époque, de cette généralisation de la règle de réduction, est qu'elle rend compte de l'évolution diachronique à long terme des mécanismes d'insertion d'argument phrastique comme des changements de forme morphologique et non comme des changements de structure informative des énoncés, comme je l'indique dans § 5. Autrement dit, la généralisation de la théorie des opérateurs que

j'ai proposée permet de lever le problème d'inconsistance et ce faisant de décrire l'évolution des phrases à argument phrastique du sanskrit védique au français moderne comme une évolution de formes morphologiques permettant de condenser l'information linguistique, ce qui n'implique en rien une évolution de la complexité des structures grammaticales des langues concernées. Je reviens de façon plus concrète sur les questions de complexité avant de conclure.

J'ai envoyé à Harris les articles de ce numéro de *Langage* avant leur publication pour qu'il puisse éventuellement faire des commentaires. Nous avons échangé une correspondance. Ensuite nous nous téléphonions de temps en temps. Son dernier coup de téléphone a été pour m'annoncer lui-même, sans émotion, qu'il allait mourir dans les deux ou trois jours et qu'il avait été heureux de me rencontrer.

2. Ce que vise à décrire Harris (1982) dont ne rendent pas compte les grammaires formelles ni la typologie. Ce dont la typologie cherche à rendre compte et que Harris (1982) ne prend pas en considération.

Le principal objectif de Harris (1982) est de rendre compte de la façon dont une langue structure l'interprétation littérale de ses énoncés. Il considère à juste titre que les propriétés lexicales, en particulier les réductions de redondances rendent caduque la structuration des énoncés à partir des catégories traditionnelles des grammaires transformationnelles.

Aucune des théories linguistiques, hormis celle de Harris (1982), ne traite de la question des ellipses syntaxiques, qui ne peuvent être confondues avec les implicatures pragmatiques, et sur laquelle je reviens de façon détaillée plus bas. En l'absence d'un traitement de cette question, aucune grammaire ne peut rendre compte de la structure d'interprétation des énoncés dans leur ensemble dans aucune langue du fait de propriétés d'ellipse d'argument de verbes transitifs et de noms opérateurs transitifs, soit indéfinis soit référentiels selon les constructions (voir §4). De façon générale, la théorie des constructions de Croft (2001) est intéressante en particulier parce qu'elle a donné lieu à de nombreuses études typologiques descriptives montrant l'insuffisance des conceptions syntaxiques traditionnelles avec leurs structures hors contexte, qui sont utilisées finalement de façons assez voisines en grammaire générative et en typologie. La syntaxe telle qu'elle est définie en grammaire générative et en typologie ne peut analyser l'ellipse grammaticale et la théorie des constructions non plus puisqu'elle a pour principe de rendre compte du sens des procédés contextuels de transitivisation ou d'intransitivisation et de la valence des éléments lexicaux selon les constructions où ils s'emploient. De ce point de vue la théorie des constructions et la théorie des opérateurs ont des défauts et des qualités inverses du fait même de leurs méthodes. On peut cependant élargir ces deux approches pour ne conserver que leurs qualités, ce que j'esquisserai ici.

Les grammaires fonctionnalistes de la typologie utilisent des catégories ou fonctionnalités prédéfinies souvent inadéquates mais décrivent de nombreuses langues orales inconnues (et en voie de disparition) mettant en évidence des procédés de grammaticalisation généralement contraints, souvent inédits, tous porteurs d'une sémantique grammaticale. La typologie analyse surtout la sémantique véhiculée par les différents procédés non lexicaux des langues : morphologies et grammaticalisations d'éléments lexicaux, question fondamentale puisqu'elle est le principal aspect de la diversité des langues et de leurs évolutions. Bien que la typologie actuelle défende souvent l'iconicité de la sémantique lexicale et de la sémantique grammaticale en guise de propriétés universelles de langage, l'analyse détaillée des procédés de grammaticalisation dans des langues, et pas seulement le découpage lexical, reflète les conceptions culturelles des locuteurs, que Harris n'a pas traitée et sur laquelle je reviens également en détail à la fin de cet article. De façon extrêmement intéressante, Spinoza (1677¹ 1968²) précise pourquoi il propose lui-même une grammaire de l'hébreu, alors qu'il en existe déjà plusieurs, en montrant que l'interprétation théologique et philosophique des Écritures ne peut être indépendante de la langue dans laquelle elle est formulée et en montrant également que le lecteur non familier de l'hébreu doit être averti des différences grammaticales et non pas seulement avoir accès à une traduction. Il montre en particulier de façon précise que la notion de temps est très différente en latin et en hébreu, l'hébreu a ce que nous appellerions, une notion d'aspect et non de temps proprement dit, la notion de « présent » n'étant qu'un point à la limite d'une convergence entre l'accompli et le prospectif. Il explique comment l'actualisation subjective de l'énonciation par le locuteur se substitue à un présent qui affecterait une chose ou un être. Spinoza (1968) montre au

chapitre 5 que la conception du temps dans les Écritures est liée au fait qu'il n'y a pas de « verbe » au sens latin du terme mais plutôt différentes sortes de « noms ». Il montre ainsi au chapitre 10 que les « prépositions » sont des sortes de « noms » car elles ont un pluriel. Le pluriel de l'équivalent de « entre » exprime non pas la relation d'une chose individuelle avec une autre mais les « intervalles » spatiaux ou temporels entre ces choses.

3. Propriétés de portée variable d'opérateurs selon la construction

Une autre question non traitée par Harris ni par les autres approches grammaticales bien qu'apparaissant dans toutes les langues et mettant en question toutes les grammaires finitistes, est celle de la portée variable de mots sur d'autres selon les constructions.

Par exemple, le sens lexical d'un verbe d'une proposition principale peut contraindre le mode du verbe subordonné avec l'interprétation réalisé ou irréalisé, comme en français le mode indicatif avec *espérer* dans (1) ou souhaiter dans (2), mais irréalisé dans (2) et (3). Une négation peut modifier ces propriétés lexicales d'une façon qui dépend du sens lexical de façon subtile et qui ne peut se représenter de façon simple parce qu'elle implique des relations de dépendance à plusieurs termes (4):

- (1) *Jean (espère + a dit) que tu viendras.*
- (2) *Jean souhaite que tu viennes.*
- (3) *Jean n'espère plus que tu viennes.*
- (4) *Jean n'a pas dit que tu viendras.*

D'une façon générale, les parties du discours ne peuvent décrire qu'une toute petite partie des propriétés de portée des mots et des structures d'interprétations, dont une partie dépend de propriétés d'éléments grammaticaux comme l'inflexion de mode et dont l'essentiel relève de restrictions de sélection de toutes sortes d'éléments lexicaux opérateurs. Adjectifs, adverbes, prépositions et noms opérateurs sont susceptibles de portée variable selon les constructions. Les opérateurs lexicaux sont des mots qui s'interprètent selon leurs restrictions de sélection sur d'autres mots. Par exemple, les adjectifs *vif* et *petit* n'ont pas la même portée dans (5) et (6) :

- (5) *Jean a un chandail bleu trop vif.*
- (6) *Jean a un chandail bleu trop petit.*

La préposition *dans* n'a pas la même portée ou plus précisément ne s'applique pas aux mêmes arguments dans (7) et (8) :

- (7) *Jean a bu son café dans un fauteuil.*
- (8) *Jean a bu son café dans un verre.*

Ces propriétés de portée des mots opérateurs mettent en jeu des restrictions de sélection lexicales à plusieurs termes et pas seulement le type syntaxique : opérateur/ nom simple, ni seulement les catégories grammaticales traditionnelles et leurs dépendances simples. Les propriétés de portée variable des mots opérateurs (adjectifs, adverbes, prépositions, verbes, certains noms) selon les constructions sont un aspect important de la complexité grammaticale. On peut ramener cette question à celle de l'ellipse en considérant (7) et (8) comme des formes réduites de (9) et (10) :

- (9) *Jean₁ a bu son café, Ø₁ étant dans un fauteuil.*
- (10) *Jean a bu son café₂, Ø₂ étant dans un verre.*

La structure syntaxique de ces exemples est relativement simple, en revanche la reconstruction des ellipses ou de la structure argumentale de l'opérateur *dans* des constructions (7) et (8) suppose une connaissance lexicale a priori non finie, vu les possibilités non limitées de combinaisons des opérateurs à argument opérateur. Par exemple dans (11) et (12) le premier opérateur *dans* opère sur le second opérateur *dans* selon la structure de (7) et (8) respectivement, alors que *avec* a une portée sélective dans (13) et (14) :

- (11) *Dans ce restaurant, on vous_x sert le café dans un fauteuil_x.*
 (12) *Dans ce restaurant, on vous sert le café_x dans un verre_x.*
 (13) *Dans ce restaurant on_x vous sert un café avec un smoking_x.*
 (14) *Jean_x a bu son café dans un fauteuil_x avec le sourire_x.*
 (15) *Jean a bu son café_x dans un verre_x avec un nuage de crème_x.*
 (16) *Jean_x a bu son café_y dans un verre_y avec le sourire_x.*

Les énoncés (5) à (16) sont généralement considérés par les linguistes comme des phrases simples par opposition à des énoncés à argument phrastique comme : (17) *Jean veut que tu viennes* ou (18) *Jean aime être dans un fauteuil*. Alors que la structure d'interprétation de (17) et (18) peut être décrite dans une grammaire finitiste si on dispose d'un lexique contenant des informations sur l'induction lexicale du mode ou sur les propriétés lexicales de « projection lexicale » (i.e. identification du sujet de l'infinitive avec un argument particulier du verbe de la principale), ce n'est pas le cas de (5) à (16) dont les structures d'interprétation dépendent directement de la construction et non pas de propriétés lexicales indépendantes du contexte.

Les propriétés de portée variables de ces exemples (5) à (16) ne sont pas régies par des propriétés lexicales que Harris appelle des probabilités d'occurrence. Ce terme cache en fait un certain scientisme (voir § 6.1) bien qu'il soit moins réducteur que la notion d'acceptabilité chomskienne. Un énoncé comme : *Les idées immatures dorment furieusement* fait intervenir une sélection de *idées immatures* par *dormir* qui n'est pas « improbable » au sens de la statistique mais qui correspond à une extension lexicale. Une fois assertée, elle est banalisée et perd son caractère « improbable ».

D'un autre côté, la notion « d'acceptabilité » se confond avec la validation propositionnelle et conduit à rejeter une phrase grammaticale comme : *Les idées vertes dorment furieusement* au motif qu'elle serait ininterprétable. Cette vue s'avère scientiste parce que arbitraire et limitée puisqu'un contexte fournit une interprétation ordinaire à de telles phrases. Il suffit de lui associer une première phrase simple telle que : *Les idées vertes sont des idées immatures que le sommeil permet d'élaborer. Les idées vertes dorment furieusement !* De façon plus générale, la notion chomskienne d'acceptabilité se révèle aberrante comme critère de grammaticalité lorsqu'on considère les énoncés d'un univers de discours scientifique assez technique comme la biologie ou la physique pour lequel un locuteur non biologiste ou physicien ne peut avoir de jugement sur la validité des énoncés. C'est précisément ainsi que de nouvelles disciplines scientifiques se créent avec un usage lexical spécifique. Un énoncé grammatical tel que : *les antigènes produisent des anticorps* n'a aucun sens en biologie mais un chercheur pourrait peut-être le transformer en question dans un contexte approprié.

Si on revient aux exemples précédents, une portée variable peut être ambiguë comme dans : *Dans ce restaurant on vous sert en smoking*, qui s'interprète soit comme *le service se fait en smoking* soit comme *ce restaurant ne sert que des clients en smoking*.

Le fait qu'un opérateur lexical requérant un argument élémentaire ne soit pas porteur de marques d'ancrage (TAM) et d'accord, peut rendre sa portée indéfinie comme *jeune* dans (19) par opposition à (20) où l'accord de genre spécifie la portée de l'opérateur *petite* : (19) *Laure a rencontré Luc jeune* (ambiguë) ; (20) *Laure a rencontré Luc petite*.

Ces quelques exemples de portée variable selon le contexte suffisent à établir que les structures d'interprétation ou structures syntaxiques des combinaisons d'opérateurs lexicaux ne peuvent être définies en dehors des constructions. De ce point de vue la théorie des constructions de Croft (2001) est très intéressante mais elle n'est pas suffisante car dépourvue d'analyse de l'ellipse grammaticale, ce qui ne permet pas d'interpréter correctement par exemple des éléments lexicaux à argument indéfini ou référentiel omis sans changement de valence. Je reviens sur cette question dans la prochaine section.

4. Conception distributionnelle du sens, absence d'iconicité et absence de toute forme de traduction propositionnelle d'énoncés pour analyser le sens linguistique en dehors d'univers de discours spécialisés. L'ellipse grammaticale comme propriété universelle.

Pour Harris comme pour Sapir, le sens d'un mot opérateur comme *aimer* en français dépend de ses restrictions de sélection. Les structures d'insertion lexicales, c'est-à-dire pour Harris les structures syntaxiques, avec les réductions et ellipses sous-jacentes doivent rendre compte du sens des combinaisons de mots dans les assertions, le sens des constructions. Cette élaboration d'une notion syntaxique des réductions et cette élaboration du sens lexical en général manquent à la Théorie des Constructions de Croft. Par exemple *aimer* en français vient du latin *amor* dont le sens en latin ancien est *faire l'amour* mais n'en est pas une simple traduction car le sens d'être amoureux n'est possible parmi d'autres qu'avec des êtres animés. Comme beaucoup d'éléments lexicaux initialement à argument(s) simple(s), *aimer* en français, comme des verbes de sens voisin dans les langues romanes et germaniques, a aussi innové la possibilité de se construire directement avec un argument phrastique : *j'aime regarder la lune ; j'aime aimer*, ce qui étend son sens lexical. En war, langue austro-asiatique, on utilise plusieurs mots de racines distinctes pour exprimer *aimer* selon que l'objet est animé ou inanimé et selon qu'on exprime un sentiment amical, une amitié ou l'intimité qu'elle soit amoureuse ou maternelle, donc aucun d'eux n'a le sens exact d'*aimer*, distinct lui aussi des deux mots anglais *like* et *love*. D'une façon plus générale, Harris comme Sapir rejette la notion d'iconicité lexicale. Cette absence d'iconicité exclut d'emblée toute forme d'analyse propositionnelle des énoncés dans un cadre général. Je reviens plus bas sur ce point.

Les grammaires de constituants syntagmatiques utilisées actuellement en typologie ou en grammaire générative reflètent une conception très simple de la grammaire des langues en terme technique de complexité. Les dépendances syntaxiques sont limitées à des dépendances linéaires. Elles considèrent leurs catégories syntaxiques et morphologiques : parties du discours, flexions ou bases verbales avec TAM et accord du verbe avec un ou plusieurs arguments, arguments sujet/ agent, objet etc. et l'ordre des mots, comme des notions nécessaires, voire suffisantes, pour définir la grammaire des langues. Ces grammaires limitent la notion de prédication aux structures argumentales (ou actanciennes) des verbes. Cette notion traditionnelle de prédication intègre la structure argumentale de l'élément lexical verbal aux éléments morphologiques d'ancrage comme la flexion ou l'affixation des TAM et l'accord argument(s)-verbe. Ainsi contrairement à Harris (1982) *description* n'est pas traité dans les syntaxes de constituants syntagmatiques comme un élément lexical transitif ni même comme un prédicat dans : *la description de Luc de ce paysage*, ou dans : *la description de ce paysage est de Luc*. Pourtant toutes les nominalisations sont assertables avec une copule ou avec un verbe dit support ou « light verb », en fait un auxiliaire de nom comme *faire* ou *donner* dans : *Luc a fait une description de ce paysage*, *Luc a donné une description de ce paysage*.

La plupart des grammaires actuelles traitent la morphosyntaxe, la sémantique et la pragmatique indépendamment, à des niveaux différents utilisant des catégories de description différentes. Récemment certains linguistes revendiquent pour des langues orales la nécessité de structures essentiellement informationnelles c'est-à-dire pragmatiques avec une syntaxe réduite. Harris (1982) est le premier linguiste à avoir tenté de définir la syntaxe d'une langue comme l'ensemble de ses structures d'interprétation, utilisant en cela le terme « syntaxe » au sens des logiciens et informaticiens sans chercher à calquer la sémantique des calculs propositionnels sur la structure d'interprétation des énoncés d'une langue. Je montre plus bas que la tentative de Harris de réduire la syntaxe aux structures lexicales est intenable. En réalité les oppositions que dressent les approches linguistiques entre morpho-syntaxe, sémantique et pragmatique sont parfois arbitraires et mieux vaut revenir au terme plus neutre de « grammaire » en précisant le sens linguistique que l'on cherche à analyser et la nature des contraintes de la grammaire qui sont plus ou moins fortes selon qu'elles tiennent au cœur de la grammaire ou à des options discursives ou stylistiques.

Harris (1982 et 1991) généralise de façon considérable la représentation du sens distributionnel des éléments lexicaux et grammaticaux, ou selon une terminologie plus actuelle, le sens des constructions, en généralisant la notion de structure argumentale à divers éléments lexicaux : noms, adjectifs, prépositions, conjonctions et adverbes qui imposent des restrictions de sélection lexicale sur un ou plusieurs arguments. Les arguments des verbes comme ceux des autres éléments peuvent être élémentaires ou phrastiques selon leurs restrictions de sélection.

Un des principaux avantages des grammaires de Harris pour la prise en compte du sens des constructions lexicales par opposition aux représentations syntagmatiques et à la théorie des Constructions, est de rendre compte de propriétés d'ellipses grammaticales ou lexicales, présentes dans les différentes familles de langues, à l'intérieur de la représentation grammaticale. Par exemple, les verbes transitifs peuvent se construire sans objet apparent sans pour autant s'intransitiver, c'est-à-dire sans dédoublement de leur sens lexical, comme dans (21) ou (22) :

(21) *Jean a décidé de manger Ø.*

où l'argument agent de manger doit être identifié à l'agent de décider et où l'objet de *manger* est indéfini, ce que Harris précise en considérant que l'indéfini *quelque chose* est l'argument implicite de *manger* dans (21), réductible comme tel à une forme zéro. Ceci distingue *manger* dans (21) de *partir* dans (22) qui conserve un argument agent implicite, identifié obligatoirement à l'agent de décider mais n'a pas d'objet :

(22) *Luc_x a décidé de Ø_x partir.*

De fait, les opérateurs lexicaux transitifs conservent leur structure d'interprétation lexicale transitive en cas d'ellipse. Un argument implicite, objet direct ou indirect ou sujet, est soit indéfini comme dans (21), soit dans certains contextes est susceptible de devoir être reconstruit syntaxiquement. Par exemple dans (23) et (24) :

(23) *Jean₁ a promis à Luc de Ø₁ venir le voir.*

venir a bien un argument sujet et celui-ci doit obligatoirement être identifié au sujet de *promettre*, c'est-à-dire *Jean* ; c'est une propriété de projection lexicale de *promettre* différente de celle de *demander* :

(24) *Jean₁ a demandé à Luc₂ de Ø₂ venir le voir.*

La notion morphosyntaxique d'ellipse est également nécessaire pour des réductions « de mots appropriés » selon la terminologie de Harris, comme les réductions métonymiques, bien connues des grammairiens, par exemple dans (25) :

(25) *Jean a bu Ø toute la bouteille.*

On doit interpréter (25) comme : *Jean a bu le contenu de toute la bouteille*. Ici c'est d'un point de vue lexical que *bouteille* ne peut pas s'interpréter comme argument objet de *boire*. En revanche dans :

(26) *Jean a cassé la bouteille.*

bouteille est bien l'argument objet lexical et syntaxique de *casser*. Ces réductions métonymiques sont des processus très courants dans probablement toutes les langues. Elles concernent aussi des animés, comme dans (27) :

(27) *Jean a charmé toute la salle.*

D'autres formes d'ellipses sont des éliminations de redondances lorsqu'un opérateur est très attendu dans un contexte comme dans (28) :

(28) *L'orchestre a commencé Ø la symphonie.*
Ø = à jouer

par opposition à (29) qui ne serait pas assertée sans un contexte très particulier :

- (29) *L'auditoire a commencé Ø la symphonie.*
 Ø = à écouter

Les ellipses concernent aussi l'argument phrastique des verbes ou noms opérateurs à argument phrastique, comme dans (30) :

- (30) *Jean est rentré chez lui. L'erreur de Jean Ø nous a étonnés.*

La reconstruction syntaxique de l'argument non élémentaire de *erreur* dans la construction (30) est obligatoire du fait de la transitivité lexicale de *erreur* et du fait que l'argument élémentaire (i.e. le sujet) de *erreur* est *Jean*, qu'on identifie au sujet de la première sous assertion de (30) : *Jean est rentré chez lui*.

- (31) *Jean est rentré chez lui. Cette erreur Ø nous a étonnés.*

Le démonstratif *cette* déclenche la recherche d'une co-référence. L'argument phrastique non instancié de *erreur* est identifié au seul candidat possible dans la co-assertion, c'est-à-dire le premier argument de la co-assertion : *Jean est rentré chez lui*. Dans (32), le déterminant indéfini ne bloque pas l'identification des arguments de *erreur*. En revanche dans (33) le déterminant indéfini bloque l'interprétation référentielle et l'argument phrastique implicite de *erreur* reste indéfini. A moins que la co-assertion ne comporte un troisième membre de phrase qui permette une reconstruction comme dans (34) où le déterminant indéfini *une* permet de désambiguïser la référence de l'argument phrastique de *erreur* en sélectionnant l'élément phrastique non encore connu au moment de l'assertion de *une* :

- (32) *Jean est rentré chez lui, une erreur Ø qui nous a étonnés.*
 (33) *Jean est rentré chez lui. Une erreur Ø nous a étonnés.*
 (34) *Jean est rentré chez lui. Une erreur Ø_x nous a étonnés. La porte était restée ouverte_x.*

La reconstruction de : *erreur de Jean d'être rentré chez lui* n'est due ni à la sémantique de *erreur*, ni à la pragmatique du discours de (30), (31) et (32) mais à la structure argumentale des éléments lexicaux, qui peut demeurer implicite selon les constructions, et à divers emplois coréférentiels.

La question du changement contextuel de structures argumentales se pose cependant aussi dans la théorie des opérateurs de Harris où elle n'est pas traitée correctement. Par exemple les noms déverbaux opérateurs expriment un procès quand ils sont sélectionnés par un verbe à argument phrastique (argument opérateur) mais beaucoup sont ambigus du fait d'une lexicalisation (figement) avec interprétation de résultat de l'action et interprétation de nom simple quand ils sont sélectionnés par un verbe à argument simple. Dans (35) emballage, comme *emballer* dans (36), est un opérateur à deux arguments *lampe* et *Luc* lui-même argument d'un opérateur à argument phrastique *prendre du temps*, tandis que dans (37) *emballage* est l'argument élémentaire de *se froisser* :

- (35) *L'emballage de la lampe par Luc a pris du temps.*
 (36) *Que Luc ait emballé la lampe a pris du temps.*
 (37) *Cet emballage s'est froissé.*

Ce processus grammatical de lexicalisation d'opérateurs verbaux nominalisés avec divers suffixes qui peuvent aussi s'employer métonymiquement comme noms élémentaires est aussi courant que la métonymie dans les langues du monde. Son traitement suppose une information lexicale *ad hoc* mais surtout la représentation des structures argumentales et le déclenchement de reconstructions éventuelles selon le (ou les) type (s) syntaxiques d'opérateur ou de noms simples de chaque élément lexical explicite dans la construction. Le type d'un opérateur est défini par le nombre d'arguments et leur statut élémentaire ou opérateur.

5. L'inconsistance des hypothèses de Harris sur le caractère finitiste de sa grammaire des opérateurs. Ce qu'apporte cette inconsistance à la compréhension de la complexité linguistique et au changement diachronique et comment la décrire.

L'ellipse d'un argument phrastique peut ne pas être restructurable dans le contexte d'une co-assertion et rester indéfinie comme dans (39), tout comme un argument élémentaire comme dans (38) :

(38) *Luc est en train de manger Ø.*

(39) *Luc est en train de comprendre Ø.*

Comprendre n'est pas plus intransitivé dans (39) que *manger* dans (38), son argument phrastique pourrait devoir être restructuré dans une co-assertion qui référencerait cet argument implicitement. Mais s'il ne peut être restructuré, il a une interprétation indéfinie de *quelque chose* :

(40) *Luc est en train de manger quelque chose.*

(41) *Luc est en train de comprendre quelque chose.*

Cependant ce *quelque chose* n'a pas la même structure d'interprétation dans (40) et (41) du fait du statut élémentaire et non élémentaire respectivement de leur second argument, comme le montre la possibilité d'asserter (42) mais pas (43) :

(42) **Luc est en train de manger quelque chose qui est que ce poisson est trop salé.*

(43) *Luc est en train de comprendre quelque chose qui est que ce poisson est trop salé.*

Définir le statut syntaxique de l'indéfini *quelque chose* ou de son ellipse dans (40) et (39) et (41) n'est pas possible dans le cadre des syntaxes de Harris (1982 et 1991). Lorsque *quelque chose* est employé comme opérateur mais qu'il n'a pas d'argument phrastique explicite c'est-à-dire fini, son ellipse ne peut être définie dans aucune des théories de Harris car selon le contexte de co-assertion, il pourra opérer sur une phrase simple ou sur une phrase comportant un enchaînement variable d'arguments phrastiques par exemple : *Luc est en train de comprendre que tu sais que Jean regrette que ce projet soit annulé.*

Aucune grammaire de constituants syntagmatiques ne peut non plus décrire ce phénomène. Il existe une solution algébrique à cette question, qui consiste à définir les opérateurs lexicaux dans un lambda calcul typé au second ordre, c'est-à-dire où les opérateurs et arguments sont des variables, hors contexte et dont la structure s'analyse dans le contexte d'une construction. Les opérateurs indéfinis opèrent sur des points fixes quand ils ne peuvent pas être identifiés contextuellement, ce qu'on pourrait gloser par *quelque chose* opérant sur lui-même : *quelque chose de quelque chose*. Certains verbes à argument phrastique non explicite se construisent avec un quantifiant ce qui implique un argument phrastique indéfini s'appliquant à lui-même comme dans (45) ou, s'il y a un contexte, une référence au contexte qui modifiera la valeur de l'indéfini :

(44) *Elle comprend tout Ø.*

(45) *Elle comprend toute chose.*

Certains opérateurs lexicaux ont des structures d'interprétation qui peuvent dépendre en partie de contraintes avec d'autres mots dans la construction où ils sont insérés. Dans un sens technique, ils ont des types syntaxiques variables : ce sont des variables d'opérateur. Cette solution technique est détaillée dans Daladier (1986 et 1990). Cette extension de la théorie des opérateurs de Harris rend compte de structures syntaxiques non finies tout en garantissant des reconstructions finies d'ellipses ou de réductions. De plus, et surtout, cette solution rend compte de changements diachroniques, qui permettent de comprendre comment des langues de familles différentes produisent, au cours de leurs évolutions, des formes morphologiques variées de dépendance phrastique. Autrement dit, de façon qui me semble fascinante, cette extension non finie de la structure syntaxique à partir d'éléments indéfinis permet de comprendre pourquoi certains aspects de la structure morphosyntaxique de familles de langues différentes changent diachroniquement.

Par exemple, la genèse de la complétivation dans des langues indo-européennes se fait à partir de structures corrélatives où s'identifient deux sortes de pronoms indéfinis référentiels, notamment en sanskrit védique, voir Renou (1952 : 386-9). Ces formes indo-européennes anciennes pourraient se traduire en français par des co-assertions comme dans (46) où les deux différents éléments indéfinis correspondant au français *quelque chose* et *cette chose* sont en corrélation :

(46) *Votre ébranlement des hommes est quelque chose ; cette chose est la manifestation de votre force.*

Le sanskrit védique a un indéfini *yá* 'quelque chose, 'celui qui ...'. L'élément *yá-* peut s'employer comme un article indéfini et comme pronom relatif. *yád* 'qui-' est la forme neutre du relatif, littéralement 'chose qui/ que', correspond au *que* complétif du français également morphologiquement lié au pronom relatif (comme c'est le cas également en latin). Cet élément *yád* pourrait se traduire par *ce qui est* ou par *que* dans la traduction en français (48) ou (49) de l'exemple sanskrit védique (47) :

(47) *yád dha vo bálam jánām acūcyavitāna* I 37 12, Renou (1952 : 387)

(48) Vous avez ébranlé les hommes, **ce qui est** une manifestation de votre force.

(49) **Que** vous ayez ébranlé les hommes est une manifestation de votre force.

J'ai proposé dans Daladier (2001) une analyse en terme de réductions de différentes formes de co-assertions contenant des indéfinis pour analyser les structures d'interprétation des déterminants, des relatives et des complétives en français. Ainsi, comme *ya(h)* en sanskrit védique, le *que* complétif en français s'apparente morphologiquement à *que* pronom relatif et à la reduplication dans l'indéfini *quelque*. On peut aussi analyser grammaticalement : synchroniquement aussi bien que diachroniquement la complétivation à partir d'une corrélation d'éléments opérant sous un indéfini *quel (que)* et considérer (50) et (51) comme des assertions qui ont même structure grammaticale et même structure d'interprétation et où (51) est une forme réduite de (50) :

(50) *Luc est en train de comprendre quelque chose ; quelque chose est que ce projet est impossible...*

(51) *Luc est en train de comprendre que ce projet est impossible.*

Ce type de décomposition syntaxique se révèle intéressant pour analyser les différentes formes que peut prendre un argument non élémentaire : complétive, infinitive ou nominalisation. Je montre également dans Daladier (2001) que l'analyse de l'insertion des articles définis et indéfinis en français, dont l'interprétation dépend des constructions dans lesquelles ils sont insérés : indéfini spécifique/ non spécifique/ générique ; défini référentiel ou générique, fait intervenir très précisément la même complexité structurelle que l'analyse des propositions relatives ou des complétives. Mon analyse rend compte aussi de la genèse diachronique de la catégorie des « déterminants » définis et indéfinis : *un(e), des, le, la, les*, et ce au même titre que les « pronoms » relatifs et que la « conjonction » complétive. L'insertion des déterminants n'est pas décrite dans la théorie des opérateurs de Harris et ne peut l'être : dans sa théorie ils ne sont ni des marqueurs d'arguments, ni des arguments, ni des opérateurs. Ils sont cependant porteurs d'information pertinente et là aussi cette information dépend de la construction : générique pour l'indéfini et le défini, spécifique ou non pour l'indéfini, référentielle pour le défini.

6. Questions posées par la règle de réduction linguistique de Harris

6.1 Questions formelles

Les réductions de formes s'appliquent aussi bien à des éléments qui ont une structure élémentaire qu'à des éléments qui ont une structure phrastique, notamment à des formes pronominales comme *cela*, dénotant une coréférence à un nom simple (52) ou à un élément phrastique (53) :

(52) *Ils mangent du poisson ; cela sent mauvais.*

(53) *Ils mangent du poisson ; cela m'a étonnée.*

Les réductions à des formes zéro concernent aussi des formes indéfinies *quelque chose*, *choses* de type élémentaire ou de type phrastique. Je reviens plus bas sur les problèmes techniques que posent les réductions d'indéfinis à argument phrastique.

Les indéfinis prennent souvent des formes zéro lorsqu'ils sont utilisés en position objet mais pas en position sujet en français : *Jean a trop mangé (de choses/ Ø)*. *Quelque chose est tombée*. *Quelque chose nous a étonnés*. Dans certaines langues un argument sujet peut ne pas être explicite. En particulier un pronom référant au locuteur ou à son interlocuteur est généralement omis en war, austro-asiatique.

Les réductions à des formes zéro ne changent pas la structure argumentale des opérateurs, verbes ou noms, dans les exemples déjà vus. Des constructions causatives qui ajoutent un argument à un verbe peuvent aussi subir des réductions dont l'analyse fait intervenir une complexité accrue. Par exemple *manger* sans objet explicite n'est pas intransitif dans les exemples (54) et (55). Dans (54) *souris* est l'objet de *manger* tandis que dans (55) l'objet de *manger* est indéfini :

(54) *Jean a fait manger les souris par son chat.*

(55) *Jean a fait manger son chat avant nous.*

Les réductions à des formes zéro ne s'appliquent pas qu'aux arguments, elles peuvent aussi s'appliquer à un opérateur dont l'argument reste en place. C'est le cas par exemple des réductions métonymiques comme (53) qui serait dérivée de la co-assertion (57) où les deux occurrences de *quelque chose* sont identifiées :

(56) *Il a bu la bouteille.*

(57) *Il a bu quelque chose ; ce quelque chose est le contenu de la bouteille.*

De la même façon, *il dort toute la journée* ne fait pas intervenir une transitivation de *dormir* mais la réduction d'un opérateur lexical à valeur temporelle comme *pendant*. Ces exemples permettent de comprendre pourquoi dans (58), *manger* pourtant transitif ne s'applique pas à *nuît* mais à un indéfini sous-jacent avec un opérateur *pendant* réduit à une forme zéro, comme dans (57) :

(58) *Les chats mangent la nuit.*

(59) *Les chats mangent Ø, Ø, la nuit*

(60) *Les chats mangent des choses, pendant, la nuit.*

En revanche, avec un opérateur lexical transitif comme *chanter* qui accepte à la fois la sélection lexicale de *nuît* comme second argument et son interprétation temporelle c'est-à-dire une réduction métonymique temporelle (i.e. la réduction d'un opérateur comme *pendant*) et l'ellipse de son second argument, (61) est ambiguë contrairement à ce que prédiraient toutes les grammaires de constituants :

(61) *Luc chante la nuit.*

Le terme « vraisemblance » ou « probabilité » d'occurrence d'un argument lexical tel que l'utilise Harris (1982) n'est pas un terme technique permettant des calculs statistiques pour reconstruire la structure d'interprétation d'un énoncé, c'est-à-dire l'ordre d'insertion des opérateurs et arguments lexicaux explicites ou implicites, en dehors de corpus dans des univers de discours spécifiques (voir § 7). Dans un univers de discours où il serait question de transe et de démons capables de modifier l'écoulement, ou plus exactement la perception humaine du temps, on pourrait asserter *nuît* comme second argument de *manger* :

(62) *Les démons ont mangé la nuit.*

L'opérateur indéfini *quelque chose* à argument phrastique dans (63) a une instanciation finie. Son argument phrastique est l'opérateur *venir* opérant sur *Jean*, argument élémentaire. Dans certaines constructions, les ellipses grammaticales mettent en jeu des dépendances multilinéaires récursives non finies qu'on peut restreindre à des représentations récursivement énumérables par des points fixes sur

des indéfinis, voir Tarsky (1955). J'applique cette solution dans Daladier (1990) pour des énoncés contenant un indéfini *quelque chose* ou une ellipse comme dans (64) ou (65) portant sur une variable à valeur dans la classe infinie mais récursivement énumérable des énoncés. On notera que (65) n'est pas synonyme de : *Luc pense de travers*.

(63) *Luc pense quelque chose de nouveau qui est que Jean ne viendra pas.*

(64) *Luc pense quelque chose $\emptyset$ de nouveau.*

(65) *Luc pense qu'il est en train de penser $\emptyset$ de travers.*

Une catégorisation intuitionniste du lexique avec des variables de types doit donc distinguer non seulement mot élémentaire et mot opérateur mais également, ce que ne prévoyait pas Harris, opérateur à argument(s) élémentaire(s) et opérateur à argument(s) non élémentaire(s) (i.e. à valeur dans la classe récursivement énumérable des énoncés). Ceci s'applique à des éléments grammaticaux, voir § 6.2 et § 8 et à des éléments lexicaux notamment verbes et noms. Par exemple *erreur* requiert un argument non élémentaire qui peut être une complétive, un indéfini ou ne pas être explicite. Ceci permet de reconstruire grammaticalement la structure d'interprétation de $\emptyset$ dans le contexte de co-assertion de la construction (32), comme le ferait un locuteur, sans implicature pragmatique *ad hoc*.

(32) *Luc est rentré chez lui, une erreur $\emptyset$ qui aurait pu lui coûter cher.*

$\emptyset$ = *que Luc soit rentré chez lui.*

6.2 Questions sur la nature des combinaisons entre éléments lexicaux et éléments grammaticaux et le problème des catégories de description

Un demi-siècle après les principales publications de Quine et de Chomsky, les hypothèses de lois universelles de la pensée rationnelle supposées sous-jacentes au langage, telles qu'elles sont présentées dans Chomsky (1968), n'ont pu faire l'objet d'aucune vérification cognitive en neurosciences. Elles relèvent d'une idéologie scientiste, comme le montrent avec profondeur et humour Backer et Hacker (1984) dans leur pamphlet sur la logique et le langage au pays d'un roi nu et dans Backer et Hacker (1980) où ils analysent les raisons déjà données par Wittgenstein sur différents aspects qui distinguent fondamentalement la logique et le langage, notamment la non iconicité dans les langues, par exemple les noms de couleur. Plus généralement, les langues ne traduisent pas des conceptions d'une rationalité supposée universelle, ni même une phénoménologie de l'esprit avec son ontologie supposée également universelle. De telles conceptions se sont avérées de simples idéologies. Les langues dans toutes les cultures permettent d'exprimer des mensonges, des plaisanteries, des théologies et des cosmogonies qui diffèrent ainsi que des poèmes qui prolongent ou transforment le sens des mots et créent des images mentales, tout aussi bien que des théories scientifiques qui sont vérifiables, plutôt que vraies, en fait vérifiables à un certain moment c'est-à-dire vérifiables jusqu'à ce qu'une nouvelle théorie plus générale, n'émerge de plusieurs théories précédentes ou d'outils d'observation plus performants.

Brouwer et Heyting ont proposé de penser et d'intégrer la subjectivité dans une mathématique intuitionniste où les objets sont construits d'après la perception qu'en a le mathématicien, après avoir montré l'impossibilité d'ontologies prédéfinies des objets mathématiques. Dès les années 1920, Brouwer fonde une théorie de la topologie en explicitant la subjectivité de sa construction des objets mathématiques initiaux de sa théorie (voir aussi Stig (1994 : VII) qui publie la correspondance et carnets personnels de Brouwer). De nombreux physiciens ont explicité la subjectivité de leurs constructions des observations. Finalement, comme en algèbre et en topologie pour les mathématiques, Spinoza, Wittgenstein et Sapir ont montré, différemment, mais tous s'intéressant de façon concrète à la grammaire, l'impossibilité d'ontologies universelles en linguistique, en philosophie du langage et en anthropologie, en montrant la nécessité de penser la subjectivité et ses variations à l'intérieur de la description de leurs objets linguistiques, philosophiques et anthropologiques.

Ceci peut être précisé ici sur un exemple arithmétique simple. Au lieu de considérer les nombres comme des primitives ontologiques en mathématiques, Church (1941: 51-58 ; 60-62) représente leurs différentes classes (entiers, réels, ordinaux transfinis) à partir des opérations d'énumération en lambda calcul qui permettent de les construire. D'un point de vue historique et anthropologique, Menninger

(1934) montre que la notion de nombre cardinal en base dix ne peut certainement pas être utilisée comme primitive ontologique pour reconstruire les systèmes de nombres en linguistique comparative. Il montre qu'une autre notion de nombre qu'il appelle « grouping » est très courante dans de nombreuses civilisations anciennes écrites comme dans de nombreuses langues orales actuelles. Cette notion dépend de ce qui est compté et fait intervenir différentes bases de numération. Cette notion de « grouping », ou comme je le précise « d'unité de compte » dans Daladier (2015b) est une notion qui précède celle de nombre cardinal, qui demeure sous-jacente aux systèmes de nombres cardinaux actuels en austro-asiatique. J'analyse un système de nombres à plusieurs bases d'énumération existant encore dans les langues Pnaric-War-Lyngam (PWL) parallèlement aux cardinaux. Je montre que les noms de nombres et les bases de numération de ce système (PWL) permettent de dériver les noms d'un grand nombre de cardinaux austro-asiatiques d'un système commun de « groupings ». Ceci pose la question de la nécessité et de la reconstruction de notions culturelles pour la comparaison et la reconstruction linguistique. Là aussi, une iconicité au-delà des cultures et de la diachronie ne peut être qu'une vue de l'esprit.

Dans sa grammaire des opérateurs qui remplace la grammaire transformationnelle, Harris souhaite intégrer la métalangue dans la langue, c'est-à-dire se passer de catégories de description pour rendre sa grammaire indépendante d'hypothèses cognitives, comme les hypothèses d'innéité que devait faire Chomsky (1968) pour assurer la cohérence de sa théorie. En utilisant le minimum de catégories grammaticales, la grammaire des opérateurs de Harris ne dépend pas d'hypothèses cognitives ou logiques supposées être des lois préexistantes au langage et à la pensée (Daladier, 1990 : 58-59).

Après un long débat avec la grammaire dite générative, Maurice Gross propose en fin de carrière un lexique-grammaire de phrases quasi figées, qui se passe de catégories grammaticales puisque la sémantique grammaticale n'y joue plus de rôle significatif. Cette option est sans doute la plus minimaliste qu'on ait jamais conçue au sens technique d'un principe génératif élémentaire mais malheureusement aussi la plus minimaliste sur l'interprétation du sens grammatical et du changement linguistique dans les familles de langues.

Dans sa grammaire des opérateurs, Harris propose une grammaire qui tente d'éviter l'analyse des formes et de la sémantique des grammaticalisations d'éléments lexicaux. Sa grammaire est lexicalisée en faisant implicitement l'hypothèse qu'à un moment où s'est constituée une morphologie dans une langue, cette morphologie grammaticale porte une information équivalente à une information lexicale. Cette hypothèse hasardeuse comme on le verra plus bas, est intéressante par son côté provocateur, innovant une sorte de révolution dans la philosophie du langage et ses ontologies.

On peut se demander pourquoi les langues fabriquent et innovent de la morphologie grammaticale et pourquoi la morphologie change au cours de l'évolution. A l'heure actuelle, il existe probablement bien peu de langues à structure entièrement isolante, s'il en existe, en revanche une majorité de langues ont une morphologie avec affixes et flexions. Avant de revenir en détail sur ces questions §8 et §9, voyons comment Harris a imaginé pouvoir contourner la question de l'inconsistance d'une représentation strictement finie des opérateurs lexicaux.

7. Le codage de l'information dans un univers de discours spécifique peut se faire simplement et indépendamment de la représentation grammaticale des énoncés qui expriment cette information. La simplicité ou complexité de l'information est indépendante de la grammaire de cette langue.

Harris, son frère et sa belle-sœur avaient défini le codage informationnel d'un corpus de textes en anglais posant une question en biologie et la résolvant. Harris et al. (1989) a montré que l'information de ce gros corpus d'articles en anglais et en français pouvait s'analyser à partir d'une grammaire informationnelle finie basée sur des structures très simples de mots clefs opérateurs et d'arguments spécifiques, ayant des emplois lexicaux bien définis. Je montre dans mon chapitre de Harris et al. (1989) que ce codage est indépendant des langues dans lesquelles sont rédigés les articles de biologie du corpus, et de leurs grammaires, parce que les mots clefs et leurs restrictions de sélection lexicale sont identiques dans cet univers de discours spécialisé. Précisément, un univers de discours spécialisé se caractérise par l'emploi particulier d'un lexique spécialisé. La conséquence était un résultat inattendu pour Harris et pour les auteurs ayant rédigé des chapitres sur l'interface structure

morpho-syntaxique/ structure informationnelle. Je montre dans mon chapitre que l'analyse de la structure morpho-syntaxique anglaise ou française comme préliminaire à la traduction de la structure informationnelle, réalisée par Harris, Gottfried et Ryckman était inutile. On peut utiliser directement le codage de l'information pertinente comme « pattern matching » sur les énoncés pour obtenir directement l'information pertinente qu'ils véhiculent, sans traiter les particularités morphosyntaxiques de la formulation elle-même, ni traduire les textes du français à l'anglais. Le terme technique de « pattern matching » et la méthode empirique utilisée sont décrits dans mon chapitre. Lorsque Harris a reçu mon chapitre, il m'a répondu que ma conclusion lui paraissait essentielle par une lettre enthousiaste datée du 18 mai 1983.

Depuis, les moteurs de recherche de Google ont fait universellement fortune sur le même principe de structures sémantiques simples définies sur de grosses bases de données permettant de court-circuiter l'analyse grammaticale des énoncés et leurs traductions. Il est inutile à une machine de comprendre un énoncé dans une langue pour comprendre son sens, il suffit de combiner le sens de mots clefs prédéfinis pertinents à un univers de discours. La simplicité ou complexité de l'information dans un univers de discours est donc indépendante de la complexité de la grammaire d'une langue la véhiculant.

Cette conclusion a été reprise par Harris dans son cycle de conférences à Columbia en 1986 où il tentait de délimiter une nouvelle philosophie de l'esprit, détaillée dans Harris (1991), allant d'une structure universelle des langues fondée sur leurs éléments lexicaux, à une formulation scientifique de l'information : « La complexité notoire de la grammaire dont la majeure partie est créée par les réductions, n'est pas due à la complexité de l'information et n'est pas nécessaire à l'information » (Harris 1988, trad. 2007 : 45). Harris distingue donc le fonctionnement d'une langue du fonctionnement des objets qu'elle permet d'exprimer et conclut que sa grammaire des opérateurs pour décrire les langues peut se passer de catégories métalinguistiques comme peut s'en passer sa grammaire des opérateurs pour décrire la structure de l'information dans des textes scientifiques. La raison en serait pour lui que les lois du langage et de la pensée seraient auto-organisées et non créées par un agent extérieur (Harris 1988 : 113 et Harris 1990 : 19). Cette conclusion se révèle purement idéologique, l'existence ou non d'un Créateur est indépendante des structures linguistiques, comme elle l'est des structures de langages mathématiques. Les structures informatives simples d'un corpus n'impliquent en rien les grammaires des langues utilisées pour formuler cette information. Sa conception trop simple des structures lexicales dans le langage ordinaire s'avère inconsistante et le fait que l'information puisse être codée de façon simple dans un univers de discours spécialisé ne fournit pas de réponse à cette question. Du reste, l'information ne peut pas toujours être codée de façon simple, y compris en mathématiques. L'utilisation par tout un chacun de nombres fractionnaires suffit à montrer que nous pouvons aussi utiliser des structures informatives non finies dans le langage ordinaire.

Le paradoxe du menteur de Kripke a eu un effet dévastateur en philosophie du langage : l'énoncé : *Je mens*, n'est ni vrai ni faux comme proposition logique bien qu'il ait une valeur informative simple comme énoncé linguistique. Différentes sortes de logiques modales ou intentionnelles ont tenté de contourner ce paradoxe. Un demi-siècle après les principales publications de Kripke, de Quine et de Chomsky et après les moteurs de recherche de Google, les hypothèses de lois universelles d'une pensée logique qui seraient sous-jacentes au langage, telles qu'elles sont présentées dans Chomsky (1968), n'ont pu faire l'objet d'aucune vérification cognitive en neurosciences. Elles relèvent d'une idéologie scientiste, comme le montrent avec humour Baker et Hacker (1984) (cf. 6.2) Les langues permettent d'exprimer des mensonges, des plaisanteries, des théologies, des cosmogonies, des poèmes, qui parfois transforment le sens des mots en leur associant des images mentales, tout aussi bien que des théories scientifiques. Le fonctionnement linguistique ne dépend ni du contenu de l'information dans des univers de discours spécialisés, ni d'une rationalité logicienne universelle.

Finalement, il reste à découvrir le fonctionnement linguistique à travers la diversité des langues et de l'évolution dans les familles de langues.

8. De la représentation des éléments lexicaux et grammaticaux par des opérateurs sans catégories prédéfinies à une typologie inductive pour l'analyse du sens grammatical dans différentes langues et au cours de l'évolution de familles de langues

Les théories de Harris sont limitées par deux grandes questions pour la description des structures d'interprétation des langues. La première relève de l'inconsistance de structures lexicales finies et de façon liée de la nécessité de rendre compte du caractère variable de la portée des opérateurs lexicaux selon la construction dans laquelle ils sont insérés. La seconde question réside dans l'absence de prise en compte des sémantiques grammaticales mises en place dans les différentes familles de langues, souvent à travers des formes différentes de morphologie. Les différentes formes grammaticales sont également intervenir des opérateurs à portée variable selon le contexte. Ces opérateurs ont des domaines de restriction de sélection à la fois sur des éléments lexicaux et sur d'autres éléments grammaticaux, comme je le montre plus bas.

Nous connaissons bien les langues à morphologie flexionnelle, avec morphologie affixale, distinguant verbes, noms, noms simples et noms d'actions, adjectifs etc. et toutes sortes de grammaticalisations. Les langues à morphologie isolante n'ont pas de morphologie spécifiant des parties du discours. Cependant l'usage nominal ou verbal de mots s'exprime par leurs propriétés de construction avec différentes particules.

Il est déjà visible dans l'état actuel des connaissances en typologie qu'il n'y a pas de sémantique grammaticale universelle mais des continuum et variations sur certains thèmes étudiés comme le genre, l'indéfini ou la focalisation d'argument(s) selon les marquages d'accord verbal ou d'autres formes discursives plus libres de focalisation. De même, le découpage lexical a des variations intéressantes à la fois dans les différentes langues et au cours du changement diachronique.

Harris ne décrit pas la structure argumentale ni la sémantique d'éléments non lexicaux, pour des éléments grammaticaux ou grammaticalisés. Pourtant ces structures ne se ramènent pas à des prédictions lexicales ou à des actes performatifs sous-jacents.

La prise en compte des structures argumentales de particules, de flexions ou d'éléments grammaticalisés c'est-à-dire la prise en compte de leurs restrictions de sélection change la nature des théories de Harris et de façon plus générale la nature des théories linguistiques dites formelles. Par exemple l'auxiliation des verbes et celle des noms avec des verbes dits « supports » ou *light verbs*, la voix passive ou la voix moyenne, dans différentes langues, ne peuvent pas se ramener à des réductions sous-jacentes d'éléments lexicaux, ni à des éléments indépendants du contexte parce qu'elles ont leurs contraintes de sélection propres. Ce n'est d'ailleurs pas par accident que les langues produisent des grammaticalisations différentes de verbes comme auxiliaires, de déictiques, de pronoms, de particules, d'affixes et éventuellement de flexions.

L'anglais comme le français et beaucoup de langues de familles différentes, ont des auxiliations distinctes, verbales et non verbales, avec des verbes grammaticalisés. En particulier les constructions dites à verbes supports avec des verbes grammaticalisés qui se construisent avec des noms, sont en réalité de véritables systèmes d'auxiliation. Dans des langues à morphologie flexionnelle comme le français ou l'anglais, ces systèmes d'auxiliation sont complémentaires des auxiliations verbales en ce qu'elles grammaticalisent toutes sortes de valeurs de subjectivité et d'aspect, voire de combinaisons des deux. Des énoncés comme : *J'ai jeté un regard sur la théorie* ou *j'ai porté mon regard sur la théorie* ne sont pas synonymes et ne se bornent pas à porter une morphologie d'ancrage et l'accord sujet-verbe. Ils apportent aussi une information subjective.

La notion d'auxiliation nominale avec des verbes grammaticalisés se construisant avec des noms est précisée dans Daladier (1999). Cette auxiliation nominale est particulièrement développée en sanskrit Brahmanique, avec différents verbes notamment un verbe signifiant *faire*, avant de se lier en un système de temps-aspect verbal proprement dit en sanskrit classique, comme le montre Whitney (1924:386-391). Ceci est à mettre en relation avec l'existence d'une notion de phrase nominale en sanskrit védique (infra). Je montre dans Daladier (2012) pour le *war* qui n'a pas de distinction morphologique entre nom et verbe, ni voix, mais des particules assertives qui se construisent avec des opérateurs lexicaux, qu'il existe un riche système d'auxiliation utilisant plus de 25 verbes grammaticalisés. Ce système d'auxiliation apporte des valeurs aspectuelles et subjectives aux énoncés, qui s'apparente à la fois aux auxiliations verbales et nominales de langues comme les langues romanes et germaniques. En *war*, ces auxiliaires ne convoient aucune morphologie d'ancrage et ne sont donc

pas des verbes supports. Les particules d'ancrage opèrent à la fois sur les éléments lexicaux et sur les autres éléments grammaticaux assertifs : auxiliaires, préfixe d'action réciproque ou marquage différentiel focalisant un ou des arguments, notamment. Le système d'auxiliation en war associe des valeurs d'aspect et de subjectivité référant au locuteur ou à un argument agentif ainsi que pour certains des changements de valence. Il associe des valeurs plus ou moins actives ou passives sans système de voix. Finalement, bien que morphologiquement très différent, le système flexionnel verbal du français, avec ses systèmes de voix et d'auxiliation, et ses constructions moyennes ou réflexives peut être comparé sémantiquement au système d'opérateurs du système assertif du war où éléments d'ancrage et auxiliaires sont indépendants. Les éléments opérant dans le système grammatical de chaque langue ne peuvent être assimilés ou intégrés à une notion simple de prédication lexicale.

En austro-asiatique, le vieux mon et le khmer préangkorien, langues écrites aux 8^{ème} et 9^{ème} siècles sont des langues avec des bases verbales mais Jenner et Pou (1981) montrent pour le khmer que les affixes s'appliquaient, s'appliquent toujours, à des fonctions à la fois nominales et verbales, avec des valeurs de subjectivité extrêmement riches qu'ils renoncent à lier pour les fonctions d'ancrage à strictement de l'aspect même modalisé. Ces valeurs d'ancrage référant à la subjectivité du locuteur et/ou de son interlocuteur semblent se retrouver dans différentes familles de langues sous différentes formes. Un des traits typologiques particulièrement intéressants du war est la façon dont s'exprime l'ancrage des énoncés à l'intérieur d'un système structuré de niveaux de formes grammaticales imposant des contraintes de sélection sur d'autres particules de niveau inférieur et sur des éléments lexicaux, avec de très riches valeurs. Au sommet de cette hiérarchie, le war a des particules qui situent causalement et temporellement des énoncés dans un discours (récit, explication etc.) ou un dialogue et qui constituent aussi des particules d'ancrage. Pour des énoncés plus simples, le war a des particules positives et négatives dont les fonctions assertives (déclaratif, interrogatif, impératifs, exclamatifs) incluent les fonctions aspectuelles d'ancrage. Ces particules peuvent être considérées comme jouant en partie la fonction verbale de l'énoncé en lui associant une valeur qui peut être aspectuelle ou renvoyer à la subjectivité du locuteur ou de son agent ou encore à celle de son interlocuteur et qui de plus peut être positive ou négative. En revanche, il n'y a aucune forme d'accord entre l'opérateur lexical utilisé verbalement et son ou ses arguments. Le dernier niveau de ce système de dépendance assertive comporte deux particules de marquage différentiel des arguments de l'opérateur lexical utilisé en fonction verbale. Ce marquage focalise les arguments pour certains rôles sémantiques de façon voisine de ce que font les formes d'accord verbe-arguments pour des langues munda qui ont développé un système verbal. Cependant le war comme le pnar et le lyngam n'ont aucune tendance à développer un système verbal et le khasi est la seule langue du groupe, que je définis comme pnaric-war-lyngam, voir Daladier (2011), à avoir développé un système verbal. En khasi, un clitique référant au sujet et auquel sont affixés des marques d'ancrage précède l'opérateur lexical qui se comporte donc comme un verbe. En addition à son système verbal, le khasi a conservé le marquage différentiel des arguments du pnar.

Dans les langues asliniennes, temiar et semlai et dans les langues munda du sud : kharia, gorum et bonda, on observe un système vestigiel de marquage différentiel des arguments pour des rôles agentif ou bénéfactifs à l'intérieur de systèmes verbaux assez différents. Du point de vue de l'évolution des langues austro-asiatiques, le système de dépendance assertive du war, du pnar et du lyngam semble être très conservateur.

En war, les énoncés comportent généralement des particules d'ancrage des énoncés, qui en plus de leurs valeurs spécifiques, ont une valeur de prise en charge de l'énoncé par le locuteur ou son agent. Lorsque l'énoncé a une valeur purement factuelle, il n'y a pas de particule d'ancrage. Ces énoncés ne comportent aucun élément qui puisse être interprété comme ayant une fonction verbale et en particulier le marquage différentiel des arguments ne s'y applique pas. L'équivalent en war de *Lum est le fils de Maia*, ne comporte aucune particule d'ancrage, aucune copule et aucune forme d'accord focalisant *Lum*. L'énoncé war se distingue par l'absence d'accord sujet-verbe de l'énoncé en français, sa copule et sa flexion de présent. Le war a une seule forme, exprime exactement de la même façon donc, ce que le français exprimerait sous les deux formes différentes : *le fils de Maia*, *Lum* et *Lum est le fils de Maia*. En revanche, si le locuteur dénie ou corrige cette information, le war énonce l'information avec une particule d'affirmation ou une particule d'aspect potentiel négatif, ce qu'on pourrait traduire par : *(ce n'est) pas le cas (que) Lum (soit) le fils de Maia* où les mots entre parenthèses ne sont pas exprimés en war. La situation serait analogue en pnar et en lyngam.

De façon très intéressante, l'existence de formes différentes d'énoncés selon que le locuteur prend en charge ou non l'information en pnaric-war-lyngam, groupe conservateur dans la genèse des systèmes verbaux austro-asiatiques, peut être comparée avec l'évolution des langues sémitiques et avec celle du sanskrit en indo-européen. Cohen (1984 : 591) conclut son ouvrage sur l'évolution du système verbal en sémitique sur le fait que les énoncés « nominaux » expriment une information factuelle et ne focalisent pas l'argument agentif alors que les énoncés verbaux à copule correspondants focalisent le sujet. Renou (1952 : 356-357) montre que le sanskrit védique avait développé deux types d'énoncés, énoncés verbaux et énoncés nominaux avec un adjectif apposé à un nom. Ces énoncés nominaux expriment des généralités par opposition aux énoncés verbaux qui expriment des valeurs temporelles, modales ou subjectives référant au locuteur ou à son sujet.

Les langues chamito-sémitiques, indo-européennes et austro-asiatiques ne fondent pas les énoncés uniquement sur la fonction verbale et la notion d'énoncé se modifie diachroniquement, en partie différemment. Je montre dans Daladier (2015a) qu'en war le marquage différentiel des arguments les focalise pour des rôles sémantiques d'agent non attendu et de bénéficiaire particulier qui se substituent aux fonctions d'accord d'un verbe avec un ou plusieurs arguments dans d'autres langues austroasiatiques, comme certaines langues munda.

Le pnar, le lyngam et le war ont chacun développé un système grammatical de dépendance assertive avec certaines particules et affixes et certaines caractéristiques sémantiques propres. Pour ces trois langues, les principaux aspects de la prédication assertive se marquent morphologiquement indépendamment de l'opérateur lexical, seuls les préfixes causatifs et le préfixe d'action collective sont associés à l'opérateur lexical dans ces langues. Le marquage différentiel en war focalise soit le caractère inattendu d'un agent pour le locuteur, soit un bénéfice particulier ou détriment malencontreux pour un acteur avec l'empathie du locuteur ou de son agent. On retrouve dans des langues austro-asiatiques ayant développé un système verbal et qui ne sont plus en contact, la focalisation d'arguments pour des rôles sémantiques voisins, agentifs ou bénéficiaires particuliers, soit dans l'utilisation de clitiques dans la base verbale, soit dans un marquage différentiel des arguments. On trouve en particulier un marquage différentiel des arguments qui pourrait être un vestige d'un système pré-verbal austro-asiatique proche du war, dans certaines langues munda et dans les langues asliennes, voir (Daladier 2015a). Des arguments à valeur de bénéficiaire et d'agent sont aussi marqués dans des langues austronésiennes, voir Ross (2002) et Blust (2002) qui y voient un système non conventionnel de voix.

Des questions sur les catégories grammaticales pertinentes et leur évolution ont été également posées pour les langues tibéto-birmanes. Thurgood (1985) et LaPolla (1993) ont mis en évidence l'existence de systèmes pré-verbaux. LaPolla comme Li et Thomson, considèrent que certaines langues n'ont pas de syntaxe, au sens de structures syntagmatiques, mais une organisation pragmatique. Certains linguistes actuels insatisfaits à juste titre des syntaxes syntagmatiques, revendiquent l'intérêt de descriptions pragmatiques qui permettraient de rendre compte de leurs « analyses informationnelles ». Cependant, ils ne rendent pas compte de la façon dont se grammaticalisent et se structurent leurs analyses informationnelles.

Dans une perspective où on ne réduit pas la syntaxe à une structure simple syntagmatique et à ses parties du discours, la question se pose très différemment puisqu'on peut considérer comme ici une grammaire qui décrit les structures d'interprétation de formes grammaticales contraintes lexicalement et grammaticalement tout en ayant parfois des valeurs subjectives. Ces formes grammaticales contraintes s'opposent à des formes discursives non contraintes de focalisation par exemple. Ainsi la focalisation syntaxique d'un argument marqué ou d'un argument qui s'accorde avec le verbe à la voix active ou passive s'oppose à la focalisation par permutation en tête de phrase, en français : *On a volé Luc. / Luc a été volé. / C'est Luc qu'on a volé.* Harris ne rend pas compte et ne peut rendre compte dans ses grammaires des variations de l'ordre des mots pour leurs variations de valeurs discursives.

Il semble intéressant de construire des grammaires induites par les données, comportant plusieurs niveaux structurels qui intègrent les structures d'interprétation contraintes et les formes grammaticales optionnelles ainsi que des formes grammaticales de variations discursives. Cette perspective n'oppose ni syntaxe et sémantique, ni syntaxe et pragmatique mais considère au contraire un continuum dans les structures d'interprétation des énoncés avec une gradation des contraintes des éléments lexicaux et grammaticaux. Elle permet en outre de situer l'observation grammaticale à un

niveau à la fois suffisamment détaillé et suffisamment abstrait pour observer des régularités sous-jacentes à l'évolution diachronique et à la diversité des langues.

L'approche inductive de la typologie et de ses catégories pertinentes permet de faire émerger des processus internes de grammaticalisation. Elle permet aussi de reconstruire une histoire et une géographie aux langues orales des diverses familles permettant de séparer les phénomènes de contacts de langues avec calques des phénomènes d'évolution interne. Elle pourrait permettre de faire apparaître les points communs du sens grammaticalisé et aussi les facteurs de divergence au-delà des variations dans les familles de langues.

De son côté, la typologie traditionnelle qui fait apparaître des pans entiers de sémantiques grammaticales à l'occasion de descriptions de langues non encore connues, pêche souvent par la non pertinence de ses catégories, généralement issues de la morphologie des langues les mieux étudiées, qu'elle tente d'étendre. Ce faisant elle manque parfois certaines généralisations. L'analyse dans différentes familles de langues, ou d'états diachroniques de langues, de thèmes typologiques ou de traits (ordre des mots) ou encore de types de morphologie, est certainement nécessaire mais certainement aussi non suffisante. L'interdépendance de différents aspects d'une grammaire sont aussi à prendre en compte. Par exemple certaines valeurs liées à l'expression d'une voix verbale peuvent se traduire dans d'autres langues comme le *war* sans opposition *verbo-nominale* par un marquage différentiel des arguments et une auxiliation d'opérateurs lexicaux.

L'analyse du statut d'opérateur de particules ou d'éléments grammaticalisés selon les constructions, tout comme l'analyse du statut d'opérateur des éléments lexicaux selon leurs constructions permet donc une approche inductive de la catégorisation pour la typologie des langues. Elle permet une description homogène des auxiliations de verbes et de noms du *sanskrit* au français. Elle permet aussi l'intégration de valeurs pragmatiques dans la morpho-syntaxe de langues austro-asiatiques (voir Daladier 2012 et 2015a).

La complexité des langues ne résulte pas seulement des dépendances phrastiques, celles-ci ne sont en réalité qu'un petit aspect de phénomènes beaucoup plus généraux de réseaux de dépendances : dépendances entre éléments lexicaux opérateurs, dépendances entre des éléments opérateurs grammaticaux entre eux et dépendances de dépendances entre opérateurs grammaticaux et lexicaux.

Les structures d'interprétation (i.e. hiérarchie et dépendances des calculs à effectuer) des langages de programmation sont depuis plus de trente-cinq ans construits avec des modèles non finitistes. La garantie de terminaison des calculs, même s'ils sont effectués en parallèle et non séquentiellement pour optimiser les temps de calcul, est assurée par des mécanismes pragmatiques ajoutés à ces structures. Ceci devrait poser la question de façon plus radicale et surtout de façon plus constructive de la pertinence des grammaires de constituants finitistes pour analyser la structure d'interprétation des langues. Puisque les langages de programmation nécessitent des grammaires non finitistes pour exécuter les calculs en parallèle, comment se pourrait-il que les analyses et descriptions de langues puissent se faire dans des grammaires à structures strictement finitistes ?

D'un point de vue empirique, la typologie fait de plus en plus apparaître le caractère arbitraire des catégories prédéfinies, comme le montre en particulier Haspelmath (1997), (2001), (2007).

Dans la perspective esquissée ici où on évite les catégories prédéfinies, une grammaire est un continuum de contraintes grammaticales interagissant entre elles et avec les éléments lexicaux. L'analyse grammaticale d'un énoncé consiste à hiérarchiser ses éléments grammaticaux et lexicaux explicites et implicites pour en produire la structure d'interprétation. Les éléments grammaticaux font intervenir différentes formes de morphologie (souvent combinées) selon les langues, particules isolantes, affixes, flexions et grammaticalisations. Dans la diversité de ces formes, elles incluent différentes valeurs de subjectivité qu'il ne faut pas confondre avec une pragmatique purement discursive peu ou pas contrainte grammaticalement et non contrainte lexicalement.

Du point de vue de l'élaboration d'une grammaire incluant plusieurs types fondamentaux d'ellipse grammaticale, la grammaire des opérateurs de Harris 1982 était très en avance sur les autres approches même si elle était insuffisante. En revanche elle a complètement méconnu le phénomène des grammaticalisations d'éléments lexicaux avec la sémantique des grammaticalisations, des flexions, des affixes et des particules tout autant structurée que celle des opérateurs lexicaux. On peut repenser la représentation de ces phénomènes de grammaticalisation avec les organisations morphologiques dans lesquelles ils s'insèrent à l'intérieur d'une représentation applicative globale comportant à la fois les structures argumentales des éléments lexicaux en contexte et celles des éléments grammaticaux en

contexte. Les éléments lexicaux et grammaticaux interagissent, les contraintes distributionnelles sont aussi bien lexicales que grammaticales, comme on le voit pour le passif.

Ces structures argumentales font donc intervenir des contraintes à plusieurs termes sur des éléments grammaticaux et sur des éléments lexicaux. Elles peuvent être représentées par des structures de treillis et non par les structures linéaires non contextuelles des grammaires de classes de mots. Les relations de dépendance des opérateurs grammaticalisés, ou particules grammaticales opérateurs dans un sens plus général, sont typiquement des relations de dépendances à plusieurs termes et non linéaires comme le sont les arbres simples des grammaires de constituants syntagmatiques. Ce point me semble fondamental pour comprendre différents aspects empiriques, spécifiques aux langues ou non, de la complexité grammaticale. Cette complexité s'associe aux propriétés d'ellipse et de portée variable selon le contexte qu'on trouve sans doute dans toutes les langues. L'inconsistance de la théorie des opérateurs de Harris est de ce point de vue un résultat négatif très intéressant.

Les questions issues de l'extension proposée ici de la grammaire des opérateurs de Harris, permettent de repenser la notion de construction grammaticale ainsi que son évolution diachronique, différente dans différentes familles de langues. Le but serait de reformuler de façon inductive la question des universaux typologiques à l'intérieur d'organisations grammaticales holistiques plutôt que comme listes de thèmes indépendants tels que : ordre des mots, voix, auxiliation, marquage différentiel des arguments, expression de l'indéfini etc...

L'analyse diachronique et synchronique d'un groupe de langues austro-asiatiques, le groupe *pnaric-war-lyngam*, doit être envisagée dans une perspective plus ouverte que ce que proposent les typologies actuelles pour prendre en compte des dépendances à plusieurs termes de particules grammaticales entre elles et sur des éléments lexicaux et également pour rendre compte d'un continuum de valeurs grammaticales qui vont de certaines fonctions « verbales » à des valeurs traditionnellement considérées comme pragmatiques mais qui sont tout aussi contraintes.

Les conceptions de Sapir (1949), Spinoza (1677¹ 1968²), Wittgenstein (1974²), Backer and Hacker (1980), Kripke (1982), Brouwer (1975) et Heyting (1955), chacun pour une raison propre à sa pensée mais toutes de façons complémentaires, ont mis en évidence l'impossibilité d'une ontologie universelle, soit pour le langage, soit pour la philosophie du langage et l'anthropologie, soit pour les mathématiques. Ces différentes approches mettent au centre de leurs méthodologies la subjectivité et en particulier la nécessité d'une construction de la représentation des objets d'observation, toute « réalité » préalable à cette construction ne pouvant être que dogmatique.

Conclusion

La tentative de Harris d'analyser une langue sans catégories, l'intégration de la métalangue dans la description d'une langue, s'est révélée être une utopie, intéressante pour plusieurs raisons. La première raison tient à la simplicité arbitraire à la fois de sa conception de structures lexicales indépendantes du contexte et de sa conception des ellipses grammaticales. Les ellipses grammaticales concernent en particulier des phénomènes référentiels de mots simples mais aussi d'éléments phrastiques avec ou sans pronom explicite et les ellipses d'éléments indéfinis, élémentaires mais aussi phrastiques. Bien que devant être profondément redéfinie comme on l'a vu ici, la notion d'ellipse grammaticale s'avère cependant très intéressante et novatrice car largement sous-estimée dans les autres approches.

Un autre aspect de l'insuffisance de la grammaire des opérateurs de Harris (1982) est que l'évolution des langues avec leurs différents processus de grammaticalisation, leurs sémantiques grammaticales et leurs morphologies, n'a rien d'un épiphénomène mineur. Ces processus de grammaticalisation jouent un rôle très important dans la sémantique des langues à côté de la sémantique des éléments lexicaux et de leurs structures contextuelles d'interprétation. On en a un aperçu dans la description du sanskrit par les grammairiens et philosophes Panini et Patanjali en fonction de la morphologie et de la sémantique grammaticale du sanskrit, dans Renou (1957). Pour des langues non indo-européennes, la différence peut être beaucoup plus importante encore, voir par exemple Jenner et Pou (1981). La tentative de Harris (1982) de gloser les formes grammaticales par des éléments lexicaux s'est révélée incohérente même en anglais, les paraphrases étant approximatives et contextuelles et cette tentative n'a pas eu de suite.

De cette utopie, il demeure l'intérêt de représenter des éléments lexicaux sans classes de mots prédéfinies, par des opérateurs, à condition d'étendre cette notion d'opérateur à des variables d'opérateurs pour rendre compte de structures contextuelles d'interprétation. De plus, la représentation par des opérateurs doit être étendue aux formes grammaticales. Là aussi des structures d'opérateurs sans catégories prédéfinies sont nécessaires pour comprendre les relations de dépendance entre éléments grammaticaux entre eux et avec certains éléments lexicaux. En particulier, les morphologies d'ancrage varient profondément selon les langues, par leurs formes : flexions, affixes, clitiques dans une base verbale, particules isolantes non associées à l'élément à fonction verbale mais surtout par leurs sémantiques et leurs contraintes avec les éléments de systèmes différents d'auxiliation, éventuellement de voix ou même de négation assertive comme dans le groupe pnaric-war-lyngam. Pas plus que la sémantique lexicale, la sémantique grammaticale n'est iconique dans les langues. Cependant, les parentés structurales et sémantiques qu'on peut observer à travers des systèmes grammaticaux différents sont sans doute d'autant plus intéressantes pour comprendre les mécanismes communs sous-jacents au langage. Le système d'ancrage du war n'est pas lié à une opposition de fonctions nominales et verbales contrairement à ce qu'on observe dans la plupart des langues actuelles. Certaines de ses valeurs sont très différentes de celles qu'on trouve dans le système d'ancrage d'une langue comme le français. Cependant les valeurs du système d'auxiliation du war contiennent en bonne partie celles des deux systèmes d'auxiliation qu'on trouve dans une langue comme le français : le système d'auxiliation des voix verbales actives et passives et le système d'auxiliation nominale dit « à verbes supports », voir Daladier (1999 et 2012). A ceci, on peut associer l'existence, dans des familles de langues différentes comme le sémitique et l'indo-européen, de formes d'énoncés soit verbales avec flexion d'ancrage, soit nominales sans flexion d'ancrage, qu'on peut comparer aux deux formes d'énoncés en war sans cette opposition mais avec ou sans particule d'ancrage. Ces systèmes morphologiquement très différents ont donc en commun d'avoir des énoncés sans prédication d'ancrage pour exprimer une information factuelle (non prise en charge par le locuteur).

Finalement, la reconstruction des structures d'interprétation dans une grammaire qui minimise la métalangue pour éviter les catégories prédéfinies mal adaptées doit faire intervenir divers processus de réduction de façon beaucoup plus importante que ne l'avait envisagé Harris, à la fois quantitativement et qualitativement. Certaines demeurent indéfinies ce qui me semble une propriété très intéressante. Aucune des conceptions grammaticales actuelles : syntaxes formelles ou grammaires fonctionnelles typologiques, ne rendent compte de façon générale ni des structures d'interprétation variables contextuellement d'éléments lexicaux non élémentaires, ni d'ellipses référentielles sans pronom explicite, faute de décrire les phénomènes d'ellipse d'indéfinis et la portée variable d'éléments lexicaux et grammaticaux, qu'on retrouve semble-t-il dans toutes les familles de langues et à différents stades d'évolution.

L'approche post-harrissienne, mais toujours sapirienne, que j'ai rapidement esquissée ici sans entrer dans les aspects techniques, a pour but de mieux comprendre de grandes phases d'évolution du sens grammatical et leurs structurations successives dans des familles de langues différentes. Ainsi, la plupart des langues austro-asiatiques ont développé des systèmes verbaux sans temps, avec aspects et différents codages de la subjectivité du locuteur et de son interlocuteur. Quelques-unes sont restées essentiellement isolantes, la plupart ont évolué en langues à structures dérivationnelles (avec affixes). Le war, comme le pnar et le lyngam, a continué à développer sa structure surtout isolante et quelques affixes, avec des grammaticalisations de valeurs dites pragmatiques de plus en plus riches, à l'intérieur même de son système grammatical de dépendance assertive. Bien que le war n'ait pas de complétives, ni de relatives, ni d'infinitives, ni d'accord sujet-verbe, ni la moindre forme de conjugaison ou de base verbale, du point de vue de l'économie des structures d'interprétation, sa grammaire n'est pas plus « primitive », ni plus simple dans un sens technique de structures récursives, que des grammaires flexionnelles comme celles des langues romanes. De nombreuses fonctions communes s'y trouvent représentées autrement et certaines fonctions lui sont propres avec leurs grammaticalisations spécifiques. Les ellipses grammaticales sont particulièrement riches et la sémantique grammaticale tout aussi riche.

Au lieu de considérer que les langues se complexifient au cours des siècles (ce qui s'avère inexact dans un sens technique), on peut voir les changements diachroniques plutôt comme des sortes de restructurations morphologiques utilisant éventuellement différentes formes de réductions et faisant intervenir des grammaticalisations, lexicalisations et figements morphologiques. Ces restructurations

se font sur des structures grammaticales non linéaires, non nécessairement finies, faisant intervenir des treillis de dépendance à plusieurs termes entre éléments lexicaux et éléments grammaticaux.

Ces dépendances complexes d'opérateurs lexicaux et grammaticaux, explicites ou non, qu'on observe dans des langues aussi différentes que le français et le war, peuvent se représenter par des structures d'opérateurs assez voisines. Ces structures sont adaptées à des comparaisons de grammaticalisations apparemment très différentes aussi bien en synchronie qu'en diachronie.

Références

- Backer, Gordon P., Hacker, Peter M. S. 1980, *Wittgenstein: Understanding and Meaning*, Oxford/Blackwell 1984, *Language, Sense and Nonsense*, London: Blackwell
- Bisang, Walter, 2008, "Underspecification and the noun/verb distinction: Late Archaic Chinese and Khmer." In A. Steube, Ed. *The discourse potential of underspecified structures*. Berlin, Mouton de Gruyter. 55–83.
- Blust, Ronny, 2002, "Notes on the history of 'Focus' in Austronesian Languages." In F. Wouk and M. Ross, Eds. *The history and typology of western Austronesian voice systems*. Canberra, Pacific Linguistics. 63–78.
- Brouwer, Luitzen E. J., 1975. Collected works éd. by A. Heyting, North Holland
- Chomsky, Noam., 1968, *Language and mind, trad. française : Le langage et la pensée*, 1970, Payot
- Cohen, David. 1984, *La phrase nominale et l'évolution du système verbal en sémitique*, Société de Linguistique de Paris LXXII
- Croft, William, 2001, *Radical Construction Grammar*, Oxford University Press
- Daladier, Anne, 1986, *Deriving the Literal Interpretation Structures of a Natural Language in a Second Order Lambda Calculus*, 50p. Rapport Technique LADL, CNRS
- 1990, Une représentation applicative des énoncés et de leur dérivation, *Les grammaires de Harris et leurs questions*, A. Daladier ed., *Langages* 99
- 1999, Auxiliation des noms d'action, *Les auxiliaires*, *Langages* 135 : 87-108
- 2001, Détermination, relativation, complétivation : dépendances variables d'éléments pronominaux sous des paramètres d'assertion, *Actes du XXIIème Congrès de linguistique et philologie romane*, U.L.B. Bruxelles
- 2011, The group Pnaric-War-Lyngngam and Khasi as a branch of Pnaric, *Journal of South East Asian Linguistics Society* 4-2, 169-206
- 2012, Graded Passive and Active values in Serial Constructions in Kudeng War in: Hyslop, Morey, Post Eds., *North East Indian Linguistics* 4, chap. 18, Cambridge Press India
- 2015a, Differential marking of arguments and the grammaticalisation of subjectivity values in War, chap. 6, Linda Konnerth, Stephen Morey, Priyankoo Sarmah, Amos Teo *North East Indian Linguistics*, 7. Canberra, Australian National University: Asia-Pacific Linguistics Open Access
- 2015b, The Counting Unit system of Pnar, War, Khasi, Lyngngam and its traces in Austroasiatic Composite Cardinal Systems, chap. 12, in Linda Konnerth, Stephen Morey, Priyankoo Sarmah, Amos Teo eds., *North East Indian Linguistics* 7, Canberra, Australian National University: Asia-Pacific Linguistics Open Access
- Harris Zellig S., 1982, *A Grammar of English on Mathematical Principles*, New-York: Wiley-Interscience
- 1988, *Language and Information*, Bampton lectures, Columbia U. Press (traduction française : 2007, Paris, CRL, par Amr H.Ibrahim et Claire Martinot)
- 1990, La genèse de l'analyse des transformations et de la métalangue, *Les grammaires de Harris et leurs questions*, A. Daladier ed., *Langages* 99, 9-21
- 1991, *A theory of Language and Information* London: Oxford Clarendon Press Haspelmath, Martin, 1997, *Indefinite pronouns*, Oxford: OUP
- 2001, Non canonical marking of core-arguments in European languages, in: Aikhenwald, A. Dixon, RMV, Onishi, M., eds. *Non-canonical marking of subjects and objects*, Benjamins
- 2007, Pre-established categories don't exist: consequences for language description and typology, *Ling. Typology* 11.1:119-132

- Heyting, Arend, 1955, *Les fondements des mathématiques, intuitionisme, théorie de la démonstration*, Paris : Gauthier-Villar
- Jenner, Philip N., Pou, Saveros, 1981, *A lexicon of Khmer Morphology*, Hawaiï University Press: Mon-Khmer Studies IX-X
- Kripke Saul A., 1982, *Naming and Necessity*, Oxford: Blackwell
- LaPolla, Randy J., (1992). "On the Dating and Nature of Verb Agreement in Tibeto-Burman." *Bull. of SOAS*, 35, 2, 298-315
1993. "Arguments against subject and direct second argument as viable concepts in Chinese", *Bull. of the Institute of History and Philology* 63.4, 759-813
- Menninger, Karl, 1934. 1969². *Number words and number symbols. A cultural history of numbers*, Cambridge Mass.: M.I.T. Press
- Renou Louis, 1952, *Grammaire de la langue védique*, CNRS et IAC
- 1957, *Terminologie Grammaticale du Sanskrit*, Paris : Champion
- Sapir, Edward. 1921, *Language*, New York: Harcourt, Brace & Company
- 1949, *Selected Writings*, ed. by D. Mandelbaum, University of California Press
- Spinoza, B., 1677¹ 1968² (pour la traduction en français), *Abrégé de grammaire hébraïque*, Paris : Vrin
- Tarsky, A., 1955, A lattic-theoretical fixpoint theorem and its applications, *Pacific J. Math.* 285-30
- Thurgood, G., 1985, Pronouns, verb agreement systems and the sub-grouping of Tibeto-Burman, in: *Linguistics of the Tibeto-Burman area: the state of the art*. Canberra: A.N.U.
- Van Stigt, W.P. 1990, *Brouwer's intuitionism*, Studies in the history and philosophy of mathematics, North Holland
- Witney, W. D., 1924¹ 1994⁷, *Sanskrit Grammar*, Delhi: Motilal Barnassidas
- Wittgenstein, Ludwig, 1974², *Philosophical Grammar*, Oxford: Blackwell
- Wouk, F and Ross M. Eds., 2003, *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics