

Aspects constructifs des grammaires de Zellig Harris

In: Langages, 25e année, n°99, 1990. pp. 57-84.

Citer ce document / Cite this document :

Daladier Anne. Aspects constructifs des grammaires de Zellig Harris. In: Langages, 25e année, n°99, 1990. pp. 57-84.

doi : 10.3406/lgge.1990.1592

http://www.persee.fr/web/revues/home/prescript/article/lgge_0458-726X_1990_num_25_99_1592

ASPECTS CONSTRUCTIFS DES GRAMMAIRES DE ZELLIG HARRIS

À partir d'un commentaire sur quelques questions, on essaiera de mettre en évidence le caractère constructif des analyses de Harris pour représenter les mécanismes d'interprétation propres aux langues naturelles. Les questions abordées seront : la notion de transformation et ses modifications permettant une meilleure intégration de la métalangue dans la langue ; l'enjeu de cette intégration et l'intérêt de cette approche par rapport à d'autres ; la construction commune des structures syntaxiques et des structures d'interprétation ; la séparation entre l'analyse des mécanismes d'interprétation dans une langue considérée dans son ensemble et l'analyse des mécanismes d'interprétation dans un univers de discours particulier.

0. Constructivisme

En analogie avec le constructivisme introduit en logique mathématique par Brouwer et Heyting, une forme de constructivité en linguistique a été développée par Harris. Cette idée me semble essentielle pour comprendre l'enjeu de son approche et je donnerai ici un aperçu sommaire de l'approche constructiviste en mathématiques¹ avant d'essayer de montrer dans cet article comment l'œuvre de Harris constitue par différents aspects une tentative pour définir une approche constructive de la linguistique.

Pour les constructivistes, des domaines comme l'algèbre et l'analyse, et leurs objets comme la notion d'ensemble, n'ont pas d'existence indépendante des descriptions qu'en donnent les mathématiciens. Les objets et les relations entre les objets de ces domaines doivent donc être l'objet de constructions explicites et la vérité ou la fausseté des énoncés mathématiques dépend de la connaissance (cohérence) que les mathématiciens produisent pour se les représenter. Un énoncé est vrai si on peut en construire une preuve et faux si sa preuve conduit à une contradiction. Les objets mathématiques s'apparentent à des algorithmes, avec en particulier la propriété de se terminer (principe de Markov). L'attitude des constructivistes s'oppose à celle des platoniciens pour qui les objets

1. Les ouvrages présentant les thèses constructivistes s'adressent à des logiciens. Cependant on trouvera une brève histoire du constructivisme avec d'importantes références bibliographiques dans le récent ouvrage de A. S. Troelstra et D. Van Dalen : *Constructivism in Mathematics. An Introduction*, Vol. 1, North-Holland, 1988. Dans cette présentation je ne ferai pas de distinction entre intuitionisme et constructivisme.

mathématiques et leurs propriétés ne sont que « redécouverts » par les mathématiciens. Ces objets existent dans une réalité préalable à l'activité mathématique, et leurs propriétés dans la ou les représentations mathématiques correspondent exactement à leurs propriétés dans cette réalité.

On peut comparer l'opposition entre constructivisme et platonisme en mathématique à l'opposition entre nominalisme et réalisme chez les grammairiens du Moyen Âge. Cette opposition est toujours présente dans les conceptions actuelles du langage. Pour les réalistes, la linguistique doit permettre d'accéder à une réalité préalable à l'activité de langage, cette réalité étant commune aux locuteurs, c'est-à-dire universelle. Les propriétés sémantiques des objets linguistiques seraient homologues aux propriétés des objets correspondants dans cette réalité, une fois débarrassés de leurs particularités et autres idiomatismes qui les attachent, en les limitant, à des langues particulières. Je reviendrai plus en détails sur cette question au paragraphe 6.

L'intérêt de ces oppositions me semble surtout lié à la façon dont elles rendent possibles certaines questions et les méthodes qui leurs sont associées, plutôt que d'autres. Par exemple, en logique mathématique, la position platonicienne ou réaliste s'est incarnée dans le développement de la sémantique, la théorie des modèles, notamment avec Tarski et Kripke. La position constructiviste, moins suivie, a développé des approches syntaxiques, notamment les théories de la démonstration. Récemment, des informaticiens se sont intéressés à l'analogie entre la construction et la vérification d'une preuve et la construction et la vérification d'un programme. La logique linéaire de Jean-Yves Girard² réanalyse en terme de mécanismes plus primitifs les opérations logiques habituelles et donne une interprétation informatique profonde de la façon dont les opérations d'un programme s'exécutent dans le temps. Je reviendrai plus en détails sur ces questions dans mon deuxième article où je montrerai que, de même qu'il existe une sémantique de la programmation, qui explicite pour la rationaliser la façon dont les opérations s'exécutent, on peut s'intéresser à une sémantique d'opérations qui permettrait d'analyser les procédés que les langues mettent en œuvre pour formuler et véhiculer des informations.

1. La notion de transformation en linguistique

La notion de transformation est issue des méthodes de classification des linguistes structuralistes. Dans un premier temps, Harris, le premier à utiliser cette notion en linguistique, a systématisé ces méthodes de classification, puis a montré que si on voulait analyser les mécanismes de formation des phrases pour l'ensemble d'une langue, on pouvait considérer la diversité considérable de ces

2. Jean-Yves Girard a donné un aperçu de la logique linéaire dans : « La logique linéaire », *Pour la science* (version française de *Scientific American*), Avril 1990. Les lecteurs déjà familiarisés avec le λ -calcul pourront trouver une bonne introduction aux applications de la théorie de la démonstration à l'informatique, au calcul des types du second ordre et à la logique linéaire dans : J.-Y. Girard, Y. Lafont, P. Taylor : *Proofs and Types*, Cambridge University Press, 1989.

phrases comme une diversité apparente. Dès l'origine, la transformation est un outil de régularisation, introduite avec une notion de phrase simple, la phrase noyau devant permettre de ramener des phrases quelconques à des combinaisons de phrases simples. Il s'agit au départ de décrire des faits tendant à l'exhaustivité tout en réduisant l'utilisation de listes *ad hoc*. La recherche de méthodes de régularisation correspond donc, comme le montre Tom Ryckman dans ce numéro, à la recherche des propriétés de la formulation d'informations en langue naturelle et non pas à la recherche de lois cognitives ou logiques supposées sous-jacentes au langage. D'autres notions de transformation linguistique ont été proposées par des chercheurs ayant d'autres buts.

Certains des logiciens s'intéressant aux langues naturelles ont pris conscience qu'on ne pouvait pas traduire directement, et d'une façon simple, les phrases en propositions. L'aspect qui semble avoir le plus retenu leur attention est le fait que les règles de bonne formation assignent des positions d'occurrence fixes aux prédicats et arguments dans les formules logiques alors que les contraintes de concaténation des éléments linguistiques sont moins simples. Les grammaires catégorielles de Lambek, par exemple, tiennent compte de cet aspect. Cependant ces grammaires, ainsi que les grammaires de constituants syntagmatiques définies dans le cadre des langages Context-Free de Kleene et Chomsky, ne suffisent pas à rendre compte des contraintes de formation des langues naturelles. Lambek, dans son article de 1961 : « On the calculus of syntactic types », *Twelfth Symposium on Applied Mathematics* (R. Jakobson ed.), p. 166-178, en vient à la conclusion que ce type de grammaire ne parviendra jamais à rendre compte de l'ensemble des phrases d'une langue, du fait du caractère elliptique de la formulation en langues naturelles. Il termine son article en espérant que les grammaires transformationnelles, qui commencent à se développer à cette époque, permettront d'apporter une solution à ce problème. Ses arguments ont été développés en 1964 par Bar-Hillel dans *Language and Information* (Addison-Wesley). Deux approches transformationnelles se sont développées, d'une part celle de Harris puis celle de Gross et de son école, d'autre part celle de Chomsky et des linguistes dits générativistes. Pour ces derniers, les langues naturelles ont deux aspects totalement indépendants. D'un côté, des contraintes de formation, la syntaxe ; de l'autre, les conditions dans lesquelles on peut assigner une valeur de vérité aux phrases, la sémantique. La syntaxe et la sémantique ont chacune leur typologie de règles et leurs catégories d'analyse. Cette attitude a de nombreux inconvénients descriptifs, déjà soulignés par Gross et son équipe, par exemple certaines ellipses responsables de violation des règles de constituants résultent de mécanismes d'interprétation. On peut également s'interroger sur les visées d'une telle approche. On reviendra à propos des sous-langages sur l'inadéquation d'identifier la sémantique des langues naturelles avec les sémantiques utilisées par les logiciens. On remarquera dès à présent que le résultat espéré de la composante sémantique des générativistes serait de traduire des phrases isolées en formules isolées d'un calcul de prédicats. Cet aspect pourrait surprendre dans la mesure où les logiciens ne s'intéressent pas aux formules des calculs de prédicats pour elles-mêmes, mais s'ils sont syntac-

ticiens, aux démonstrations, c'est-à-dire aux conditions dans lesquelles une formule peut être déduite d'une autre ou plus généralement, comment des hypothèses doivent être introduites pour obtenir une conclusion dans des conditions permettant de limiter les étapes de calcul. S'ils sont sémanticiens, ils s'intéressent aux modèles dans lesquels certains raisonnements peuvent être validés. Or, la notion même d'unité d'analyse des logiciens, la formule bien formée, est relative à leurs objectifs. Ses règles de formation sont pertinentes pour l'analyse des notions de déduction ou pour l'étude de la validation dans des modèles. Mais pour de nombreux linguistes, dont Harris, les principales propriétés des langues et du langage sont indépendantes des conditions dans lesquelles un raisonnement peut être formulé.

La plupart des approches formalisées en linguistique multiplient les composantes de leurs grammaires, chacune introduisant sa typologie de règles et ses catégories de description pour pallier les insuffisances des composantes initiales, qui sont conservées, faute semble-t-il d'alternatives formelles qu'on pourrait appliquer telles quelles à la linguistique. À l'inverse, l'élaboration harrissienne de la notion de transformation a consisté à essayer de construire une représentation appropriée aux contraintes empiriques et à revenir, à chaque reformulation de la grammaire, sur le choix des catégories initiales de description, jusqu'à faire disparaître les catégories auxiliaires dans la dernière de ces reformulations (cf. *A Grammar of English Principles* [GEMP]³ où pour l'analyse d'une langue dans son ensemble, les structures de phrases sont directement définies par l'application des entités lexicales les unes sur les autres, comme on le montrera au paragraphe suivant). Les éléments lexicaux ont une interprétation fonctionnelle qui est celle de la théorie des types du premier ordre définie par Alonzo Church dans : *The Calculi of Lambda-Conversion*, Annals of Mathematics Studies 6, Princeton University Press, 1941 (voir la note du paragraphe 2). Sans entrer dans les détails techniques de ce calcul, on peut remarquer que sa capacité descriptive est la même que celle de la grammaire catégorielle de Lambek. Il existe cependant une différence fondamentale entre la grammaire de Harris et celle de Lambek. La grammaire de Harris intègre toute une conception transformationnelle. Les règles de typage qui vérifient la bonne formation ne s'appliquent qu'après reconstruction de l'information linguistique implicite qui permet de définir la structure d'interprétation d'une phrase.

Dès l'étape distributionnelle de ses travaux, Harris minimise l'emploi de notions extra-linguistiques, en particulier l'emploi de catégories auxiliaires, pour que, dans une mesure qui s'est développée à chacune des étapes suivantes, chaque pas d'une dérivation de phrase puisse être formulé en terme de phrase et soit donc validable selon l'intuition du locuteur. Dans les premières pages de « Transformational Theory », article important dans la genèse de la notion de transformation, Harris montre qu'on peut intertraduire les représentations de

3. Les initiales entre crochets renvoient à la bibliographie des ouvrages de Harris, insérée après son article.

phrase en termes d'analyse en constituants, d'analyse en chaîne ou d'analyse transformationnelle. Chacune de ces représentations met l'accent sur des propriétés linguistiques différentes. L'analyse en constituants dépend plus que les autres de l'ordre de concaténation des mots. L'analyse en chaîne analyse les phrases en terme de phrases noyaux et adjonctions, l'analyse transformationnelle en terme de phrases noyaux et d'opérations : permutations, réductions et insertions d'opérateurs tels que : négation, adverbes, conjonctions. Les deux dérivations suivantes sont obtenues pour analyser : (1) *Un échantillon qu'un jeune naturaliste peut obtenir directement a souvent de la valeur*, la première dans la grammaire en chaîne, l'autre dans la grammaire transformationnelle :

[noyau]	(2) <i>un échantillon a de la valeur</i>
[adjonction 1]	<i>un échantillon + qu'un naturaliste peut obtenir</i>
[adjonction 2]	<i>un naturaliste + jeune</i>
[adjonction 3]	<i>peut obtenir + directement</i>
[adjonction 4]	<i>souvent + (2)</i>

L'adjonction, notée +, est une règle de concaténation. La relation entre la phrase (1) et la phrase (2) n'apparaît pas dans l'analyse en constituants. L'analyse transformationnelle va plus loin dans la décomposition en terme de phrases :

	(3) <i>un échantillon a de la valeur</i>
	(4) <i>un naturaliste obtient un échantillon</i>
	(5) <i>un naturaliste est jeune</i>
[souvent (3)]	(6) <i>un échantillon a souvent de la valeur</i>
[relativ. (4) (5)]	(7) <i>un naturaliste qui est jeune obtient un échantillon</i>
[qui est → E]	(8) <i>un naturaliste jeune obtient un échantillon</i>
[permu.]	(9) <i>un jeune naturaliste obtient un échantillon</i>
[peut (9)]	(10) <i>un jeune naturaliste peut obtenir un échantillon</i>
[directement (10)]	(11) <i>un jeune naturaliste peut obtenir directement un échantillon</i>
[relativ. (6) (11)]	(1) <i>un échantillon qu'un jeune naturaliste peut obtenir directement a souvent de la valeur</i>

L'analyse en chaîne fait intervenir une phrase noyau et cinq adjonctions tandis que l'analyse transformationnelle fait intervenir neuf phrases dont trois phrases noyau et six transformations. Elle fait donc intervenir neuf étapes de dérivation au lieu des cinq de la grammaire transformationnelle mais chaque étape est une phrase, qui peut être en tant que telle soumise à l'intuition du locuteur, et ceci permet de représenter en partie le langage de description dans la représentation de la langue. Autrement dit, une partie de la métalangue s'intègre à la langue décrite. Harris tente de généraliser cette propriété et j'y ferai référence dans ce qui suit sous la terminologie de l'hypothèse de la métalangue dans la langue. Je montrerai dans mon deuxième article qu'elle doit être affaiblie à l'hypothèse de la représentation de la métalangue dans la représentation de la langue.

Pour Harris comme pour ses prédécesseurs structuralistes, Bloomfield et surtout Sapir, ce sont les propriétés distributionnelles des mots dans chaque langue et non les entités extra-linguistiques qui déterminent la structure syntaxique et les structures d'interprétation des langues. Sapir et indépendamment Benveniste ont montré à propos d'aspects différents des langues comment les emplois lexicaux déterminent des découpages conceptuels propres aux différentes langues.

Si le problème du choix des catégories d'analyse est lié à la conception de la grammaire et à la nature de la sémantique à laquelle on s'intéresse, il est lié également aux propriétés formelles et en particulier à la capacité descriptive des grammaires. Les propriétés formelles de tel ou tel dispositif et ses mérites comparés avec d'autres dispositifs sont périodiquement l'objet de polémiques passionnées, surtout dans le cadre des systèmes informatisés de l'Intelligence Artificielle, mais les hypothèses sur les choix initiaux sont rarement discutées.

Les discussions sur la capacité descriptive des systèmes formels sont vides de sens si elles ne sont pas associées à la discussion du choix des catégories d'analyse. Les travaux du L.A.D.L. ont montré que de nombreuses relations de dépendance empiriques ne peuvent pas être analysées en termes de dépendances simples (grammaire context-free ou grammaires context-sensitive) si les règles de dépendance sont formulées à partir de catégories syntagmatiques, mais le peuvent si on les analyse en faisant intervenir les propriétés lexicales des autres mots de l'énoncé pour définir leur statut syntaxique. Je reviendrai sur cette question dans mon deuxième article.

La plupart des théories linguistiques font l'hypothèse que les règles de constituants syntagmatiques, sous leurs différentes variantes, fournissent une bonne approximation de la bonne formation en langue naturelle, ce qui s'est révélé inexact. Ces règles imposent un carcan préliminaire aux formes de composition des expressions lexicales qu'on cherche justement à mettre en évidence. Ces théories ajoutent ensuite des composantes ayant leur propre typologie de règles pour essayer de remédier aux inconvénients des règles d'insertions lexicales initiales. Il paraît plus judicieux de construire les représentations lexico-syntaxiques qui permettent de représenter de façon comparable les contraintes empiriques observées dans différentes langues.

Lorsqu'il a défini la notion de transformation comme une opération définie sur la structure des phrases, Harris se proposait d'utiliser les transformations pour formuler les propriétés empiriques des phrases. Les transformations devaient être l'outil principal pour l'étude des langues et leur comparaison devait permettre d'envisager l'étude du langage. La notion de phrase était définie à partir de celle de transformation. Une séquence de mots est une phrase si elle est en relation transformationnelle avec une phrase. Si cette dernière n'est pas elle-même une phrase noyau, il existe un nombre fini de transformations et de phrases dérivées qui la mettent en relation avec une phrase noyau. La compositionnalité des phrases est donc en partie définie par leurs relations avec d'autres phrases. Pour le reste, elle est définie par la compositionnalité des phrases noyaux.

La notion harrissienne de transformation pose des problèmes différents selon qu'on s'intéresse à l'analyse d'une langue dans son ensemble ou à l'analyse d'un sous-langage. Avant d'aborder cette distinction et la définition d'un sous-langage aux paragraphes 3 et 5 respectivement, il sera question ici de la notion de transformation dans le cadre de l'analyse d'une langue prise dans son ensemble. La façon dont Harris conçoit les transformations est indissociable de son objectif d'intégrer le langage de description d'une langue dans la langue elle-même (i.e. la métalangue dans la langue). Cette notion a été élaborée au cours de plusieurs étapes, dont l'enjeu semble avoir été chaque fois une amélioration de l'intégration de la métalangue dans la langue. On peut associer à cette préoccupation la recherche d'une sémantique constructive en affinant et en décomposant la description des mécanismes syntaxiques propres aux langues naturelles.

On peut distinguer deux grandes étapes dans l'élaboration de la notion de transformation. La première s'étend grosso modo des années cinquante au début des années soixante dix, la seconde sur les vingt années suivantes. Deux ouvrages les représentent assez bien : *Mathematical Structure of Language* [MSL] et *A grammar of English on Mathematical Principles* [GEMP]. Au cours de la première étape, les transformations sont définies comme des opérations qui s'appliquent à des structures de phrase et qui produisent d'autres structures de phrase. La notion de structure de phrase, mais non la notion de phrase elle-même (voir ci-dessus), préexiste donc à la notion de transformation. Les phrases noyaux sont construites comme des séquences de mots concaténés soumises à des conditions sur leur composition en termes de catégories grammaticales traditionnelles (verbe, nom, compléments du verbe, etc.). Les autres phrases sont construites à partir des phrases noyaux. Les transformations expriment principalement des contraintes sur les opérations de concaténation et sont soumises à des conditions d'application lexicales. Les transformations recouvrent des opérations de natures différentes, elles peuvent s'apparenter à l'insertion lexicale ; par exemple, à partir de deux phrases, on peut obtenir une troisième phrase en insérant une conjonction, cette insertion est non contrainte s'il s'agit d'une conjonction de coordination, elle est contrainte lexicalement s'il s'agit d'une conjonction de subordination. Les transformations peuvent être des opérations de permutation, par exemple le passif, également soumises à des contraintes lexicales. Elles peuvent aussi être des opérations de réduction, qui décrivent les ellipses. Au cours de la seconde étape, les opérations définies sur la structure des phrases disparaissent et toutes les contraintes empiriques sont réexprimées, voir paragraphe suivant.

Que ce soit dans le cadre de [MSL] ou dans celui de [GEMP], les transformations ont pour but de permettre de décomposer une phrase quelconque en termes de phrases canoniques. La dérivation transformationnelle d'une phrase jusqu'à bonne fin (i.e. l'existence d'une décomposition en phrases canoniques de cette phrase), fournit donc une preuve de sa bonne formation. Cependant l'aspect probablement le plus intéressant des transformations est qu'elles fournissent des méthodes internes à la langue qu'on veut analyser pour

induire des structures d'interprétation. En effet, les transformations sont soumises à des conditions d'application lexicales et / ou syntaxiques ; éventuellement des faits lexicaux ayant une certaine généralité servent de critères de classification pour obtenir des listes exhaustives, permettant dans certains cas d'obtenir ensuite une propriété transformationnelle qu'on peut à son tour tester sur des listes exhaustives en dégagant ses conditions d'application (voir Gross : *Méthodes en syntaxe*, Herman 1975 et Boons, Guillet, Leclère : *La structure des phrases simples en français*, Droz, 1976, en particulier l'exemple des constructions « croisées »). Lorsque les conditions sont satisfaites, on doit donc pouvoir appliquer automatiquement une transformation. L'analyse des propriétés lexicales sur des données exhaustives du français a été développée par Maurice Gross et l'équipe du LADL, initialement dans le cadre transformationnel de [MSL]. Cependant, ces propriétés débouchent sur une organisation du lexique français dont certains aspects sortent du cadre de [MSL]. On peut mentionner la mise en évidence de relations de dépendances lexicales à trois termes ou plus, qui ne peuvent être décrites dans les grammaires existantes, la remise en cause de certaines des catégories d'insertion lexicale traditionnelles pour la formulation des relations transformationnelles, l'emploi de prédicats sémantiques pour la formulation de nouveaux appariements de phrases.

Une notion de phrase noyau a pu être concrètement définie par Maurice Gross grâce à sa méthode de classification lexicale. Cependant, comme il l'esquisse dans son article, ses dérivations ont des objectifs sensiblement différents (voir également Alain Guillet : « Morphologie et syntaxe : quelques exemples d'interactions », *Recherches Linguistiques* 3, Université Paris 8, 1975 et Annie Meunier : « Sur les bases syntaxiques de la morphologie dérivationnelle, *Linguisticae Investigationes*, Vol. 1, Benjamins, 1977). L'analyse des phrases complexes ne se fait en général pas uniquement en termes de combinaisons de phrases simples mais peut utiliser des prédicats sémantiques. Cette notion est définie dans : « Les bases empiriques de la notion de prédicat sémantique », *Langages* 63, 1981, et n'est pas compatible avec l'hypothèse de la métalangue dans la langue. Par exemple, bien que le verbe *ranger* ne puisse être asserté avec un argument phrastique, il existe un prédicat sémantique *ranger* qui joue une sorte de rôle causatif permettant d'analyser (a) à partir de (b) :

- (a) *Max range les boîtes dans l'armoire*
- (b) *Max range [les boîtes sont dans l'armoire]*

Une analyse de l'ensemble des constructions locatives du français a été proposée selon cette méthode par Alain Guillet et Christian Leclère dans : *La structure des phrases simples en français. Verbes à complément direct et verbes à complément locatif*. Droz, Genève, 1990.

La notion de prédicat sémantique et l'assimilation des notions d'acceptabilité et d'assertabilité font de la notion de transformation élaborée par Maurice Gross une notion nouvelle. On remarquera qu'elle contient en les unifiant certains aspects des dérivations de la grammaire en chaîne et certains aspects des dérivations transformationnelles de [MSL] en leur ajoutant une spécificité lexicale.

Harris a développé, surtout dans le cadre transformationnel de [GEMP], l'utilisation des propriétés syntaxiques pour induire des hypothèses de reconstruction de structures informatives. Cet aspect est particulièrement intéressant, cependant sa méthode pose des problèmes qui amènent à modifier le cadre formel sous-jacent à [GEMP]. Je montre dans le deuxième article de ce numéro que ces problèmes sont dus d'une part à l'absence de représentation d'informations lexicales simultanément aux représentations syntaxiques. D'autre part à une incompatibilité entre la notion d'opérateur lexical et les réductions linguistiques, qui conduit à redéfinir ces derniers.

2. Transformations dans [GEMP] et reformulation des structures d'insertion lexicales (élimination des catégories auxiliaires)

La reformulation des structures d'insertion lexicales à partir des opérateurs lexicaux de [GEMP] a plusieurs avantages sur lesquels je reviendrai dans les paragraphes suivants. Peut-être faut-il commencer par le fait qu'elle rend explicite une hypothèse fondamentale des méthodes distributionnelles : l'idée que le sens des mots peut être défini comme l'ensemble de leurs propriétés distributionnelles. Dans [GEMP], le fait pour un mot, ou plus précisément pour une expression lexicale, d'imposer ou non des restrictions de sélection lexicales correspond au fait que cet élément lexical est ou non représenté comme un opérateur. Par exemple, le fait d'analyser certaines prépositions comme des opérateurs dont les arguments sont d'une part un autre opérateur et d'autre part un argument élémentaire correspond à l'existence de restrictions de sélection imposées par ces prépositions sur d'autres mots à statut d'opérateurs, verbes ou nominalisations. Cette analyse permet de préciser comment le sens d'un mot est déterminé par ses différents emplois. Le sens d'un mot opérateur dépend du sens de ses arguments, le cas échéant du sens de ses arguments d'arguments jusqu'à ce que sa descendance se termine par des arguments de statut élémentaire.

Un ordre applicatif sur les insertions lexicales est nécessaire pour exprimer ce que Harris appelle la structure d'entrée des phrases, c'est-à-dire l'ordre dans lequel les mots s'appliquent les uns aux autres pour produire la structure d'interprétation de la phrase. Par exemple, un mot comme *dort* requiert la présence d'un autre mot comme *quelqu'un* pour être interprété. De la même façon, un mot comme *longtemps* requiert la présence d'une phrase ; on dit que *dort* opère (est un opérateur) sur *quelqu'un* et que *longtemps* opère sur une phrase, par exemple *dort* lui-même opérant sur *quelqu'un*, ce qui fournirait la structure applicative de *quelqu'un dort longtemps* : *longtemps (dort (quelqu'un))*. Les mots : *quelqu'un*, *dort*, *longtemps*, doivent être distingués du point de vue de leur statut *syntactique*, leur statut applicatif ⁴ (le nombre et le caractère élémentaire

4. Le statut applicatif des opérateurs recouvre la notion de type dont plusieurs spécimens ont été définis en logique. Cette notion a été introduite par Bertrand Russel pour éviter le paradoxe qui mettait en question le fondement des ensembles (voir l'article d'André Lentin). La notion de type a également été redéfinie par Alonzo Church après la découverte par Kleene et Rosser d'un paradoxe

ou non de leurs arguments) pour exprimer la structure d'interprétation des phrases, dans des cas simples comme l'exemple précédent mais également pour rendre compte des propriétés de formulation elliptiques des langues naturelles. Par exemple, le fait que l'interprétation de *dormir* requiert toujours un argument élémentaire est nécessaire pour interpréter les phrases suivantes où l'argument élémentaire non explicite de *dormir* est obligatoirement identifié à l'un des arguments d'un opérateur de niveau supérieur à *dormir*, selon des contraintes lexicales de « projection » définies par Maurice Gross dans les tables de *Méthodes en Syntaxe*, pour chaque verbe à complétive, comme dans (1) et (1'). Ces contraintes de projection peuvent être étendues à d'autres opérateurs à argument de statut non élémentaire, comme la préposition *pour* dans (3). Ces contraintes de projection ne sont pas limitées à une relation entre un infinitif et l'opérateur dont il est l'argument immédiat comme dans (2) et (3) où la réduction de l'argument élémentaire de *dormir* n'est pas déclenchée par son opérateur direct. De plus, plusieurs éléments lexicaux induisant des contraintes de projection peuvent occuper dans un même énoncé. Les contraintes de projection devront être hiérarchisées selon les propriétés lexicales des éléments qui les induisent. Dans un grand nombre de cas, les identifications qui prennent place à l'intérieur d'une sous-expression ont priorité sur celles qui font intervenir une sous-expression de niveau supérieur comme dans (2) et (4), la notion d'opérateur doit rendre compte de cette sorte d'information :

- (1) *Luc a ordonné à Jean de Ø dormir.*
- (1') *Luc a promis à Jean de Ø dormir.*
- (2) *Luc a supplié Jean de le Ø laisser Ø dormir.*
- (2') *Marc a supplié Paul de Ø dormir. Tu me laisseras Ø dormir.*
- (3) *Pour Jean, Ø dormir est un plaisir.*

dans son Lambda-Calcul. La complétude déductive et la complétude fonctionnelle rendaient son calcul inconsistent. Ce paradoxe est connu comme le paradoxe de Curry, qui en a donné une formulation dans le cadre de sa logique combinatoire. En collaboration avec Hindley et Seldin, il a ensuite élaboré plusieurs formes de fonctionnalité qui sont compatibles avec des formes plus ou moins faibles de déduction, dans le tome 2 de *Combinatory Logic*, North-Holland, 1972. La théorie des types simples de Church consiste à restreindre de façon drastique la notion de fonction utilisée dans son calcul. Des notions de fonction plus puissantes et n'introduisant pas de paradoxe ont été définies ensuite, en définissant des calculs de type d'ordre supérieur ou de façon équivalente en construisant des modèles algébriques, par exemple la théorie des catégories de Lambek, ou le modèle d'enchâssement de treillis continus de Dana Scott. Dans un cadre constructif, il faut mentionner les théories de Salomon Feferman, de Per Martin-Löf et de Girard (le système F puis la logique linéaire). Ces deux systèmes de types sont présentés dans l'ouvrage précédemment mentionné. Plusieurs de ces théories sont utilisées pour construire des langages de programmation. Les lecteurs familiarisés avec le calcul des prédicats pourront se reporter au livre de Hindley et Seldin : *Introduction to Combinators and λ -Calculus*, Cambridge University Press, qui présente un aperçu assez complet et remarquablement pédagogique de ces questions. Pour les autres lecteurs, une introduction donnant accès aux sources de ces questions est le livre de Jean Van Heijenoort, *From Frege to Gödel : A Source Book in Mathematical Logic, 1879-1931*, Harvard University Press. La notation utilisée par Harris et les exemples analysés dans [GEMP] peuvent être représentés dans la théorie des types simples de Church définie dans : *The Calculi of Lambda Conversion*, Princeton University Press, 1941.

(4) *Pour le commissaire, Luc a ordonné à $\overline{\text{Jean de } \emptyset}$ dormir.*

Ces quelques exemples suffisent à indiquer que les reconstructions ne font pas intervenir que le statut applicatif des mots, et ne sont pas limitées à la relation entre un opérateur et son argument selon la définition qu'en a proposée Harris, mais également des contraintes lexicales (contraintes de « projection » associées à *ordonner*, *promettre*, *supplier*, *pour*) et la hiérarchisation de ces contraintes. Les reconstructions peuvent dans bien d'autres situations faire intervenir des combinaisons de contraintes lexicales et syntaxiques. Ces reconstructions concernent également des arguments non élémentaires. Ainsi, c'est parce qu'on fait l'hypothèse que *espoir* requiert toujours un argument non élémentaire en plus de son argument élémentaire qu'on peut reconstruire l'interprétation implicite de :

Luc a dit qu'il viendra et j'ai bon espoir pour Jean $\emptyset$

Dans ce contexte, *J'ai bon espoir pour Jean $\emptyset$* doit être reconstruit : *J'ai bon espoir pour Jean qu'il (Jean) viendra.*

La grammaire des opérateurs de Harris distingue deux ordres sur les mots. L'ordre dans lequel les mots d'une phrase s'appliquent les uns aux autres pour constituer sa structure d'interprétation n'est pas l'ordre séquentiel dans lequel ces mots sont dits ou écrits, cet ordre séquentiel s'obtient à partir de l'ordre applicatif par des opérations de « linéarisation ». L'ordre dans lequel les opérateurs sont linéarisés doit rendre compte des contraintes observées sur la concaténation. Cet ordre ne doit pas être confondu avec l'ordre applicatif des opérateurs qui doit rendre compte d'autres propriétés syntaxiques et des structures d'interprétation. Une phrase donnée, non ambiguë, a une structure applicative, c'est-à-dire une structure d'interprétation, unique. Mais cette structure applicative peut souvent être linéarisée de plusieurs façons, par exemple : *Jean a dormi longtemps*, et *Jean a longtemps dormi*, donc deux linéarisations de la même structure informative indiquée plus haut. Par contre *Jean a mangé un tigre* et *Un tigre a mangé Jean* ont des structures applicatives différentes, qui font intervenir une permutation des arguments élémentaires de *manger* (voir plus bas). La principale nouveauté de [GEMP] par rapport aux autres théories transformationnelles est que l'insertion lexicale ne se fait plus par l'intermédiaire de catégories auxiliaires comme N, V, AUX, NEG, Conj, etc. comme dans les théories dites génératives ou par l'intermédiaire de catégories auxiliaires et d'opérateurs grammaticaux, donc métalinguistiques, comme dans [MSL]. La structure d'interprétation des phrases s'exprime directement par l'application ordonnée de leurs éléments lexicaux explicites ou reconstruits. Les transformations ne sont donc plus des opérations de restructuration sur la structure des phrases mais seulement des opérations de substitutions de formes zéro (ou de formes morpho-phonétiques appropriées) dans les structures applicatives restant identiques des phrases d'une dérivation. Ces réductions correspondent aux éliminations de redondance, à la formulation elliptique de l'information, propre à une langue.

Les cinq points suivants pourraient servir d'introduction à la lecture de [GEMP] :

1. La structure applicative, la structure d'insertion lexicale et la structure d'interprétation d'une phrase sont un seul et même objet.

2. Dans la plupart des cas, cet objet met en jeu des reconstructions d'éléments morphologiquement réduits, éventuellement à des formes zéro. Ces réductions font partie de la formulation elliptique de l'information propre aux langues naturelles ; elles sont des substitutions de formes morphologiques réduites (pronoms, formes partielles, formes condensées en affixes, suffixes ou morphologie verbale, formes zéro) à des formes lexicales pleines, elles ne modifient pas les structures applicatives. La reconstruction d'une structure informative se fait en utilisant le statut applicatif des entités lexicales. Lorsque les arguments d'un opérateur manquent, on induit l'hypothèse qu'ils ont été l'objet d'une réduction et qu'ils figurent implicitement dans la structure informative. La reconstruction des structures informatives, ordre d'insertion des opérateurs et statut applicatif de ceux-ci, peut être formalisée d'une manière purement syntaxique mais la reconstruction du contenu lexical exact des éléments réduits nécessiterait le plus souvent d'autres procédures lexicales, ou des procédures qui feraient intervenir simultanément des contraintes lexicales et syntaxiques, procédures que ne fournit pas la grammaire des opérateurs.

3. Une structure applicative peut être linéarisée sous la forme d'une phrase, ou le cas échéant de plusieurs phrases s'il y a plusieurs façons de concaténer les mots, en lui appliquant des règles de linéarisation. Par exemple, la structure applicative : *longtemps < a dormi < Jean*, peut être linéarisée de plusieurs façons : *Jean a longtemps dormi*, *Jean a dormi longtemps*. Une phrase ambiguë a plusieurs structures applicatives qui rendent compte de chacune de ces interprétations. Chaque structure d'interprétation peut elle-même être éventuellement linéarisée de plusieurs façons. Ces règles n'ont pas été explicitement formulées pour l'ensemble d'une langue mais elles sont relativement simples à imaginer pour le français et l'anglais. Les règles de linéarisation sont définies sur les structures applicatives c'est-à-dire qu'elles dépendent du statut applicatif des mots. Elles peuvent aussi dépendre de traits lexicaux ; par exemple en français, le sujet ou plus précisément l'argument de rang d'un opérateur se concatène à gauche si cet opérateur a une forme verbale, à droite s'il a une forme nominale : *Jean a décrit...*, *la description de Jean...*

4. Les différentes étapes de la reconstruction d'une structure applicative peuvent, elles aussi, être linéarisées comme phrases, ce qui correspond à l'idée que la décomposition syntaxique d'une phrase est elle-même paraphrasable en phrases. L'ensemble de ces étapes de reconstruction est une dérivation transformationnelle. Les phrases reliées transformationnellement ont donc la même structure syntaxique.

5. Quand une même structure applicative peut être linéarisée de plusieurs façons, les phrases correspondantes ne constituent pas une dérivation et ne sont pas transformationnellement reliées.

Dans une série de conférences données en 1974 à l'université de Vincennes, qui ont été ensuite reprises et traduites dans : *Notes du Cours de Syntaxe*, Le Seuil, 1976, Harris a proposé une étape préliminaire de la conception qu'il a développée dans [GEMP]. Dans cette étape préliminaire, peut-être devrait-on dire hybride entre [MSL] et [GEMP], la notion de transformation recouvre à la fois les opérations de permutation sur la concaténation des mots et des opérations de substitution de formes réduites ou de formes zéro dans des structures applicatives. Cette notion de transformation ne permettait donc pas de définir la structure syntaxique comme étant la structure d'interprétation d'une phrase. La permutation d'un mot dans la structure applicative et la permutation d'un mot dans la structure de concaténation d'une phrase n'ont pas du tout le même effet linguistique, mais elles n'étaient pas distinguées alors. Par exemple, une permutation de l'ordre applicatif des arguments de *mange* produit une modification de la structure informative, ainsi (a) et (b) ont pour linéarisations respectives (A) et (B) :

- (a) *a mangé (Jean, un tigre)*
- (b) *a mangé (un tigre, Jean)*
- (A) *Jean a mangé un tigre.*
- (B) *Un tigre a mangé Jean.*

Les règles de linéarisation utilisent le statut applicatif des mots mais aussi des informations lexicales. Dans certains cas, l'utilisation de propriétés de linéarisation peut faciliter la reconnaissance de la structure applicative d'une phrase.

On peut résumer ces cinq points de la façon suivante. Vérification de la bonne formation et reconstruction de la structure informative d'un texte doivent être deux aspects d'un même processus. Pour une langue prise dans son ensemble, c'est-à-dire indépendamment des univers de discours spécifiques où on peut l'utiliser, les phrases sont toutes soumises au même critère de compositionnalité, la satisfaction des relations opérateurs / arguments des entités lexicales, relations qui définissent en partie les phrases bien formées. Lorsqu'on veut analyser une phrase existante, ces relations permettent également d'induire l'existence d'une réduction sous-jacente si ces dernières sont apparemment non satisfaites. Dans le cas de l'analyse d'un sous-langage, le caractère applicatif de la grammaire fournit aussi une méthode de reconstruction de l'information implicite (voir plus bas).

Pour devenir des phrases, en plus de la satisfaction des relations opérateurs-arguments, les structures applicatives doivent satisfaire les règles de linéarisation, elles doivent également satisfaire des règles morpho-phonémiques (règles d'accord), elles doivent enfin satisfaire des conditions sur l'insertion des temps conjugués, qui sont encore insuffisamment décrites.

3. Structures syntaxiques et structures informatives

Harris utilise l'expression « structure informative » plutôt que « structure sémantique » pour éviter toute confusion avec l'emploi de ce terme dans les

analyses non constructives, c'est-à-dire la majorité des analyses actuelles. Si on met à part cette approche constructive des mécanismes d'interprétation propres aux langues, on peut considérer qu'il existe principalement deux sortes de relation entre syntaxe et sémantique dans les approches actuelles : celle qui a été développée par les philosophes du langage et celle qui a été développée par les générativistes, et les chercheurs de l'Intelligence Artificielle.

Chez les philosophes du langage, la relation entre syntaxe et sémantique est une relation directe : la sémantique permet d'interpréter la syntaxe. Mais cette relation ne tient pas compte ou très peu de la spécificité informative des langues naturelles, les phrases étant ramenées à des formules bien formées des calculs de prédicats. Cette position est illustrée par différentes tentatives, dont les représentants les plus connus sont Montague, Kripke, Kamp et Barwise. Pour ces logiciens s'intéressant à une extension de l'application de leur outil logique, l'interprétation d'une phrase se ramène aux conditions dans lesquelles on peut lui assigner une valeur de vérité. Comme logiciens et héritiers de la conception de Tarski de la vérité, ces chercheurs ont surtout développé la sémantique de la logique mathématique, c'est-à-dire la théorie des modèles, reléguant l'intérêt pour la syntaxe au culte des grands ancêtres (contrairement à la tradition des logiciens constructivistes). Ils se sont donc intéressés à des modèles qui étendent certains traitements des modèles déjà existants, par exemple à la sémantique des mondes possibles et aux traitements de phrases traduisant des propositions paradoxales. Une fois leurs objectifs sémantiques atteints, leur syntaxe s'obtient par simple jeu d'écriture. La finalité de ces tentatives, qui diffèrent souvent entre elles par des caractéristiques techniques importantes, parfois fructueuses en logique mathématiques, repose sur des préoccupations ontologiques : quand, comment et dans quelles conditions dit-on le Vrai ? Ces questions intéressent les philosophes du langage mais ne devraient intéresser le linguiste que dans le cadre restreint de l'analyse du sous-langage que constitue la formulation en langue naturelle de la logique ou de l'ontologie. Précisons également que comme théorie, ces philosophies du langage reposent sur des primitives sémantiques issues de l'intuition de leurs auteurs sur ce que doit être une « bonne » ontologie. Par exemple la solution de Tarski au paradoxe du menteur laisse Kripke (dans : « *Outlines of a Theory of Truth* », *The Journal of Philosophy*, 1975) insatisfait parce que pour lui la vérité doit être définie par un prédicat unique.

Pour les informaticiens qui tentent de simuler des formes de « compréhension » du langage naturel, l'accent est également mis sur la sémantique, avec ici aussi l'application de modèles de la logique (classique ou non). Mais à la différence des philosophes du langage, leur expérience de textes ou de dialogues réels leur a fait abandonner la représentation syntaxique des phrases par des formules bien formées des calculs de prédicat. Ils ont donc ajouté à la composante sémantique de leur représentation linguistique, qui contient une sémantique au sens logique du terme associée à sa propre syntaxe, une composante syntaxique (au sens linguistique) préliminaire dont le but est de permettre de coupler la reconnaissance et la production de phrases et les représentations sémantiques de la logique. Cette composante préliminaire est

généralement empruntée aux grammaires des générativistes. C'est une grammaire de réécriture utilisant les constituants syntagmatiques comme catégories auxiliaires et, dans une mesure qui peut varier, des informations lexicales. Son objectif semble être principalement de rendre compte des contraintes sur l'ordre des mots dans les phrases écrites ou parlées, c'est-à-dire des contraintes de concaténation propres à chaque langue. Les linguistes générativistes mettent l'accent sur l'analyse de contraintes langagières et ont développé des opérations de transformation avec une technique de « pointeurs » appelés « traces » pour associer des formes de phrase dont les contenus sémantiques leur semblent devoir être identifiés. Les chercheurs de l'Intelligence Artificielle, eux, mettent l'accent sur une forme de compréhension effective prenant place dans des univers de discours particuliers, compréhension qui est simulée pour un échantillon d'expressions d'une langue donnée, jugé représentatif ou « utile ». Dans la mesure où de nombreux traitements en Intelligence Artificielle utilisent des règles de constituants syntagmatiques, ces deux démarches différentes ont souvent en commun l'hypothèse de l'autonomie des composantes syntaxiques et sémantiques. Les chercheurs de ces domaines considèrent que les contraintes sur la concaténation des mots décrites par les règles de constituants syntagmatiques⁵ rendent compte pour l'essentiel des contraintes sur la bonne formation. Ces hypothèses se révèlent inadéquates parce que la bonne formation en langue naturelle ne peut être décrite, ni même approximée, à l'aide de règles de réécriture s'appliquant à des catégories auxiliaires et ce pour deux familles de raisons : le caractère elliptique de la formulation d'informations en langue naturelle (ces ellipses étant dépendantes de mécanismes d'interprétation) et l'existence de relations de dépendance lexicales à trois termes ou plus que les règles de constituants ne peuvent permettre de décrire. Je reviens en détail sur ces problèmes dans mon deuxième article.

La conception de la relation entre syntaxe et sémantique que Harris a développée est fondamentalement différente. Le contenu informatif des phrases dans un sous-langage donné peut être exprimé de façon à la fois efficiente et économique en utilisant des propriétés distributionnelles de la formulation en langue naturelle, qui sont distinctes des contraintes de concaténation. Pour les raisons que Thomas Ryckman examine dans son article et sur lesquelles je reviendrai ci-dessous, l'analyse du contenu informatif des phrases recouvre deux problèmes distincts selon qu'il s'agit de phrases appartenant à un sous-langage ou de phrases appartenant à une langue considérée dans son ensemble.

5. Il existe en fait plusieurs sortes de grammaires utilisant des règles de constituants syntagmatiques. La tendance lexicaliste, illustrée par des chercheurs comme Joan Bresnan, ont tenté d'améliorer la composante syntaxique de la grammaire générative en subordonnant la représentation syntaxique à certaines formes de dépendance lexicale. Parmi ces grammaires, certaines utilisent des règles de réécriture contextuelles s'appliquant aux catégories syntagmatiques, les T.A.G. (tree adjunct grammars) de A. Joshi et ses collaborateurs. Ces grammaires améliorent, mais seulement dans une faible mesure eu égard aux problèmes non traités, la définition chomskienne de la bonne formation en langue naturelle.

4. Structures informatives d'une langue et structures informatives d'un sous-langage

Il existe une différence fondamentale entre la reconstruction des structures informatives des énoncés d'une langue considérée dans son ensemble et celles des énoncés d'un sous-langage. Dans le cadre d'un sous-langage, les phrases mettent en jeu un jugement d'acceptabilité défini, alors que dans le cas d'une langue en général, les phrases ne mettent en jeu que des jugements d'acceptabilité relatifs. On peut trouver bizarres des phrases comme (1) et (2) ci-dessous, voire même pour certains linguistes se demander si ce sont bien des phrases, pourtant il est possible de leur associer un contexte dans lequel leur interprétation devient naturelle :

(1) *Les cercles sont carrés.*

(2) *Marie lui téléphonera hier demain.*

Pour Quine, une phrase comme (1) pose un problème ontologique et devrait être interdite. Dans l'approche décrite ici, (1) n'est exclue que des sous-langages où elle est ininterprétable, c'est-à-dire ceux où *cercle* ne peut jamais occurrer dans le domaine de sélection de *être carré*. Les significations de *cercle* et de *carré* ne sont contradictoires que dans un univers de discours où elles ont été définies comme telles. Comme l'a fait remarquer Heyting dans : *Intuitionism, an introduction*, North-Holland, 1956, dans un tel univers, on ne peut même pas interpréter la négation de (1), on ne peut que constater qu'il n'est pas possible de construire un objet ayant les propriétés d'être rond et carré. Par contre, on peut imaginer un autre univers de discours où ces deux propriétés sont compatibles. Par exemple, quelqu'un fabriquant des images vidéo pourrait définir une transformation de cercle en carré et asserter une phrase comme (3) puis (1) dans ce contexte :

(1) *Les cercles ne sont pas encore carrés mais dans quelques secondes ils le seront.*

Supposons que Eva est en train de faire un film et que l'ordre dans lequel les épisodes du film sont tournés ne soit pas leur ordre chronologique dans l'histoire. Aujourd'hui mardi, Eva décide que l'épisode où Marie appelle Marc sera filmé le lendemain, mercredi, et que l'épisode qui a lieu le lendemain de ce coup de téléphone sera filmé aujourd'hui. Eva peut donc annoncer aux acteurs la phrase (2). Un aspect intéressant de la formulation dans (2) est qu'elle induit auprès d'un auditeur l'hypothèse que le locuteur est en train de définir un univers de discours qui implique plusieurs notions de temps simultanées.

Hors contexte, les phrases (1) et (2) sont peu acceptables ; selon la terminologie de Harris, elles ont une vraisemblance d'occurrence faible. Cependant, dans une langue considérée dans son ensemble, toutes les séquences de mots bien formées ont une valeur informative, ne serait-ce que pour nous permettre de les juger absurdes. Il n'en va pas de même dans les sous-langages où la bonne formation d'une phrase n'est pas une condition suffisante pour qu'elle soit interprétable, il faut aussi qu'elle soit acceptable dans ce sous-langage spécifique comme l'exemple (1) ci-dessus qui n'est pas une phrase dans le sous-langage de la géométrie scolaire. On prendra un autre exemple en biologie, la phrase :

Des antigènes ont été injectés dans les anticorps d'un lapin.

est dénuée de valeur informative. Il ne serait pas suffisant de dire que cette phrase représente une information fausse parce qu'il n'y a aucune façon d'obtenir une signification, même fausse, à partir des combinaisons de mots qui la constituent, pour un biologiste. Dans un sous-langage, une phrase est interprétable si elle est acceptable.

On peut construire une représentation, les mathématiciens ont parlé d'idéographie, qui prenne en compte seulement l'information pertinente pour représenter les phrases acceptables d'un sous-langage donné. L'existence d'une intuition d'acceptabilité définie pour les phrases d'un sous-langage va de pair avec le fait que les mots ont des restrictions d'emploi et donc des significations particulières à ce sous-langage.

La démarche de Harris peut être rapprochée de celle de Péano insistant sur l'importance des définitions et la nécessité d'une idéographie pour distinguer l'interprétation des phrases en mathématique de l'interprétation des phrases dans une langue considérée dans sa généralité. Dans plusieurs articles sur les définitions en mathématique, notamment : « Le definizioni in matematica », 1921 (repris dans : *Selected works of G. Peano*, Kennedy ed., U. of Toronto Press, 1973), Peano explique que toutes les définitions ne sont que des définitions de noms et non d'une réalité. L'existence de la chose définie n'est pas nécessaire. Les définitions utilisent des notions primitives qui ne peuvent être définies dans le langage où elles sont exprimées. Le rôle des définitions est de faire apparaître explicitement quelles sont les notions primitives, comment les notions primitives sont combinées pour produire des notions dérivées et comment ces notions sont reliées entre elles. La sémantique d'une langue dans laquelle serait formulée l'arithmétique ne doit pas être confondue avec la sémantique de l'arithmétique. Les termes *un*, *deux*, *trois*, *multiplication*, *point*, *ligne*, *plan* appartiennent au langage ordinaire mais la question de savoir si ces termes peuvent être définis nécessite une analyse du langage, un répertoire de tous les mots et flexions grammaticales des premières pages des textes où sont exprimées les sciences qui les utilisent. « Dans les expressions *cinq mètres*, *cinq cent*, l'apposition a la valeur de *multiplié par* et il est impossible de définir l'apposition à l'aide d'autres appositions. Une expression comme : *le produit de deux nombres est la somme de autant de termes égaux au multiplicande qu'il y a d'unités dans le multiplicateur*, n'apprend rien à quelqu'un qui ne sait pas déjà ce qu'est la multiplication. Un concept mathématique exprimé par l'apposition peut être exprimé ailleurs par d'autres moyens grammaticaux. L'important est qu'une fois qu'on a dégagé les primitives d'une notion et leur représentation dans le langage ordinaire ou dans une idéographie, les définitions qui utilisent cette notion ou des notions dérivées ne doivent faire intervenir que ces représentations primitives. »

L'existence d'un jugement d'acceptabilité fournit une unité d'analyse, c'est l'unité à laquelle s'applique le jugement. Par exemple, la validation s'applique à une formule bien formée du calcul des prédicats. Ce fait est à rapprocher des difficultés que pose le choix d'une unité d'analyse pour une langue prise dans son

ensemble. Dans son article, Thomas Ryckman analyse la structure des unités d'information définies par Harris pour des textes de biologie. En se reportant aux exemples qu'il utilise, on vérifiera que les phrases représentées par les unités d'information ne correspondent en général pas à la notion intuitive de phrase simple. Le choix précis de la notion de formule bien formée en calcul de prédicats est déterminée par l'expression de la validité d'une déduction. Ainsi, de même que les unités d'information ne représentent que l'information pertinente pour l'expression d'assertions acceptables en biologie, les formules bien formées ne représentent que l'information pertinente pour l'expression de déductions valides. On peut se demander comment la représentation de phrases isolées par des formules bien formées du calcul des prédicats a pu acquérir un tel prestige auprès de certains linguistes et on peut également s'interroger sur l'objectif exact de ceux qui essaient de ramener toute interprétation de textes n'appartenant pas à un sous-langage particulier à une validation de déductions. Bien entendu, il est toujours possible d'utiliser la sémantique d'un calcul de prédicats ou celle d'un langage de programmation et de traduire les manipulations acceptables d'expressions symboliques en langage naturel. C'est ce que font de nombreux chercheurs en Intelligence Artificielle lorsqu'ils réalisent des interfaces homme-machine. Les victoires et les difficultés dans ce domaine sont celles des représentations symboliques, c'est-à-dire celles des sémantiques utilisées en logique ou en informatique théorique. Représenter des propositions autoréférentielles ou des propositions qui n'ont pas toujours une valeur de vérité peut être fait dans des modèles disposant d'une forme de complétion comme dans la *Sémantique des Situations* de Barwise et Perry. Mais ce genre d'approche ne nous dit rien de nouveau sur la sémantique des langues naturelles.

Un des aspects importants de la représentation d'un sous-langage est que sa structure (sa grammaire) peut être considérablement simplifiée et pourtant précisée du point de vue de la représentation de son contenu informatif du fait de l'existence d'une notion stable d'acceptabilité (voir le paragraphe suivant). Cependant, cette représentation utilise de façon fondamentale les propriétés linguistiques de la formulation de l'information dans ce domaine particulier, c'est-à-dire les restrictions de sélection lexicales spécifiques à ce domaine (on peut même dire que dans une certaine mesure, ce sont ces restrictions qui font de ces textes un sous-langage). Pour ces raisons, cette représentation se distingue fondamentalement de celles qui sont actuellement utilisées en Intelligence Artificielle où l'analyse sémantique est indépendante de la formulation linguistique et se fait à partir de notions définies avant les analyses (concepts, « frames », « mops », réseaux sémantiques etc.) qui sont justifiées par des raisons extérieures au domaine d'analyse, en général des raisons d'ordre cognitif reposant sur les intuitions des informaticiens ou des psychologues, qui font implicitement l'hypothèse qu'il existe des mécanismes de représentation ou de conceptualisation de l'information très généraux, indépendants de la formulation de cette information.

Un autre aspect important, lié au précédent, est que, aussi bien pour la grammaire d'une langue dans son ensemble que pour la grammaire d'un

sous-langage, on distingue deux niveaux syntaxiques : d'une part la structure applicative qui permet d'obtenir la structure d'interprétation des phrases ou des unités d'information ; d'autre part, la « linéarisation » de ces structures applicatives, les unes en phrases et les autres en unités d'information, selon des règles de séquentialisation qui leur sont propres. Le premier niveau est évidemment ici le plus important. Il est donc à noter que l'ordre (séquentiel) des mots dans les phrases, qui joue un rôle si important pour déterminer les structures syntaxiques, et donc pour la formulation des contraintes syntaxiques, dans la plupart des grammaires existantes, ne joue qu'un rôle secondaire dans le dernier cadre harrissien ([GEMP] et [TLI]).

Dans le cas d'une langue prise dans son ensemble, les structures applicatives, qui sont les structures d'insertion lexicales, sont aussi les structures d'interprétation. Elles doivent rendre compte de la bonne formation moins les contraintes de concaténation et les règles d'ajustements morphologiques (e.g. règles d'accord) dans la langue analysée. Dans le cas des sous-langages, les structures applicatives des unités d'information ne sont pas les structures des phrases correspondantes. Elles rendent compte des restrictions de sélection lexicales qui sont spécifiques au sous-langage étudié. Ce sont des structures d'insertion des catégories informatives et les structures informatives du sous-langage. Elles rendent compte des combinaisons de catégories informatives acceptables dans le sous-langage indépendamment des langues particulières dans lesquelles l'information est formulée et non de la bonne formation des phrases que les unités d'information représentent. J'ai proposé dans [FIS] une grammaire des linéarisations des unités d'informations du sous-langage étudié pour le français.

5. Analyse du contenu informatif de textes appartenant à un sous-langage

Harris *et al.* dans : *The Form of Information in Science* [FIS] ont montré comment la représentation des restrictions de sélection lexicales spécifiques aux phrases d'un univers de discours permettait de définir une grammaire informative du sous-langage. Cette grammaire représente directement le contenu informatif pertinent des textes du sous-langage. Le choix des catégories informatives, celui des unités d'information comme unités d'analyse et celui de leur structure applicative résultent de cet objectif. Ces choix ont été faits à partir de plusieurs méthodes conférant différentes propriétés à la grammaire. Ainsi, celle-ci rend compte du sens contextuel des mots dans un formalisme non contextuel.

Elle fournit en même temps une unité d'analyse plus précise et plus opératoire que la notion de phrase pour le domaine étudié. Les catégories d'analyse du contenu informatif de ces textes ont été pour la plupart induites en employant des méthodes d'analyse distributionnelles, de la formulation de l'information dans ce domaine. Seules les catégories « élémentaires » c'est-à-dire celles dont le sens ne dépend pas d'autres catégories, et qui sont représentées pour cette raison comme des arguments terminaux de catégories ou de combinaisons de catégories de niveau supérieur, ont été directement introduites par des experts du domaine (i.e. de façon non constructive).

En résumé, cette grammaire est une construction simultanée de la forme des objets qu'on veut interpréter (les structures informatives) et des propriétés sémantiques (leur contenu informatif) adaptées au but de la représentation, c'est-à-dire une représentation de la connaissance « utile » et de son acquisition, formulée de façon non ambiguë mais concise. En définissant sur cette grammaire applicative des unités d'informations des règles de linéarisation et en associant aux catégories informatives un dictionnaire d'expressions lexicales dont l'organisation est adaptée aux règles de linéarisation, on vérifie que les phrases effectivement réalisées dans ce sous-langage peuvent être reconnues et produites sans passer par l'étape de règles de concaténation des mots dans les phrases. Par organisation du dictionnaire, on entend principalement deux aspects, d'une part un choix approprié d'expressions lexicales et non de mots (ces expressions correspondent en général à des constituants discontinus) et d'autre part, une marque pour les mots de statut non élémentaire avec un pointeur droit ou un pointeur gauche. Un système d'indexation pour les différentes formes d'accord ou de flexion doit être également associée au dictionnaire. Le dictionnaire et son organisation est spécifique à chaque langue.

La structure applicative des unités d'information est indépendante, et en fait généralement très différente de la structure applicative des phrases correspondantes (cette dernière définissant la structure d'interprétation des phrases non pas dans le sous-langage mais dans une langue). Il est donc important de souligner que c'est un système de règles très simples qui permet de linéariser directement les structures applicatives des unités d'information en phrases.

La représentation proposée fournit un point de vue nouveau sur la relation entre les structures applicatives sur les catégories d'information et les contraintes de concaténation sur les mots. Grâce aux contraintes sur l'ordre des arguments des catégories informatives, la grammaire informative permet de reconnaître le contenu informatif des phrases sans passer par une reconnaissance des phrases en tant que telles. Je montre au chapitre 7 de l'ouvrage cité précédemment comment la structure applicative de la grammaire informative peut être utilisée comme « pattern matcher » directement sur les phrases. Une autre méthode consiste à utiliser des transformations linguistiques pour faire figurer les mots qui sont représentés par des catégories informatives dans l'ordre où s'appliquent ces catégories mais cette méthode est beaucoup plus difficile à automatiser, si tant est qu'elle puisse l'être, d'une part parce qu'elle fait intervenir plusieurs étapes intermédiaires supplémentaires et d'autre part parce que les transformations sont soumises à des conditions d'application lexicales nombreuses et particulières à chaque langue alors que l'utilisation de la structure applicative des unités d'information comme « pattern matcher » sur les phrases est indépendante de la langue dans laquelle est formulée le sous-langage en question.

Une conséquence pratique de cette situation est que le problème de la traduction automatique peut être reconsidéré sous un jour nouveau. Plutôt que de traduire les phrases d'un texte du sous-langage formulées dans une langue A dans une langue B, on peut utiliser la grammaire informative du sous-langage,

commune par construction aux différentes langues exprimant ce sous-langage, pour représenter le texte formulé dans A et produire un texte de B. On utilise le dictionnaire de B avec son organisation propre, ce qui permet d'une part de produire un texte authentique de B ayant le même contenu informatif que celui de A sans en être nécessairement une traduction littérale. D'autre part, l'organisation du dictionnaire de B permet de linéariser les expressions lexicales correspondant aux catégories informatives selon les contraintes propres à B. Cette méthode est limitée aux sous-langages.

La grammaire des unités d'information peut également être utilisée pour reconstruire automatiquement des informations implicites, pertinentes au sous-langage, telles qu'ellipses ou antécédents de relations de coréférence alors que ces reconstructions ne sont en général pas entièrement automatisables en employant les méthodes qui s'appliquent à une langue prise dans son ensemble ([FIS] chap. 7).

La description harrissienne des sous-langages permet de prendre en considération une interaction entre description et objet décrit. Sa représentation met en évidence que si l'existence des entités initiales peut être dégagée par des expériences ou observations dans un univers de discours donné, c'est l'introduction de termes désignant des relations entre ces entités qui détermine une certaine représentation de la connaissance ou plutôt la construction d'un objet auquel on confère un statut de réalité, au moins temporairement, et qui est liée aux termes dans lesquels on l'exprime. L'apparition d'un problème dans l'univers de discours en question, qui met en jeu la cohérence de cette réalité, donc l'existence même de cet objet, est directement liée à la formulation de relations entre les primitives. C'est l'introduction de termes désignant de nouvelles relations, affinant en les subdivisant ou regroupant des catégories initiales, qui permet de résoudre le problème de cohérence en rendant possible l'émergence d'une nouvelle situation d'observation. Dans l'exemple traité par Harris, la question était : certains chercheurs observent que X est responsable de Z tandis que d'autres observent que Y est responsable de Z. La solution a été trouvée quand on a pu formuler une relation entre X et Y ainsi qu'une catégorie V les regroupant, permettant de dire : X et Y sont deux stades de V qui est responsable de Z.

Ce point de vue peut être mis en parallèle avec la façon dont apparaissent les paradoxes en mathématiques et les solutions apportées pour les dépasser.

Les paradoxes en mathématique sont parfois perçus comme des paradoxes pour la Raison. Ils ne sont pourtant tels que si on considère que les termes dans lesquels ils sont formulés appartiennent à une sémantique universelle et non à la sémantique d'un sous-langage particulier dont la connaissance est susceptible de changements. Ainsi, les langues naturelles nous permettent de formuler des contradictions dans le sous-langage des mathématiques, non pas parce que les langues naturelles sont vagues et incohérentes par nature, comme le pensent certains logiciens, mais simplement parce que la sémantique des sous-langages mathématiques en question est insuffisamment précise. Par exemple, Gödel a

montré que l'hypothèse du continu est compatible avec l'axiomatisation de Zermelo-Frankel pour la théorie des ensembles. Mais Cohen a également montré que la négation de cette hypothèse est compatible avec l'axiomatisation de Zermelo-Frankel. Il serait absurde d'en conclure que la phrase qui formule l'hypothèse du continu n'a pas le droit d'exister en tant que phrase. En revanche, le mathématicien doit reformuler les problèmes de fondation de la théorie des ensembles tout comme les biologistes ont dû reformuler les termes dans lesquels ils raisonnaient, donc leurs raisonnements, pour tenir compte de nouvelles données expérimentales. En mathématiques où la réalité d'objets, comme la notion d'ensemble, est problématique, Feferman remarque que les différents axiomes « plausibles » qui ont été proposés depuis le paradoxe de Cohen ne permettent pas de conclure la question de l'hypothèse du continu et que cette question restera probablement sans réponse parce qu'elle n'est pas définie.

On conclura ce paragraphe par une remarque sur le paradoxe du menteur. Sous sa forme la plus concise, il peut s'exprimer par (1) considérée comme une proposition :

(1) *Je mens.*

On ne peut pas assigner de valeur de vérité à (1) dans un calcul de prédicats classique. Par contre, comme le propose Tarski, on peut éviter d'avoir à assigner une valeur de vérité à (1) en définissant une stratification de langages ayant chacun son prédicat de vérité et en interdisant d'appliquer un prédicat de vérité à une proposition appartenant au même niveau de langage. Comme phrase, (1) est interprétable par un procédé de stratification des temps de l'assertion et de l'action de mentir, interne à la formulation en langue naturelle. Le temps où j'asserte (1) et le temps auquel renvoie mon action de mentir ne sont pas identifiés. Il n'est pas fortuit que le verbe *mentir* ne soit pas un verbe à complétive.

6. La structure informative des expressions bien formées dans une langue considérée dans son ensemble

Les oppositions décrites au paragraphe 4, structures syntaxiques et structures informatives, recouvrent en partie l'ancienne opposition, importante au Moyen Âge chez les philosophes et les grammairiens, entre Réalistes et Nominalistes⁶. La démarche de Harris se caractérise notamment par son

6. C'est Jean Stéphanini qui a attiré mon attention sur cette parenté il y a une dizaine d'années. Dans le recueil d'articles qu'il a présenté avec A. Joly : *La Grammaire Générale. Des modistes aux idéologues*, Presses Universitaires de Lille, 1977, on trouvera des idées intéressantes sur l'histoire de la grammaire et les éclairages qu'elle apporte aux débats actuels. Notamment, J.-C. Chevalier met en parallèle la grammaire de Sanctius, qui accorde une place très importante à l'ellipse, et la grammaire des opérateurs de Harris. La Minerve de Sanctius, traduite par G. Clérico, est également publiée aux Presses universitaires de Lille. M. Covington présente une analyse concise des controverses du Haut-Moyen Âge avec de nombreux extraits des textes originaux en traduction bilingue dans : *Syntactic theory in the High Middle Ages*, Cambridge University Press, 1984. Chacun de ces ouvrages présente une bibliographie sur les domaines concernés.

anti-réalisme et s'apparente au nominalisme, sans toutefois s'y identifier totalement. Son approche cherche à préciser quelle est l'économie informative propre aux langues naturelles. Cette économie a une sémantique, c'est elle que représentent les structures informatives. Pour les « Réalistes », le langage traduit un état mental qui exprime la réalité du monde, éventuellement perçue comme la somme des réalités des Mondes Possibles. Il est donc naturel pour eux d'avoir recours à des entités extra-linguistiques, sémantiques ou cognitives, primitives sur lesquelles s'appuient les analyses linguistiques. C'est le cas dans la philosophie analytique, dans la plupart des approches en Intelligence Artificielle et chez les Générativistes. Considérer le langage comme un véhicule de l'information signifie chez Harris que le langage ne traduit ni n'exprime directement le monde dont parle cette information. La notion de discours ou plus précisément de discours rapporté sur laquelle il revient dans *Language and Information* [LI] précise cette idée. L'information dans le langage ne procède pas des relations qu'ont les entités extra-linguistiques dont elle parle, dans la Réalité ou dans l'univers mental du locuteur. Pour Harris, l'information procède de relations entre les noms (les mots ou expressions lexicales) mis en jeu par une langue. Il s'agit essentiellement des propriétés distributionnelles c'est-à-dire les propriétés de restrictions de sélection lexicales et l'ordre applicatif qui doit permettre de définir simultanément les structures d'insertion lexicales et les structures d'interprétation des discours. Pour Harris, le sens des mots résulte donc de leurs propriétés distributionnelles et non l'inverse. Cette conception me semble indissociable de l'hypothèse qu'il existe une représentation commune des énoncés et de leur dérivation.

Dans [LI], Harris précise sa conception de la signification dans une langue en la comparant à sa notion d'information. Les lignes qui suivent tentent de résumer sa position. La signification (meaning) est définie par rapport à l'information : « On ne peut pas chercher leur structure indépendamment, puis la comparer ou la corrélérer avec la structure d'une langue. Mais on peut se demander comment les mots portent une signification et comment les phrases portent une information » ([LI], p. 56). La signification d'un mot change synchroniquement et diachroniquement, avec sa sélection d'arguments si c'est un opérateur, avec les opérateurs qui le sélectionnent si c'est un argument élémentaire. Le changement de la signification d'un mot peut s'analyser à partir des extensions ou limitations de ses emplois, c'est-à-dire des restrictions de sélection lexicale. Le sens seul d'un mot, si tant est qu'il puisse se définir indépendamment des emplois lexicaux particuliers, ne permet pas de prévoir les extensions ou restrictions qui caractérisent le changement de signification de ce mot. Harris donne l'exemple du verbe anglais *sail* (*naviguer*) [LI] p. 63), qu'on a employé avec *Zeppelin* plutôt que *float*, *fly*, *soar*, etc. Le sens seul de *sail* n'aurait pas permis de prévoir cet emploi. (Notons qu'en français, on parle de la navigation aérienne et du personnel navigant et non du personnel volant, pour des raisons probablement plus proches du contexte socio-économique des langues française et anglaise que du concept intrinsèque de navigation.) Harris indique que l'indo-européaniste Henri Hoenigswald a montré que l'extension

d'un mot à de nouveaux emplois est déterminée non par le sens de ce mot mais par la « niche » qu'il occupe dans le lexique, c'est-à-dire la façon dont les restrictions de sélection et contraintes syntaxiques du mot à un moment donné sont reliées aux restrictions de sélection et contraintes syntaxiques d'autres mots comparables. Aller du sens d'un mot à l'analyse de ces emplois n'est pas une tâche pratique accessible alors que la démarche inverse se révèle fructueuse.

Ayant introduit la notion de signification des mots (word meaning) comme étant définie principalement par leurs propriétés de combinaisons lexicales, Harris définit la notion plus abstraite de sens (sense). Cette notion définie à partir de celle de signification n'a qu'un intérêt secondaire. « La signification des mots est probablement mieux caractérisée par leur sens, c'est-à-dire les *propriétés* communes à leur référents possibles, que par leur dénotation c'est-à-dire leur extension ou ensemble de référents. Cependant, ces propriétés ne peuvent être suffisamment spécifiées pour constituer une différenciation adéquate des significations des mots. Cette signification des mots est mieux décrite par l'étude de leur distributions » ([LI], p. 64). Ainsi des mots comme *licorne* ou *centaure* ont un sens qui pose problème aux philosophes du langage parce qu'ils ont des référents dont l'extension est identiquement nulle dans la réalité. Pourtant ils ont des significations qui ne sont ni nulles ni indistinguables. Chacun d'eux peut occuper sous une sélection d'opérateurs différente et raisonnablement spécifiable. On voit que les notions de signification et de sens chez Harris ne recouvrent pas les notions de sens généralement utilisées en grammaire générative, en Intelligence Artificielle ou en philosophie du langage.

La signification des phrases ne résulte pas seulement de la signification des mots qui la constituent. Il est admis que leur signification résulte également d'un apport des fonctions grammaticales. Dans la mesure où Harris tente d'analyser les fonctions grammaticales à partir de deux mécanismes : les reconstructions d'éléments lexicaux implicites et l'application ordonnée des éléments lexicaux explicites ou reconstruits, la signification dans une langue peut s'analyser en termes des propriétés distributionnelles et de l'ordre applicatif des éléments lexicaux. La signification dans une langue peut s'exprimer dans cette langue, elle est également relative à cette langue. Dans cette conception, il n'existe pas de composante abstraite universelle à partir de laquelle on pourrait obtenir la sémantique des langues naturelles. On peut se demander s'il ne serait pas possible d'identifier la sémantique des langues naturelles et la réunion de toutes les grammaires informatives de sous-langages. En dehors d'objectifs pratiques, cette conception serait réductrice dans la mesure où elle ne tiendrait pas compte du fait que la création de nouveaux sous-langages, par l'emploi approprié de restrictions de sélection particulières pour des mots du langage ordinaire, est une propriété ouverte du langage. La sémantique d'une langue concerne l'économie des moyens mis en œuvre par cette langue pour véhiculer des informations. Elle ne doit pas être confondue avec la sémantique d'un sous-langage, qui concerne les propriétés de l'information dans ce sous-langage.

7. Le problème d'une définition de l'assertabilité pour l'analyse d'une langue dans son ensemble

On ne reviendra pas ici sur la distinction entre le problème de l'assertabilité dans les sous-langages et dans une langue considérée dans son ensemble. Les contraintes de composition d'éléments lexicaux dans les séquences de mots susceptibles d'être interprétées ne se réduisent pas aux structures de concaténation (i.e. l'ordre linéaire des éléments lexicaux dans les énoncés) et aux compatibilités lexicales qui régissent ce qu'on appelle habituellement l'acceptabilité. Pour Harris, d'autres contraintes sont mises en jeu dans la composition des expressions lexicales, celles qui régissent l'assertabilité. Par exemple, (1) et (2) sont assertables et on peut leur assigner un type de discours mais elles ont une vraisemblance d'occurrence faible, alors que (3) est inassertable et qu'on ne peut lui assigner un type de discours :

- (1) *J'ai mangé une théière.*
- (2) *L'immeuble s'est effondré pendant des heures.*
La vitre de la fenêtre s'est brisée pendant des heures.
La voiture a explosé pendant des heures.
- (3) *J'ai mangé que Luc était venu.*

On remarquera que la relation entre les éléments lexicaux qui donne leur caractère déviant aux phrases (2), et qui met en jeu l'aspect de leur verbe, dépend de la sélection d'argument de ce dernier. Les phrases (4) sont naturelles :

- (4) *La croûte terrestre de cette région s'est effondrée pendant des heures.*
La surface de glace du fleuve s'est brisée pendant des heures.
La super nova 779437 a explosé pendant des heures.

et montre l'intérêt d'une représentation applicative de l'insertion lexicale.

La question de l'assertabilité ne peut être posée sans faire intervenir la question du choix de l'unité d'analyse. En retour, la question de l'unité d'analyse est une résultante de l'ensemble d'une conception grammaticale⁷. Chacune des trois grandes étapes de l'œuvre de Harris propose une unité d'analyse qui s'apparente à une notion de phrase élémentaire, chaque fois précisée. L'analyse transformationnelle dépasse le cadre de l'analyse en chaîne dans la mesure où elle permet de ramener les « extensions » de l'analyse en chaîne à de véritables combinaisons phrastiques de phrases de base.

La notion de phrase que définissent les décompositions transformationnelles

7. Il existe des analyses très intéressantes de cette question pour différentes traditions grammaticales, voir en particulier *Langages* n° 92 (décembre 1988) sur l'emploi des parties du discours dans l'histoire de la grammaire. Sur le rôle des parties du discours et la notion de phrase issue de la proposition complète d'Aristote, on consultera également : Marc Baratin, *La Naissance de la syntaxe à Rome*, Les Éditions de Minuit, 1989. Dans la tradition des grammairiens indiens, l'ouvrage de Louis Renou : *Terminologie grammaticale du sanskrit*, Champion, 1957, fait le point sur l'emploi de la terminologie dans différentes écoles de grammairiens indiens. Louis Renou a également traduit la grammaire de Panini, diffusée à la librairie A. Maisonneuve. On remarquera que dans cette tradition, les catégories d'analyse ne sont jamais définies par rapport à une notion de phrase (contrairement aux « parties » du « discours » dans la tradition occidentale). La notion de phrase semble d'ailleurs ne correspondre qu'à un usage métrique.

est trop restrictive pour cerner la notion d'assertabilité dans la mesure où elle ne peut rendre compte de relations qui prennent place dans le cadre de l'énoncé, comme certaines ellipses anaphoriques. Par exemple, hors contexte, (5) n'est pas assertable mais elle le devient dans (6) :

(5) *Les taches d'huile sont difficiles.*

(6) *Jean a passé la matinée à enlever les taches d'huile du canapé.*

Les taches d'huile sont difficiles Ø.

La notion de discours, qui tient compte de certaines des dépendances mettant en jeu l'énoncé, est proposée dans [GEMP] mais je montrerai dans mon deuxième article pourquoi elle n'est pas suffisante. Auparavant, il faut revenir sur plusieurs des inconvénients de la notion de phrase dans le cadre de l'analyse transformationnelle.

La notion de phrase simple qui ressort de l'analyse transformationnelle a un deuxième inconvénient pour rendre compte de l'assertabilité, qui provient de la définition même des transformations. Les transformations sont des relations d'équivalence qui permettent d'apparier des phrases en préservant leur degré d'acceptabilité à une constante sémantique près. Or les degrés d'acceptabilité sont relatifs à un univers de discours, donc les transformations aussi. En particulier, l'analyse des textes scientifiques montre que les relations de paraphrase donnant lieu à une même représentation de leur structure informative s'appliquent à des familles de phrases qui n'ont pas de telles relations en dehors des relations de restrictions particulières à ce sous-langage, c'est-à-dire dans une langue considérée dans son ensemble ([FIS], chap. 7). Dans la mesure où le langage ne se réduit pas à l'ensemble des sous-langages existant à un moment donné (la possibilité de création de sous-langages ou d'extensions de sous-langages existants étant ouverte), il semble restrictif de ne définir qu'une notion de phrase relative à celle de sous-langage. Ce problème me semble être un des plus sérieux des grammaires transformationnelles.

Dans la conclusion de « Transformational Theory », Harris indique à quel point sa conception des transformations est liée à son hypothèse du caractère finitiste de l'analyse des langues : « Si deux phrases sont liées par une relation transformationnelle, leur différence peut toujours être formulée comme une somme de différences élémentaires. » Cependant, depuis une vingtaine d'années, les travaux du L.A.D.L. ont montré à quel point les transformations dépendent de conditions d'application lexicales complexes. À ce problème s'ajoute celui de la notion d'entité lexicale et toute la difficile question des expressions figées ou partiellement figées (voir M. Gross « Une classification des phrases figées du Français », *Revue Québécoise de linguistique*, Vol. 11, n° 2, Presses de l'université du Québec). Ces faits, qu'on ne peut que mentionner ici, montrent que l'hypothèse du caractère finitiste des transformations n'a pas le caractère élémentaire supposé au moment de leur introduction.

Dans la grammaire des opérateurs, l'unité d'analyse, le discours, ne correspond pas exactement à la notion de phrase. L'ensemble d'un énoncé est un

discours, constitué de combinaisons de discours. De plus, un discours ne comporte pas obligatoirement une morphologie verbale conjuguée, par exemple : *la démonstration de Jean de ce théorème* est un discours. Le discours ne dépend pas, en principe, de la notion d'acceptabilité⁸. Elle est strictement syntaxique et peut être définie, pour ce qui en fait son essence aux yeux de Harris, comme une forme de saturation syntaxique : un discours est une combinaison de mots dont toutes les restrictions de sélection syntaxiques (pas nécessairement les restrictions de sélection lexicales qui sont relatives aux sous-langage) sont satisfaites. Ce qu'on peut reformuler comme : un discours est une combinaison d'opérateurs lexicaux dont les types se réduisent à un type de discours après les reconstructions linguistiques et avant l'application des règles de linéarisation et ajustements morphologiques. De ce point de vue, elle apparaît comme une sorte de rivale antagoniste de la catégorie *λογος* d'Aristote au sens de l'*Organon* (pas au sens de la *Poétique*) et de l'*oratio* sous ses différentes formes issues du *λογος* de l'*Organon* ainsi que de bon nombre de notions de phrase et d'énoncé utilisées dans les théories linguistiques, puisque la caractéristique de son interprétation n'est pas d'être une « pensée complète » ou une forme logique définie par les conditions dans lesquelles on peut lui assigner une valeur de vérité.

Cependant, cette notion de discours ne permet toujours pas de caractériser l'assertabilité pour l'ensemble d'une langue et cette question met en jeu des problèmes empiriques de natures différentes dont quelques-uns seront examinés dans mon deuxième article.

En guise de conclusion, parmi les différentes questions positives que posent les grammaires de Harris, il y en a deux qui m'ont paru particulièrement intéressantes :

- Peut-on intégrer la métalangue d'une langue dans la représentation de cette langue en construisant les catégories de description pertinentes ?
- Peut-on caractériser l'assertabilité pour une langue dans son ensemble, considérée indépendamment des univers de discours où elle est employée ?

8. En fait la notion d'acceptabilité intervient indirectement puisqu'on l'utilise pour définir les réductions « appropriées ». Un élément lexical subit une réduction « appropriée » quand il est lexicalement indéfini (par exemple *quelqu'un*, *quelque chose*) ou que sa « vraisemblance d'occurrence », son acceptabilité, est particulièrement forte par rapport à son opérateur ou ses arguments. On a donc besoin des traits de sélection lexicale des phrases acceptables, au sens de Maurice Gross, qu'il faudrait associer aux opérateurs pour reconstruire les structures d'insertion lexicale, qui rappellons-le, sont les structures d'interprétation. Par exemple, hors contexte, la phrase :

(1) *Marc a bu la bouteille.*

n'a pas la structure : *a bu* (*Marc, la bouteille*) quoique cette structure ait bien un statut de discours, qui correspondrait à une phrase déviante hors contexte mais qui serait naturelle dans : *Max a bu la bouteille en chocolat qui avait fondu dans ses mains*. Hors contexte, (1) doit avoir la structure applicative de (2) dont elle est une forme réduite :

(2) *Marc a bu le contenu de la bouteille.*

RÉFÉRENCES

- BARATIN, M., *La Naissance de la syntaxe à Rome*, Les Éditions de Minuit, 1989.
- BENVENISTE, E., *Le vocabulaire des institutions indo-européennes*, éditions de Minuit, 1969.
- BOONS, J. P. ; GUILLET, A. ; LECLÈRE, C., *La structure des phrases simples en Français* Droz, 1976.
- CHURCH, A., *The Calculi of Lambda-Conversion*, Annals of Mathematics Studies 6, Princeton University Press, 1941.
- COLOMBAT, B. ed., *Les parties du discours*, *Langages* 92, décembre 88.
- COVINGTON, M., *Syntactic theory in the High Middle Ages*, Cambridge University Press, 1984.
- DALADIER, A., « Some arguments against NP and WH- movement rules », *N.E.L.S. 9 Conference*, Cuny Grad. Center, New-York, 1979.
- DALADIER, A., « Interpretations in natural languages without truth values nor paradoxes », *Semantics* Vol. 10, 1988.
- GIRARD, J. Y., « La logique linéaire », *Pour la science* (version française de *Scientific American*), Avril 1990.
- GIRARD, J. Y. ; LAFONT, Y. ; TAYLOR, P., *Proofs and Types*, Cambridge University Press, 1989.
- GROSS, M., *Méthodes en syntaxe*, Herman, 1975.
- GROSS, M., « Les bases empiriques de la notion de prédicat sémantique », *Langages* 63, 1981.
- GROSS, M., « Une classification des phrases figées du Français », *Revue Québécoise de linguistique*, Vol. 11, n° 2, Presses de l'université du Québec, 1982.
- GUILLET, A., « Morphologie et syntaxe : quelques exemples d'interactions », *Recherches Linguistiques* 3, Université Paris 8, 1975.
- GUILLET, A. et LECLÈRE, C., *La structure des phrases simples en français. Verbes à complément direct et verbes à complément locatif*, Droz, Genève, 1990.
- HEYTING, A., *Intuitionism, an introduction*, North-Holland, 1956.
- HINDLEY, R. et SELDIN, J., *Introduction to Combinators and λ -Calculus*, Cambridge University Press.
- KRIPKE, S., « Outlines of a Theory of Truth », *The Journal of Philosophy*, 1975.
- LAMBECK, J., « On the calculus of syntactic types », *Twelfth Symposium on Applied Mathematics* (R. Jakobson ed.), A.M.S., Providence, 1961.
- MEUNIER, A., « Sur les bases syntaxiques de la morphologie dérivationnelle, *Linguisticae Investigationes*, Vol. 1, Benjamins, 1977.
- PEANO, G., « Le definizioni in matematica », 1921 (repris dans : *Selected works of G. Peano*, Kennedy ed., U. of Toronto Press, 1973).
- RENOU, L., *Terminologie grammaticale du sanskrit*, Champion, 1957.
- STÉPHANINI, J. et JOLY, A., *La Grammaire Générale. Des modistes aux idéologues*, Presses Universitaires de Lille, 1977.
- TROELSTRA, A. S. et VAN DALEN, D., *Constructivism in Mathematics. An Introduction*, Vol. 1, North-Holland, 1988.
- VAN HEIJENOORT, J., *From Frege to Gödel : A Source Book in Mathematical Logic, 1879-1931*, Harward University Press.