Le verbe en HPSG : nucleus d'un système lexicalisé

Verbs in HPSG : Nucleus of a Lexicalized Theory

Philippe Blache

Édition électronique

URL : <http://journals.openedition.org/praxematique/2266>

DOI : 10.4000/praxematique.2266

ISSN : 2111-5044

Éditeur

Presses universitaires de la Méditerranée

Édition imprimée

Date de publication : 1 janvier 1994

Pagination : 37-56

ISSN : 0765-4944

Référence électronique

Philippe Blache, « Le verbe en HPSG : nucleus d'un système lexicalisé », *Cahiers de praxématique* [En ligne], 22 | 1994, document 2, mis en ligne le 01 janvier 2013, consulté le 08 septembre 2020. URL : <http://journals.openedition.org/praxematique/2266> ; DOI : <https://doi.org/10.4000/praxematique.2266>

Ce document a été généré automatiquement le 8 septembre 2020.

Tous droits réservés

Le verbe en HPSG : nucleus d'un système lexicalisé

Verbs in HPSG : Nucleus of a Lexicalized Theory

Philippe Blache

- 1 La tendance actuelle des formalismes linguistiques, et notamment ceux plus particulièrement destinés au traitement automatique, est à la lexicalisation. Ce mouvement existe évidemment depuis longtemps, les grammaires lexicales fonctionnelles de Kaplan et Bresnan en étant une instance emblématique. Mais il s'est généralisé vers la fin des années 80, qui ont également vu les théories basées sur l'unification devenir l'outil de prédilection des informaticiens linguistes. Prenons quelques exemples de théories récentes. La représentation de données syntaxiques au niveau du lexique est d'une importance capitale pour les grammaires d'arbres adjoints lexicalisées, le principe d'adjonction étant une forme d'insertion lexicale. Les HPSG¹ à leur tour, en s'appuyant sur des propriétés déjà présentes dans GPSG², accordent une place centrale au lexique : rôle des règles lexicales, importance du principe de sous-catégorisation, etc... Jusqu'à GB³ qui, dans sa dernière évolution (l'approche minimaliste), intègre cette tendance. On observe ainsi, au moins sur ce terrain, un rapprochement considérable de ces théories qui pourraient peut-être s'avérer moins radicalement différentes qu'on aurait pu le croire. Mais il est vrai que cette dernière question n'est en fin de compte que marginale, l'essentiel étant de disposer d'outils formels productifs.
- 2 Ces quelques remarques d'ordre général peuvent sembler sans lien direct avec le problème qui nous préoccupe dans ce volume : le verbe. Mais ce n'est qu'une première vue : repousser les informations syntaxiques au niveau du lexique revient à accorder dans la représentation une place centrale aux catégories majeures. Et parmi ces catégories, le verbe occupe une place éminemment stratégique : quelle que soit la théorie considérée, le verbe sera un élément déterminant dans la représentation des informations syntaxiques (et sémantique, nous allons le voir) contenues dans une grammaire.

- 3 Nous avons donc choisi d'illustrer le phénomène de lexicalisation par la description de la représentation lexicale du verbe et de son insertion pendant l'analyse dans le cadre de la théorie des HPSG. Nous verrons ainsi comment la description des phénomènes de rection verbale peut être généralisée à la représentation des autres constituants. Nous étudierons en particulier le problème de la sous-catégorisation en tant que phénomène emblématique des mécanismes d'analyses (ou principes) de cette théorie. Nous verrons au passage comment cette approche permet une représentation relativement naturelle des relations pouvant exister entre la syntaxe et la sémantique. Finalement, nous aborderons dans une dernière partie les problèmes d'implantation spécifiquement liés aux notions abordées.

1. Deux mots sur les HPSG

- 4 Une rapide présentation des principales propriétés des HPSG s'impose évidemment avant d'entrer dans l'analyse détaillée des particularités du verbe. Cette théorie est essentiellement issue des GPSG et fait à ce titre partie des grammaires dites d'unification. Comme dans toutes les théories appartenant à cette famille, les propriétés syntactico-sémantiques y sont représentés sous forme de traits (i.e. des couples attribut/valeur). Un de ses aspects caractéristiques réside dans la description des objets linguistiques manipulés (également appelés signes) sous forme de structures de traits. Mais à la différence des GPSG, la notion de hiérarchie syntagmatique est elle-même considérée en tant que propriété syntaxique et à ce titre représentée par un trait particulier. Ainsi, le résultat d'une analyse syntaxique en HPSG se présentera sous la forme d'une structure de traits générale contenant à la fois les caractéristiques inhérentes à un signe, mais également une représentation de la structure hiérarchique.
- 5 Entrons un peu dans les détails. Tout objet linguistique manipulé en HPSG, de quelque niveau qu'il soit (lexical, syntagmatique ou propositionnel) correspond à une structure de trait appelée *SIGNE*⁴. De plus, nous verrons ceci plus précisément pour le verbe dans la section suivante, toutes les structures de traits sont typées (cf. section 2). Il est ainsi possible de limiter la sous-spécification dans la mesure où des sous-ensembles de valeurs pertinentes (i.e. d'un type donné) seront ainsi spécifiés pour un trait donné : il est en effet peu utile d'autoriser en toute généralité l'instanciation d'un trait *VFORM*⁵ pour une catégorie nominale. Il est important à ce propos de signaler que les structures de traits typées, et plus généralement la logique attribut-valeur, constituent la base théorique du formalisme HPSG. On en trouvera une description détaillée dans plusieurs ouvrages (Johnson, 1988 ou Carpenter, 1992). La figure 1 représente, sans tenir compte de la notion de typage, la structure d'un signe. Signalons simplement d'une part que cette structure est incomplète (n'y figurent notamment pas les traits utiles à la description de la quantification) et d'autre part que cette structure peut à tout moment évoluer à condition bien entendu d'en préserver la cohérence générale. On peut par exemple signaler plusieurs travaux portant sur l'intégration de descriptions phonologiques dans le trait *PHON* indiqué dans le signe canonique (cf. Bird, 1993 ou Blache et Angoujard, 1993). Cette propriété est extrêmement intéressante et fait de HPSG une théorie linguistique ouverte, évolutive et donc vivante.

Figure 1 : Structure canonique de traits en HPSG

- 6 En ce qui concerne le formalisme grammatical, il s'appuie comme en GPSG sur une représentation distincte des informations de hiérarchie et d'ordre (formalisme DI/PL⁶) Mais à la différence de cette théorie, HPSG utilise des schémas de DI-règles très généraux : une DI-règle au sens GPSG est en fait une instance d'un des schémas proposés. Les contraintes d'ordre sont quant à elles classiquement représentées à l'aide de contraintes de préséance linéaire. Mais revenons aux schémas de DI-règles. Cette caractéristique est directement liée à la lexicalisation : il n'est en effet pas possible d'utiliser des schémas de hiérarchie sans autre information. C'est la valeur du trait de sous-catégorisation qui permettra de compléter le schéma général de bonne formation fourni par ces règles. Cette caractéristique permet de plus d'éviter une certaine redondance d'informations que l'on trouve dans la plupart des autres théories. En GPSG, par exemple, la sous-catégorisation est à la fois présente au niveau lexical (qui doit comporter la référence de la classe de valence) et au niveau grammatical dans les DI-règles. En HPSG, ces informations sont uniquement comprises au niveau lexical. La figure 2 donne l'exemple du schéma de DI-règle n° 1 (cf. Pollard et Sag 1994 :30) : il indique qu'un signe syntagmatique saturé ayant un fils TÊTE syntagmatique et un fils COMPLEMENT unique est une structure bien formée.

Figure 2 : Exemple de schéma DI

- 7 L'analyse syntaxique en HPSG consistera donc à construire ces structures de traits en instanciant les valeurs des traits tout en vérifiant leur cohérence. Ces deux processus (vérification et instanciation) sont essentiellement pris en charge par les *Principes*

d'instanciation qui vont gérer la propagation des valeurs dans la structure tout en vérifiant un certain nombre de contraintes. Ils sont en HPSG en nombre limité (HPSG94⁷ en propose 12) et ont une vocation universelle (donc très généraux). Les principaux sont le principe des traits de tête, le principe de sous-catégorisation, le principe des traits non-locaux ou le principe sémantique. Leur mécanisme de base est le *partage de structure* qui, lui, reposera sur l'unification. Nous n'entrerons pas ici dans une présentation détaillée de leur fonctionnement, en nous contentant d'en aborder quelques uns dans le cours de cet article.

2. Les types

- 8 Les structures de traits, on l'a vu, sont typées. Cela signifie que chaque signe doit faire partie d'une famille partageant un certain nombre de propriétés notamment structurelles. Pratiquement, un type est associé à un ensemble de traits le spécifiant, et réciproquement, un trait sera réputé appartenir à un type donné. Les traits devant être hiérarchisés, nous aurons de façon parallèle une hiérarchie de types (un treillis) organisant l'ensemble des descriptions. La typologie associée au verbe est bien entendu la plus détaillée et permet ainsi une description très précise des caractéristiques du constituant analysé. Elle est essentiellement constituée par deux *treillis de types* : la hiérarchie des traits de tête et la hiérarchie de sous-catégorisation. Mais ces deux treillis auront des utilisations très différentes : le premier correspondra effectivement à des traits spécifiquement verbaux tandis que le second pourra être utilisé (au moins en théorie) pour toute autre catégorie. On notera cependant que la sous-catégorisation reste éminemment emblématique d'un constituant verbal.

Figure 3 : Classification des verbes en fonction de leurs traits de tête

- 9 Le treillis de la figure 3 précise l'organisation des traits AUX et VFORM par leurs types respectifs *aux* et *vform*. Le premier indique simplement si le verbe est un auxiliaire et dans ce cas s'il est modal, copule, etc ... Le second précise la caractéristique de la forme analysée. Cette répartition est tout à fait classique.

Figure 4 : Types de sous-catégorisation

- 10 La figure 4 décrit quant à elle les types de sous-catégorisation. On distinguera tout d'abord les types nominaux des autres. On remarquera au passage que les types *saturé* et *commun* sont exclusivement nominaux tandis que les autres sont verbaux (HPSG considère en effet que le verbe sous-catégorise son sujet). Les type verbaux permettront ainsi d'établir une distinction de construction très fine.

3. Représentation lexicale

- 11 Prenons l'exemple d'une entrée lexicale en HPSG. La figure 5 décrit une structure de trait (partielle) correspondant à l'entrée du verbe *donner*. Cette structure est partielle mais comprend néanmoins les propriétés essentielles illustrant le rôle du lexique en HPSG. Passons rapidement sur les traits de tête qui, nous l'avons vu, sont essentiellement destinés à décrire les caractéristiques (le type syntaxique pourrait-on dire) d'un signe. Plus intéressant est en revanche le trait de sous-catégorisation. Celui-ci a pour valeur la liste des catégories constituant la valence du verbe. Dans notre cas, il y a sous-catégorisation d'un SN nominatif, d'un SN accusatif et d'un SP. Il s'agit donc du schéma ditransitif de *donner*.

Figure 5 : Entrée lexicale du verbe *donner*

- 12 Ces notions sont évidemment classiques et nous les retrouvons d'une façon ou d'une autre dans les autres formalismes syntaxiques. Leur présence au niveau lexical permet cependant d'éviter leur représentation dans des règles syntagmatiques (comme en GPSG par exemple). Nous pourrions ainsi, et il s'agit d'une caractéristique formelle tout à fait importante, utiliser non plus des règles syntagmatiques, ni même des DI-règles, mais directement des schémas de DI-règles (cf. paragraphe précédent).
- 13 Le principe de sous-catégorisation permet donc de vérifier la structure valencielle en imposant la saturation du signe (i. e. la réalisation des signes sous-catégorisés). Il indique que la valeur du trait SOUSCATÉGORISATION d'une tête est la concaténation du trait SOUSCATÉGORISATION du syntagme avec la liste des valeurs SYNSEM des compléments. Cette vérification reposera également sur l'unification, comme pour la plupart des autres principes. Il est donc possible (et même souhaitable) d'utiliser ce principe pour propager certaines valeurs de traits ne dépendant d'aucun des deux autres principes majeurs (principe des traits de tête et principe des traits non locaux). Nous pourrions ainsi rendre compte de phénomènes extrêmement différents comme l'accord sujet/verbe ou l'affectation de la fonction syntaxique. Le principe consiste à spécifier les valeurs de traits désirées en les attachant aux constituants sous catégorisés. On remarque ainsi que le premier syntagme nominal est spécifiée comme étant nominatif. Cette valeur du trait CAS sera instanciée correctement au SN précédant le verbe par unification lors de la vérification de sa réalisation. Mais il est également possible de transmettre des liens entre des sous-structures de l'entrée lexicale et le constituant sous-catégorisé grâce à la coindexation. Nous avons vu en effet qu'il était possible

d'indexer les sous-structures d'un signe. On exprime alors par la coindexation le partage de structure : dans ce cas, une même structure devient la valeur de plusieurs traits (de même type, évidemment).

- 14 Le second point intéressant de la sous-catégorisation est directement lié aux verbes. Il s'agit de l'utilisation de la coindexation pour la construction du schéma sémantique (i. e. du schéma relationnel) associé au verbe. Précisons, avant d'entrer dans les détails de cette structure, que nous retrouvons le même traitement de la sémantique pour les noms : ceux-ci sont également considérés comme porteur d'une relation particulière de désignation.
- 15 Chaque verbe est associé à un schéma relationnel spécifiant une relation et des valeurs pour les rôles de cette relation, permettant ainsi une description fonctionnelle de leur structure sémantique. Cette représentation est relativement classique et consiste à décrire la structure prédicative en précisant l'ensemble de ses arguments. Nous obtenons ainsi une structure de traits un peu particulière : la racine de cette structure sera le trait `CONTENU` composé des traits `RESTRICTION` et `INDEX`. Le premier permet de réduire la portée du second en spécifiant certaines propriétés. La valeur du trait `RESTRICTION` correspond en quelque sorte au *frame* associé dans une représentation conceptuelle.
- 16 Mais si cette représentation n'est pas à proprement parler originale, la particularité du traitement de ces aspects sémantiques par HPSG réside dans l'interface syntaxe/sémantique (cf. Pollard, 1989) et dans la représentation des mécanismes associés. En effet, nous voyons dans la figure 5 comment le partage de structure permet de construire dès le niveau lexical ce schéma relationnel. Mais il est également intéressant de remarquer que la spécification d'une catégorie (par exemple pour la sous-catégorisation) se fait entièrement par l'intermédiaire de traits sémantiques qui vont donc dans cette théorie jouer un rôle déterminant au niveau syntaxique. La description des relations entretenues par ces traits sémantiques avec les autres traits de la structure générale est assurée par un principe de propagation, exactement au même niveau que les autres principes d'instanciation plus directement syntaxiques. Ce principe indique (dans sa version simplifiée ne tenant pas compte des quantifications) que :

La valeur du trait `CONTENU` d'un syntagme est partagée avec celle du trait `CONTENU` de sa tête sémantique⁸.

- 17 HPSG propose ainsi un fonctionnement et une description homogène des relations inter-structurelles aussi bien syntaxiques que sémantiques. Une telle représentation influera donc sur certains aspects directement syntaxiques comme l'accord. Le fait de représenter les traits d'accord non plus à l'intérieur des traits de tête mais dans le trait sémantique `CONTENU` permettra de prendre en compte certains aspects particuliers de ce problème (comme le pluriel de politesse en français, voir Pollard et Sag 1994 :101).

4. Les règles lexicales

- 18 Nous allons prendre dans cette partie quelques exemples permettant d'illustrer le rôle joué par la lexicalisation dans la description du verbe. Nous aborderons pour cela la question du *passif* et la représentation des *verbes à montée* (cf. infra). Ces points nous

permettront notamment de présenter plus précisément le rôle des *règles lexicales* en HPSG.

- 19 Depuis les premières descriptions transformationnelles, toute théorie linguistique se doit de fournir une description du passif. Il s'agit en effet d'une des descriptions permettant d'illustrer le mieux les différences entre les approches monostratales et celles utilisant une structure profonde. GPSG propose pour cela une métarègle. HPSG, qui n'utilise pas à proprement parler de métarègles, proposera une règle lexicale. Ces dernières sont destinées à rendre compte de paramètres plus spécifiques à une langue donnée et préservent ainsi, par voie de conséquence, le niveau de généralité des autres principes décrits précédemment. Elles rendent compte de relations existant horizontalement entre différentes classes de mots comme les relations de nominalisation, passivation, flexion, etc ... Pratiquement, une règle lexicale permet de prédire une forme de sortie à partir d'une forme d'entrée, comme par exemple une relation existant entre une racine et une forme fléchie (cf. Pollard et Sag 1987 :209).

$$\left[\begin{array}{l} \text{Phon } [1] \\ \text{Part-Pas } [2] \\ \text{Syn | Loc } \left[\begin{array}{l} \text{Cat | SousCat } \langle \text{[SN]}_4, [\text{SN}]_3, \dots \rangle \\ \text{Contenu } [5] \end{array} \right] \end{array} \right] \\ \rightarrow \left[\begin{array}{l} \text{Phon } f_{pps}([1], [2]) \\ \text{SynLoc } \left[\begin{array}{l} \text{Cat | SousCat } \langle \text{[SN]}_3, \text{SP}[par]_4, \dots \rangle \\ \text{Contenu } [5] \end{array} \right] \end{array} \right]$$

Figure 6 : Règle Lexicale du Passif

- 20 La règle lexicale de passivation, décrite dans la figure 6, fonctionne de la façon suivante : une règle lexicale spécifique calculant le participe (indiquée dans le trait PHON) du verbe gère son aspect morphologique. Le contenu sémantique est quant à lui simplement reporté sans changement. L'action principale de cette règle lexicale portera sur la sous-catégorisation : le SN sujet est éliminé de la liste de sous-catégorisation et son index réaffecté à un SP[par] ajouté en fin de liste. Le SN objet, qui était le second argument de la liste de sous-catégorisation du verbe à l'actif, se retrouve ainsi premier argument⁹ de la nouvelle liste et devient donc sujet de cette forme verbale. Cette règle lexicale s'appliquera bien entendu à la classe des verbes de type *transitif*. Ce type comprend les verbes strictement transitifs (cf. figure 4), mais également les verbes sous-catégorisant d'autres compléments. La figure 7 donne l'exemple de l'application de cette règle à l'entrée verbale « voir ».

$$\left[\begin{array}{l} \text{Phon voir} \\ \text{Part-Pas vu} \\ \text{Syn | Loc } \left[\begin{array}{l} \text{Cat | SousCat } \langle \text{[SN]}_1, [\text{SN}]_2 \rangle \\ \text{Contenu } \left[\begin{array}{l} \text{Rel voir} \\ \text{Agent } [1] \\ \text{Patient } [2] \end{array} \right] \end{array} \right] \end{array} \right] \\ \rightarrow \left[\begin{array}{l} \text{Phon vu} \\ \text{SynLoc } \left[\begin{array}{l} \text{Cat | SousCat } \langle \text{[SN]}_2, \text{SP}[par]_1 \rangle \\ \text{Contenu } \left[\begin{array}{l} \text{Rel voir} \\ \text{Agent } [1] \\ \text{Patient } [2] \end{array} \right] \end{array} \right] \end{array} \right]$$

Figure 7 : Instanciation de la règle lexicale Passif pour le verbe *voir*

- 21 Terminons cette description par la représentation lexicale des *verbes à contrôle obligatoire* (comme *espérer*, *vouloir*, *persuader*) et des *verbes à montée*¹⁰ (comme *pouvoir*,

sembler). Il s'agit comme pour le passif d'un exemple traité par toute nouvelle théorie linguistique essentiellement en vue de se situer par rapport à la théorie du Gouvernement-Liage de Chomsky. La description des phénomènes de contrôle liée à ces deux types de verbes est intéressante car elle illustre parfaitement la capacité qu'ont les HPSG à représenter au niveau du lexique des phénomènes syntaxiques complexes. Ces phénomènes portent notamment sur la description du sujet sous-jacent des infinitives : la question posée est celle de la représentation du fait que le sujet non réalisé de l'infinitive partage un certain nombre de valeurs avec un des constituants régis par le verbe de ce type. Ceci sera évidemment spécifié grâce au partage d'une sous-structure. L'exemple de la figure 8 reprend un des cas décrits par Pollard (cf. Pollard et Sag 1994 :147). On remarque ainsi que le phénomène de contrôle est indiqué par l'intermédiaire du trait de sous-catégorisation : la proposition infinitive comprend à son tour un trait de sous-catégorisation ayant pour valeur un SN coindexé avec un SN sous-catégorisé par le verbe lui-même. On remarquera également une différence d'ordre sémantique entre ces deux verbes : le verbe à contrôle obligatoire (ici *persuade*) affecte un rôle sémantique à tous les constituants qu'il sous-catégorise contrairement au verbe à montée (ici *believe*).

22 Exemple :

- They persuade him to be happy
- They believe him to be happy

23 Il s'agit pour Pollard et Sag d'une des différences essentielles entre ces deux types de verbes.

24 *persuade*

25 *believe*

Figure 8 : Schémas de verbes à montée

5. Implantation

26 Nous n'entrerons pas ici dans les détails des problèmes d'implantation d'un analyseur HPSG (pour cela, se reporter à Blache, 1993). Mais il est cependant intéressant d'aborder les questions liées aux problèmes évoqués dans cet article. Ils sont, rappelons-le, relatifs aux spécificités d'une approche lexicalisée.

27 De nombreux problèmes peuvent ainsi être évoqués. Nous choisissons d'en traiter deux parmi les principaux : insertion lexicale et sous-catégorisation. Ces questions seront abordées à travers les problèmes particuliers de la représentation des connaissances et de l'implantation des mécanismes d'analyse.

5.1 Représentation des connaissances

- 28 Les HPSG, nous l'avons vu, reposent sur une représentation en structures de traits. Il existe de nombreuses façons de les représenter mais qui, du point de vue de l'implantation, auront des conséquences importantes. On abordera en effet différemment le problème selon que l'on utilise un graphe orienté ou une formule logique. Cette dernière solution présente l'avantage d'une représentation proche du support théorique des HPSG que constitue la logique attribut/valeur (décrites dans Carpenter 1992). Elle est de plus relativement naturelle dans le paradigme (classique aujourd'hui pour le traitement automatique du langage naturel) de la programmation en logique. Nous utiliserons ainsi Prolog pour la description des questions d'implantation qui suivent.
- 29 Dans cette approche, un trait est représenté par un prédicat de la forme `trait (index, valeur)` où `index` est un indice localisant la structure référencée dans la structure de traits générale. Une structure de traits peut ainsi être représentée par une conjonction de prédicats de ce type. L'exemple de la figure 9 montre la représentation prédicative d'une structure de traits. On remarquera que toutes les sous-structures sont indexées, ce qui permet donc une représentation directe de leurs chemins d'accès.

Figure 9 : Structure de traits d'un SV

- 30 La représentation prédicative de la structure de trait donnée en figure 9 est la suivante :
- 31 `Phon (1, donne...) ⌈ Syn|Loc (1, 3) ⌈ Cat (3, 4) ⌈ Tete (4, 5) ... Maj (15, N) Vform (5, Tensé) ...`
- 32 Cette représentation peut bien entendu s'implanter de différentes façons. Nous avons choisi de rester aussi proche que possible du cadre formel notamment en s'appuyant sur la notion de satisfaction de contraintes : une structure de trait, munie d'un certain nombre de propriétés formelles doit, pour être bien formée, satisfaire un certain nombre de contraintes. Ces dernières correspondent en fait aux principes d'instanciation décrits précédemment. Nous avons ainsi besoin d'une fonction d'interprétation associant à un prédicat (représentant un trait) une valeur de vérité. Cette fonction peut être formulée de la façon suivante :

L'interprétation d'un prédicat est vraie si le trait associé appartient effectivement à la structure générale construite. Elle est fausse si ce trait ne peut pas appartenir à cette structure.

- 33 Munis de cette fonction d'interprétation, nous serons alors en mesure de contrôler l'instanciation des traits en évaluant la validité des formules correspondant aux structures. En termes d'implantation, nous intégrons la fonction d'interprétation à la représentation des traits en associant à chaque prédicat un 4-uplet de la forme :

<Trait, Indice, Valeur, Interprétation>

- 34 Une structure de traits, correspondant à une conjonction de prédicats, sera représentée par une liste de ces 4-uplets. Cette représentation présente plusieurs avantages. Tout d'abord concernant le mécanisme d'ajout d'information : l'ajout d'un trait à une structure générale se fait simplement en complétant la liste correspondant au niveau concerné. Il suffit donc de connaître le nom du trait et l'indice de la structure à laquelle il appartient pour accéder à sa valeur si elle existe, mais également pour préparer l'instanciation d'un trait.
- 35 Il est en effet tout à fait possible d'ajouter à la structure un trait ne contenant pas encore de valeur précise ou même un trait pour lequel on ne peut pas encore donner de valeur d'interprétation : c'est le cas par exemple des valeurs par défaut pour lesquelles on ne sait pas au moment de leur instanciation si elles sont ou non cohérentes avec la structure courante. La question de l'insertion lexicale se résume donc à une simple question de manipulation de listes. La consistance des structures construites peut se faire quant à elle de plusieurs façons. Une méthode efficace et élégante consiste à utiliser pour cela des contraintes booléennes (cf. Blache 1993).

5.2 Sous-catégorisation

- 36 La sous-catégorisation, nous l'avons vu, est l'expression de la structure valencielle des verbes et plus généralement du schéma de rection des catégories majeures. La spécification de tels schémas nécessite l'utilisation d'un principe d'instanciation permettant d'assurer la satisfaction de ces contraintes de sous-catégorisation.
- 37 La valeur de ce trait est constituée par un ensemble de catégories. Il s'agit plus précisément d'un ensemble d'indices de catégories. Pour utiliser une métaphore dans le domaine de l'informatique, ces indices constituent des pointeurs vers une structure décrivant un signe.
- 38 La spécification d'une propriété pour un des signes sous-catégorisés entraîne donc son instanciation dans le signe référencé. C'est de cette façon que les cas seront affectés : le premier syntagme nominal spécifié dans le trait SOUSCATÉGORISATION est spécifié comme ayant le trait CAS[Nominatif] au niveau lexical. Lorsque la coindexation est réalisée (i.e. la sous-catégorisation vérifiée), la structure partielle décrite au niveau du lexique s'unifiera avec la structure correspondant au signe lui-même en affectant au passage les traits décrits par sous-catégorisation. On représentera de la même façon la plupart des relations d'accord, en particulier l'accord sujet/verbe.
- 39 Le principe de sous-catégorisation est formulé de la façon (simplifiée) suivante :
- 40 La valeur du trait de sous-catégorisation d'un signe non lexical correspond à la valeur du trait de sous-catégorisation de son fils-tête à laquelle ont été retirés les constituants correspondants à ses fils-compléments.

- 41 Ce principe repose, comme l'insertion lexicale, sur une manipulation de listes. La contrainte correspondante consistera à spécifier que la liste de sous-catégorisation (liste de constituants représentant la valeur de sous-catégorisation) d'une catégorie tête est la concaténation de la liste de sous-catégorisation de sa projection et de l'ensemble des compléments frères.

Conclusion

- 42 Les théories fortement lexicalisées comme les HPSG, les TAG ou les LFG accordent une place centrale au verbe, celui-ci étant considéré comme porteur d'informations syntaxiques (en particulier valencielles) et sémantiques importantes. Les HPSG franchissent une étape supplémentaire en permettant de plus la description des interactions existant entre syntaxe et sémantique, notamment pour ce qui concerne la définition des rôles pragmatiques associés à la structure valencielle par sous-catégorisation.
- 43 D'une façon plus générale, la description du verbe dans le cadre de théories lexicalisées permet une présentation relativement complète de la théorie elle-même.

BIBLIOGRAPHIE

- ABEILLÉ A. (1991) *Une grammaire lexicalisée d'arbres adjoints pour le français*, Thèse de Doctorat, Université Paris 7.
- (1993) *Les nouvelles syntaxes. Grammaires d'unification et analyse du français*, Armand Colin.
- BLACHE P. (1993) « Deux théories adaptées à la TAO : GPSG et HPSG », in *La Traductique*, AUPELF/UREF, Presses de l'Université de Montréal, pp. 228-240.
- (1993) « HPSG as a Constraint Satisfaction Problem », in *proceedings of NLPRS'93*.
- CARPENTER B. (1992) *The Logic of Typed Feature Structures*, Cambridge University Press.
- GAZDAR G., E. KLEIN, G. PULLUM & I. SAG (1985) *Generalized Phrase Structure Grammar*, Blackwell.
- JOHNSON M. (1988) *Attribute-Value Logic and the Theory of Grammar*, CSLI, University of Chicago Press.
- POLLARD C. (1989) « The Syntax-Semantics Interface in a Unification-Based Phrase Structure Grammar », in *proceedings of the Workshop on GPSG and Semantics*, Berlin.
- POLLARD C. & I. SAG (1987) *Information-Based Syntax and Semantics*, CSLI, University of Chicago Press.
- (1994) *Head-Driven Phrase Structure Grammar*, CSLI, University of Chicago Press.
- SHIEBER S. (1986) *An Introduction to Unification-Based Approaches to Grammar*, CSLI, University of Chicago Press.

NOTES

1. Abréviation de Head-driven Phrase Structure Grammars (grammaires syntagmatiques guidées par les têtes).
 2. Abréviation de Generalized Phrase Structure Grammars (grammaires syntagmatiques généralisées).
 3. Abréviation de Government and Binding (gouvernement-liage).
 4. On notera dorénavant les noms de traits en petites capitales.
 5. Même si celui-ci est un trait de tête.
 6. Les règles syntagmatiques classiques contiennent des informations de hiérarchie (rapport partie droite – partie gauche de la règle) et d'ordre (succession des éléments de la partie droite). Le formalisme DI/PL représente de façon séparée ces informations par des règles de domination immédiate et de préséance linéaire.
 7. On note traditionnellement HPSG87 et HPSG94 les deux versions de cette théorie.
 8. Une tête sémantique est l'adjectif s'il existe, sinon la tête (on considère en effet en HPSG que les adjectifs sélectionnent les têtes).
 9. On rappelle que l'ordre des constituants de la liste de sous-catégorisation d'un verbe va de la catégorie la moins oblique (SN sujet) vers la plus oblique (SP).
 10. Respectivement *Equi* et *Raising* en anglais. La description de ces phénomènes semble d'ailleurs plus productive pour l'anglais que pour le français.
-

RÉSUMÉS

Les théories linguistiques utilisées dans le cadre du traitement automatique du langage naturel connaissent aujourd'hui une évolution sensible en direction d'une lexicalisation des informations syntaxiques. C'est en particulier le cas de LFG, bien entendu, mais aussi des TAG ou de HPSG. Nous prendrons ici le cas de cette dernière en illustrant ce phénomène par la représentation du verbe et de ses informations afférentes. Cette catégorie est en effet emblématique dans le sens où elle est le creuset de nombreux phénomènes syntaxiques concernant par exemple la notion de sous-catégorisation ou le problème de l'insertion lexicale.

Lexicalization is a current trend for contemporary linguistic theories used in natural language processing. It is the case for example of TAG, LFG or HPSG. We describe here this phenomenon for HPSG by relying our description on the verb representation in this theory. Indeed, verbal categories illustrate very well the specification of major syntactic phenomena (such as lexical insertion or subcategorization) within a linguistic formal framework.

AUTEUR

PHILIPPE BLACHE

2LC – CNRS Sophia Antipolis pb@llaor.unice.fr