

FACULTÉ DES LETTRES

ÉTUDE TERMINOLOGIQUE ET ANALYSE
DES MODES DE FORMATION DE 50 NOTIONS
SUR LE TRAITEMENT AUTOMATIQUE
DES LANGUES NATURELLES

MARTINE DUBÉ
B.A., C. Sc. Ed., B.A. Ling.

Mémoire présenté
pour l'obtention
du grade maître ès arts (M. A.)
dans le cadre d'une entente entre l'Université Laval
et l'Université du Québec à Chicoutimi

ÉCOLE DES GRADUÉS
UNIVERSITÉ LAVAL

JUIN 1990



Mise en garde/Advice

Afin de rendre accessible au plus grand nombre le résultat des travaux de recherche menés par ses étudiants gradués et dans l'esprit des règles qui régissent le dépôt et la diffusion des mémoires et thèses produits dans cette Institution, **l'Université du Québec à Chicoutimi (UQAC)** est fière de rendre accessible une version complète et gratuite de cette œuvre.

Motivated by a desire to make the results of its graduate students' research accessible to all, and in accordance with the rules governing the acceptance and diffusion of dissertations and theses in this Institution, the **Université du Québec à Chicoutimi (UQAC)** is proud to make a complete version of this work available at no cost to the reader.

L'auteur conserve néanmoins la propriété du droit d'auteur qui protège ce mémoire ou cette thèse. Ni le mémoire ou la thèse ni des extraits substantiels de ceux-ci ne peuvent être imprimés ou autrement reproduits sans son autorisation.

The author retains ownership of the copyright of this dissertation or thesis. Neither the dissertation or thesis, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

RÉSUMÉ

Le présent mémoire comporte deux volets. Tout d'abord, une étude terminologique du traitement automatique des langues naturelles comprenant une présentation de la méthodologie utilisée pour la rédaction des dossiers, un arbre du domaine, un corpus des termes traités et enfin un vocabulaire composé de 50 dossiers de terminologie. Chaque article est constitué de différentes rubriques soit d'une entrée et de sous-entrées, d'un domaine d'emploi, d'une définition, de contextes, d'observations et, à l'occasion, d'illustrations.

Le second volet consiste en une analyse des modes de formation néologiques les plus fréquemment rencontrés dans le corpus en l'occurrence le syntagme et le calque. Une typologie des principaux cas répertoriés et des tableaux statistiques ont été dressés pour chacun d'entre eux en plus d'un commentaire sur les cas particuliers qui ont été relevés.

Martine Dubé, étudiante

Jean-Claude Boulanger, directeur

AVANT-PROPOS

La réalisation de ce mémoire a été possible grâce à la collaboration active de plusieurs intervenants. Tout d'abord, je tiens à remercier mon directeur, Monsieur Jean-Claude Boulanger, pour avoir bien voulu partager sa très grande expérience de la terminologie avec moi et pour ses conseils judicieux.

Je tiens particulièrement à souligner le travail de François Labelle, professeur de linguistique à l'Université du Québec à Chicoutimi, pour sa disponibilité et sa précieuse collaboration tout au long de la rédaction des dossiers.

Également, un merci très spécial à mon ami, René Beaudin, pour sa confiance et son optimisme face à la réussite de ce projet.

Et enfin, merci à Monsieur Yves Saint-Gelais, directeur de la Maîtrise en linguistique à l'Université du Québec à Chicoutimi, pour son appui et sa compréhension depuis le tout début de ce projet.

TABLE DES MATIÈRES

	page
Résumé	ii
Avant-propos.....	iii
Table des matières	iv
Liste des tableaux.....	vi
 INTRODUCTION.....	1
 CHAPITRE 1 PRÉSENTATION MÉTHODOLOGIQUE	3
Introduction	3
A. Phase de recherche.....	4
1. Documentation	4
2. Arbre du domaine.....	5
B. Phase de traitement terminologique	6
1. Corpus	6
2. Structure du dossier	6
2.1 Entrée et sous-entrées	7
2.2 Domaine	9
2.3 Définition	9
2.4 Contexte	10
2.5 Observation.....	12
2.6 Voir aussi.....	12
2.7 Illustration.....	13
2.8 Références.....	13
3. Index	14
 CHAPITRE 2 50 DOSSIERS DE TERMINOLOGIE SUR LE TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES.....	15
1. Arbre du domaine du traitement automatique des langues naturelles...	15
2. Corpus.....	16
3. Dossiers.....	17
4. Index français.....	99
5. Index anglais.....	103

	page
CHAPITRE 3 MODES DE FORMATION NÉOLOGIQUE EN TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES.....	107
Introduction	107
A. Syntagme.....	108
1. Découpage du syntagme lexicalisé.....	110
2. Pourcentage des résultats	112
2.1 Typologie des syntagmes.....	116
2.1.1 nom + adjectif.....	117
2.1.2 nom + joncteur + nom.....	117
2.1.3 nom + joncteur + prédéterminant + nom.....	117
2.1.4 nom + joncteur + nom.....	117
2.1.5 nom + adverbe + adjectif.....	117
2.1.6 adjectif + nom.....	117
3. Analyse	118
4. Conclusion	122
B. Calque.....	124
1. Pourcentage des résultats	126
1.1 traduction littérales.....	128
1.2 adaptations syntaxiques.....	129
1.3 adaptations morphosyntaxiques.....	129
1.4 adaptations morphologiques.....	130
1.5 adaptations partielles	130
2. Analyse	130
3. Conclusion	134
CONCLUSION.....	135
BIBLIOGRAPHIE THÉMATIQUE	137
1. Bibliographie terminologique.....	137
2. Bibliographie linguistique.....	139

LISTE DES TABLEAUX

	page
Tableau 1 Lexies simples et syntagmes.....	112
Tableau 2 Syntagmes simples et syntagmes complexes.....	113
Tableau 3 Structure syntaxique des syntagmes lexicalisés.....	114
Tableau 4 Typologie des syntagmes: modèles binaires.....	116
Tableau 5 Typologie des calques.....	127

INTRODUCTION

Dans une société axée sur les communications, l'implantation de nouvelles technologies pour accélérer le traitement de l'information s'avérait inévitable. Ces technologies, mieux connues sous le nom de science informatique, regroupent plusieurs champs d'activités que l'on pense par exemple aux logiciels, à la micro-informatique ou, tout simplement, à l'intelligence artificielle.

L'objet principal de l'intelligence artificielle réside dans la conception de programmes capables de reproduire certains aspects du raisonnement humain dont la compréhension d'une langue naturelle. Et, qui dit compréhension d'une langue naturelle pense tôt ou tard à cette faculté propre à l'être humain qu'est la communication.

C'est donc sur cette capacité de communiquer que se penchent les chercheurs en traitement automatique des langues naturelles qui travaillent à la mise au point de formalismes représentant la structure profonde d'une langue dans le but de permettre un jour la communication homme-machine. Pour ce faire, le langage utilisé par l'ordinateur doit se rapprocher le plus possible de celui que les humains parlent entre eux. Mais en plus d'avoir recours à la linguistique formelle et à la logique mathématique pour la conception de nouveaux formalismes reproduisant le langage humain, les spécialistes doivent se tourner vers les sciences cognitives et la psychologie pour tenter d'expliquer les mécanismes de compréhension qui se cachent derrière la communication. Car la faculté de la parole sous-entend une certaine connaissance du monde qui permet au cerveau humain de faire les inférences nécessaires à la communication et dont l'ordinateur doit être doté pour communiquer efficacement.

Le traitement automatique des langues naturelles est une technique relativement nouvelle mais vouée à un essor considérable à cause des multiples applications qui en découlent (enseignement assisté par ordinateur, génération automatique de textes, traduction automatique, pour ne nommer que celles-là). C'est un sous-domaine qui se prête bien à une étude terminologique puisque les technologies de pointe, souvent aux prises avec un vocabulaire encore en formation, font face à des lacunes importantes relativement à la dénomination des nouveaux concepts.

C'est précisément l'objet de la terminologie « [...] de repérer, d'analyser et, au besoin de créer le vocabulaire pour une technique donnée [...] de façon à répondre aux besoins de l'utilisateur » (Dubuc cité dans Boulanger, 1985 : 14). Avant que le vocabulaire ne soit définitivement fixé par l'usage, de nombreux cas de synonymie et d'emprunts sont rencontrés avec pour résultat un manque flagrant d'uniformité des termes utilisés pour dénommer un même concept. C'est au terminologue qu'il appartient de relever et de faire la lumière sur les termes à privilégier par l'intermédiaire des dossiers de terminologie qui sont son principal moyen de transmettre à l'utilisateur les données recueillies. Car toute langue de spécialité a obligatoirement une terminologie propre qui est à la base de la communication.

En terminologie, la théorie et la pratique sont indissociables. C'est pourquoi le présent mémoire est réparti en trois chapitres. Le premier chapitre, consacré à la présentation méthodologique, relate toutes les démarches qui ont conduit de la préparation à l'élaboration des 50 dossiers de terminologie contenus dans ce mémoire. Le second chapitre est un recueil des dossiers de terminologie proprement dits. Purement pratique, l'étape terminographique consiste à établir, et ce pour chaque terme du corpus, un dossier regroupant tous les renseignements pertinents à la compréhension d'une notion. Enfin, le troisième chapitre, plus théorique cette fois, est consacré à l'analyse des deux principaux modes de formation néologique retenus dans le cadre de cette recherche, le syntagme et le calque.

CHAPITRE 1

PRÉSENTATION MÉTHODOLOGIQUE

Introduction

Le présent chapitre s'appuie principalement sur deux ouvrages terminologiques dont l'un, intitulé Méthodologie de la recherche terminologique (Auger et Rousseau, 1978), s'est appliqué à établir les fondements de la terminologie et a contribué grandement à en faire une discipline autonome en plus de mettre par écrit un protocole de rédaction des dossiers de terminologie alors que tout était à faire dans cette discipline encore récente. L'autre ouvrage consulté, le Guide de présentation des cahiers de Néologie en marche, porte sur l'aspect purement formel de la présentation des dossiers en s'inspirant des règles qui prévalent en lexicographie afin d'uniformiser les dossiers à paraître dans les cahiers de Néologie en marche, mais aussi ceux destinés à d'autres fins. «L'organisation de la microstructure, tout en s'alignant sur le modèle de la lexicographie traditionnelle, tient compte de la répartition de l'information suivant les principes fondamentaux de la terminologie (une seule notion par dossier, regroupement des synonymes, rédaction de notes techniques, etc.)» (Cayer et Lebel-Harou, 1983 : 6).

Tout travail terminologique comporte deux phases. Une phase préparatoire, la phase de recherche, qui consiste, une fois la documentation sur le domaine recueillie, à structurer le domaine traité en sous-parties afin de constituer l'arbre du domaine qui viendra plus tard aider le terminologue à replacer les termes du corpus dans leur sous-domaine d'emploi. La seconde phase, la phase de traitement terminologique, consiste à fixer la terminologie du domaine par la production de dossiers de terminologie.

Les prochaines pages relatent la démarche qui a précédé et prévalu pendant la rédaction des 50 dossiers de terminologie que comporte ce mémoire. Nous aborderons d'abord la phase de recherche pour ensuite poursuivre avec la phase de traitement terminologique.

A. Phase de recherche

1. Documentation

Après avoir arrêté notre choix sur le traitement automatique des langues naturelles, la première démarche entreprise a été la cueillette de documents susceptibles de fournir l'information nécessaire à l'élaboration de l'arbre du domaine. Certains d'entre eux, rédigés en anglais à l'intention d'un public profane, se sont avérés d'une grande utilité comme c'est le cas pour l'ouvrage de Mary Dee Harris Introduction to Natural Language Processing et de William Gevarter Artificial Intelligence, Expert Systems, Computer Vision and Natural Language Processing dont la table des matières délimite assez bien le domaine traité à cause de la vue d'ensemble qu'elle offre du domaine.

La documentation amassée est principalement composée de monographies, d'articles de périodiques spécialisés, d'actes de colloques (par exemple, COLING), de documents internes et de mémoires en français et en anglais les derniers étant plus nombreux et surtout plus à jour que les premiers. Même si nous avons tenté de privilégier les documents français non traduits, il est apparu que ceux-ci ont subi malgré tout l'influence de l'anglais qui dans bien des cas est la langue d'origine des nouveaux concepts.

En règle générale, la documentation se situe à un niveau intermédiaire et avancé de spécialisation puisqu'elle s'adresse à des spécialistes de l'intelligence artificielle et de la linguistique. Même les ouvrages destinés à des profanes du traitement automatique des langues naturelles nécessitent toutefois certaines connaissances de l'une ou l'autre de ces disciplines. Les documents ont été sélectionnés autant que possible dans les années 1980 et

plus mais le recours à des documents parus avant pour retracer des concepts de base est apparu inévitable. Nous pensons ici à des concepts comme «grammaire des cas» et «réseau sémantique» mis au point en 1967 et «système-q», «réseau de transition augmenté», «théorie de la dépendance conceptuelle», pour ne nommer que ceux qui ont vu le jour au début des années 1970.

2. Arbre du domaine

L'arbre du domaine est une sorte d'infrastructure où les sous-parties regroupées sous un terme générique entretiennent des rapports hiérarchiques entre elles et avec l'ensemble du domaine. En terminologie, l'arbre est un plan de première nécessité parce qu'il détaille chacune des parties et replace chaque sous-partie dans son champ sémantique. Il apparaît donc que le sous-domaine phases de traitement précède le sous-domaine structure de données dans la hiérarchie puisque que le premier comporte des opérations sans lesquelles il devient impossible de passer à d'autres opérations.

Le traitement automatique des langues naturelles n'est en fait qu'un sous-domaine d'application de l'intelligence artificielle qui elle-même est un sous-domaine d'une activité de portée plus vaste: l'informatique. C'est pourquoi il arrive parfois que certains termes rencontrés fassent simultanément partie de la nomenclature de l'intelligence artificielle et du traitement automatique des langues naturelles. On pense ici à *calcul des prédicats*, *grammaire des cas*, *réseau sémantique*, *théorie de la dépendance conceptuelle* qui, même si dans un premier temps font partie de l'intelligence artificielle, trouvent leur application en traitement automatique des langues naturelles.

La phase de recherche terminée, passons maintenant à la phase de traitement terminologique.

B. Phase de traitement terminologique

1. Corpus

Le corpus est en fait la liste des termes étudiés dans le cadre de cette recherche. La précision des renseignements recueillis lors de l'élaboration de l'arbre du domaine facilite grandement le travail terminologique, mais avant tout le choix du corpus, parce que la recherche terminologique est basée sur l'étude du terme dans ses rapports avec la notion. Une fois l'étude du terme amorcée, l'arbre du domaine permet de replacer la notion couverte par le terme dans son contexte d'énonciation.

Un examen du présent corpus démontre que les substantifs, dont font partie les syntagmes, sont les seules catégories lexicales représentées.

2. Structure du dossier

Chaque terme du corpus fait l'objet d'une analyse. Le dossier terminologique renferme les renseignements recueillis sur chaque terme regroupés sous les rubriques entrée et *sous-entrées, domaine, définition, contexte, *observation, *voir aussi, *illustration, les rubriques précédées d'un astérisque étant facultatives. Les rubriques sont rédigées à simple interligne.

Les dossiers sont classés en ordre alphabétique absolu et sont numérotés de 1 à 50, le numéro du dossier figurant à la rubrique entrée est placé devant l'entrée proprement dite. Le titre des rubriques est au singulier même si ces-dernières comportent plusieurs éléments. Il doit être écrit en minuscules avec une capitale et un trait de soulignement.

exemple: Entrée : 1. analyse ascendante

La rédaction des dossiers est soumise à certaines règles de présentation qui seront détaillées dans les prochaines pages.

2.1 Entrée et sous-entrées

L'entrée présente le terme étudié dans le dossier. Elle est précédée du numéro de dossier et suivie de l'indicatif lexico-grammatical. L'entrée précède les sous-entrées et toutes deux précèdent l'entrée anglaise. Cette dernière n'est cependant pas suivie de l'indicatif lexico-grammatical parce qu'elle n'est mentionnée qu'à titre de référence au concept dans sa langue d'origine.

exemple: 19. dominance immédiate, n. f.
 abrév. DI, n. f.
 immédiate dominance
 abrév. ID

Les sous-entrées, pour leur part, offrent les termes en concurrence avec l'entrée principale en commençant avec celles qui sont le plus près de l'entrée principale comme la variante orthographique (var.) qui fait état des différentes graphies rencontrées pour un même terme.

exemple: 47. système-q, n. m.
 var. système-Q, n. m.

Vient ensuite l'abréviation (abrév.) sous forme de sigle français ou anglais:

exemple: 42. réseau de transition augmenté, n. m.
 abrév. ATN, n. m.

ou de réduction d'une unité:

exemple: 21. génération automatique de textes, n. f.
 abrév. génération de texte, n. f.
 génération automatique, n. f.
 génération, n. f.

En dernier lieu apparaît le synonyme (syn.) qui renvoie à des termes concurrents ou à un même référent. Lorsqu'ils sont nombreux les synonymes apparaissent en fonction de leur intérêt linguistique et terminologique par rapport à l'entrée principale. Les abréviations et

les variantes orthographiques sont considérées comme des sous-catégories de la rubrique synonyme.

exemple: 45. schéma, n. m.
 syn.: frame, n. m.
 cadre, n. m.
 prototype, n. m.

Les sous-entrées sont placées en retrait sous la quatrième lettre de l'entrée principale et sont également suivies de leur indicatif lexico-grammatical en abrégé. De même, quand il y a plus d'une catégorie de sous-entrée, la catégorie et le terme correspondant sont alignés sur la catégorie et le terme précédents. On procède de la même façon lorsqu'une sous-entrée possède elle-même une abréviation.

exemple: 49. traitement automatique des langues naturelles, n. m.
 abrég. traitement des langues naturelles, n. m.
 abrég. TLN, n. m.
 syn. traitement automatique du langage naturel, n. m.
 abrég. traitement du langage naturel, n. m.

Lorsque le terme en entrée, en sous-entrée ou en entrée anglaise est un syntagme, il ne subit pas d'inversion. Le syntagme est traité comme un terme simple pour ce qui est de l'attribution de l'indicatif lexico-grammatical. Ainsi, le syntagme *grammaire discontinue contextuelle* sera suivi de l'indicatif n. f. puisque le genre est ici régi par la base du syntagme en l'occurrence *grammaire*.

Si une entrée principale ou une sous-entrée comporte un syntagme trop long pour tenir en une seule ligne, on utilisera le nombre de lignes nécessaire en décalant les termes apparaissant sur les autres lignes de trois espaces par rapport au début du syntagme sur la ligne précédente.

exemple: 34. grammaire syntagmatique généralisée, n. f.
 abrég. G.S.G., n. f.
 syn. grammaire de structure de phrase
 généralisée, n. f.

Lorsqu'un terme en entrée est proposé par l'auteur, la mention terme prop. l'indicatif lexico-grammatical.

exemple: 22. grammaire de cases, n. f., terme prop.

Voici les indicatifs lexico-grammaticaux et leur abréviation rencontrés dans les dossiers:

f.	féminin
n.	nom
m.	masculin
pl.	pluriel

La catégorie et le genre sont séparés par un espace.

2.2 Domaine

Le domaine comporte deux mentions. Le domaine générique ou le sujet traité et le domaine spécifique ou le sous-domaine dans lequel le terme s'imbrique. Ces derniers sont séparés par un trait oblique.

ex.: traitement automatique des langues naturelles/logiciels

Dans le cadre de cette étude le domaine générique étudié est le traitement automatique des langues naturelles à l'exception du n° 49 où le domaine générique traitement automatique des langues naturelles devient spécifique et intelligence artificielle le générique.

ex.: intelligence artificielle/traitement automatique des langues naturelles

2.3 Définition

Tout comme en lexicographie, la définition tient en une seule phrase et commence par une capitale et finit par un point. Elle regroupe les éléments d'information les plus pertinents, les autres détails étant reportés en observation. Les définisseurs sont de la même catégorie

lexicale que le terme défini; un substantif définit un substantif et ainsi de suite. Par exemple,

Entrée : 40. registre, n. m.

Définition : Variable des réseaux de transition augmentés contenant des informations globales sur la phrase analysée.

Cependant le terme de l'entrée principale n'est pas repris dans la définition à l'exception du syntagme dont le définisseur peut reprendre le premier élément constituant. Par exemple,

Entrée : 42. réseau de transition augmenté

Définition : Réseau de transition de la puissance d'une grammaire transformationnelle qui permet d'accumuler de l'information dans des registres et de formuler des conditions en tenant compte de l'information contenue dans ces registres.

La plupart des définitions contenues dans cette étude ont été élaborées avec la collaboration de François Labelle qui a agi à titre de spécialiste du domaine.

2.4 Contexte

Le contexte est une rubrique complémentaire à la définition qui atteste de l'existence des termes présents en entrée et en sous-entrée. Il cerne le sens complet d'un terme à l'intérieur d'un champ sémantique. En général, on trouve autant de contextes que d'items dans le bloc d'entrée. Si tel n'est pas le cas, les explications sur la provenance de ces termes sont reportées dans la rubrique Observation.

Un seul terme peut comporter plusieurs contextes. Le contexte reprend les entrées et les sous-entrées en suivant leur ordre d'apparition; le contexte de l'entrée principale apparaît en premier lieu de même que les contextes extraits de périodiques. Viennent ensuite les contextes provenant de monographies et en dernier lieu les contextes traduits.

Il arrive parfois qu'en l'absence d'un contexte tiré d'une source française, le contexte proposé ait été fabriqué de toutes pièces par l'auteur et les spécialistes à partir d'informations diverses ou qu'il soit une traduction de l'anglais. Dans le premier cas, aucune mention bibliographique n'apparaît et dans le second cas la mention traduit de apparaît suivie de la référence (voir 1. 7. 8).

exemples:

La grammaire discontinue statique utilise des graphes d'analyse plutôt que des arbres d'analyse comme le fait la grammaire discontinue pour structurer les données.

Une grammaire sémantique ressemble à une grammaire syntaxique à l'exception que les catégories de mots (noms, verbes) sont remplacées par des classes sémantiques (bateau, propriétés du bateau) (traduit de William B. Gevarter, Artificial Intelligence. Experts Systems. Computer Vision and Natural Language Processing, 1985, p. 116).

Dans la présente étude, les contextes définitoires et descriptifs sont les plus fréquents parce qu'ils sont porteurs de la substance du terme. Les contextes proviennent généralement de la documentation parue dans les années 1980 à l'exception des termes *réseau de transition augmenté* et *système-q* qui comportent des contextes puisés dans des documents parus au début des années 1970, faute de documents plus récents.

Fait à noter, lorsqu'il y a plus d'un contexte, et ceci vaut également pour les observations, ils sont numérotés.

Les interventions du terminologue dans un contexte s'effectuent entre crochets et apparaissent comme suit:

- [sic] pour indiquer que le texte comporte une erreur typographique, orthographique, etc.
- [...] pour supprimer une partie du texte.
- [texte] pour indiquer tout ajout au texte original.

2.5 Observation

De nature linguistique, terminologique, technique et encyclopédique, l'observation consiste à apporter des renseignements additionnels visant à faciliter la compréhension d'un terme ou d'une notion. C'est d'ailleurs dans ce même ordre qu'elles apparaissent dans le dossier et elles sont numérotées. Le terme est souligné tandis que la notion ne l'est pas.

Lorsqu'un terme apparaissant en entrée ou en sous-entrée n'a pas de contexte, la provenance de ce terme est indiquée en observation. Il s'agit ici d'une observation terminologique. Les observations techniques sont la plupart du temps traduites des documents de présentation du concept rédigés par le concepteur lui-même ou proviennent des mêmes sources que les contextes en complément de l'information que contiennent ces derniers. Quoiqu'il en soit, elles ont été formulées par des spécialistes. Voici un exemple d'observation traduite tirée de la même source que le contexte et qui fournit des détails complémentaires au contexte.

exemple:

Observation: La grammaire lexicale fonctionnelle ne contient aucune règle transformationnelle. Les règles syntagmatiques et les règles lexicales en particulier sont les pivots de cette théorie.

2.6 Voir aussi

La rubrique Voir aussi qui paraît immédiatement après la rubrique Observation reporte le lecteur à une ou d'autres notions connexes pour lui permettre une meilleure compréhension de la notion étudiée.

exemple:

Entrée : 45. schéma, n. m.

Voir aussi : case, 14.; réseau sémantique, 44.; script, 46.

Le terme auquel on renvoie est suivi d'une virgule, du numéro de dossier et d'un point. Les renvois sont séparés entre eux par un point virgule.

2.7 Illustration

Apparaissant à la toute fin de l'article, la rubrique Illustration comporte des figures et des schémas relatifs à la notion contenue dans le dossier afin d'illustrer soit une définition, un contexte ou une observation ou tout simplement de permettre au lecteur de visualiser une notion au complet le but du dossier de terminologie étant de décortiquer la notion en parties constituantes afin d'en faciliter la compréhension.

Lorsqu'un dossier renvoie à une illustration contenue dans un autre dossier, le renvoi s'effectue de la façon suivante:

Illustration : voir grammaire syntagmatique généralisée, 34.

2.8 Références

Les références citées dans les définitions et les contextes diffèrent selon qu'elles proviennent de monographies ou d'articles de périodiques.

Dans le premier cas, elles contiennent le prénom et le nom de l'auteur (s'il y a lieu), le titre de l'ouvrage souligné, l'année de publication et la page entre parenthèses. Tous les autres renseignements (collection, lieu, maison d'édition, éditeur, etc.) sont reportés dans la bibliographie.

exemple: (Maryse Condillac, Prolog: fondements et applications, 1986, p. 128).

Dans le second cas, elles comportent le titre du périodique souligné, le volume, le numéro, l'année et la page entre parenthèses.

exemple: (Revue québécoise de linguistique, vol. 14, n° 2, 1985, p. 35).

3. Index

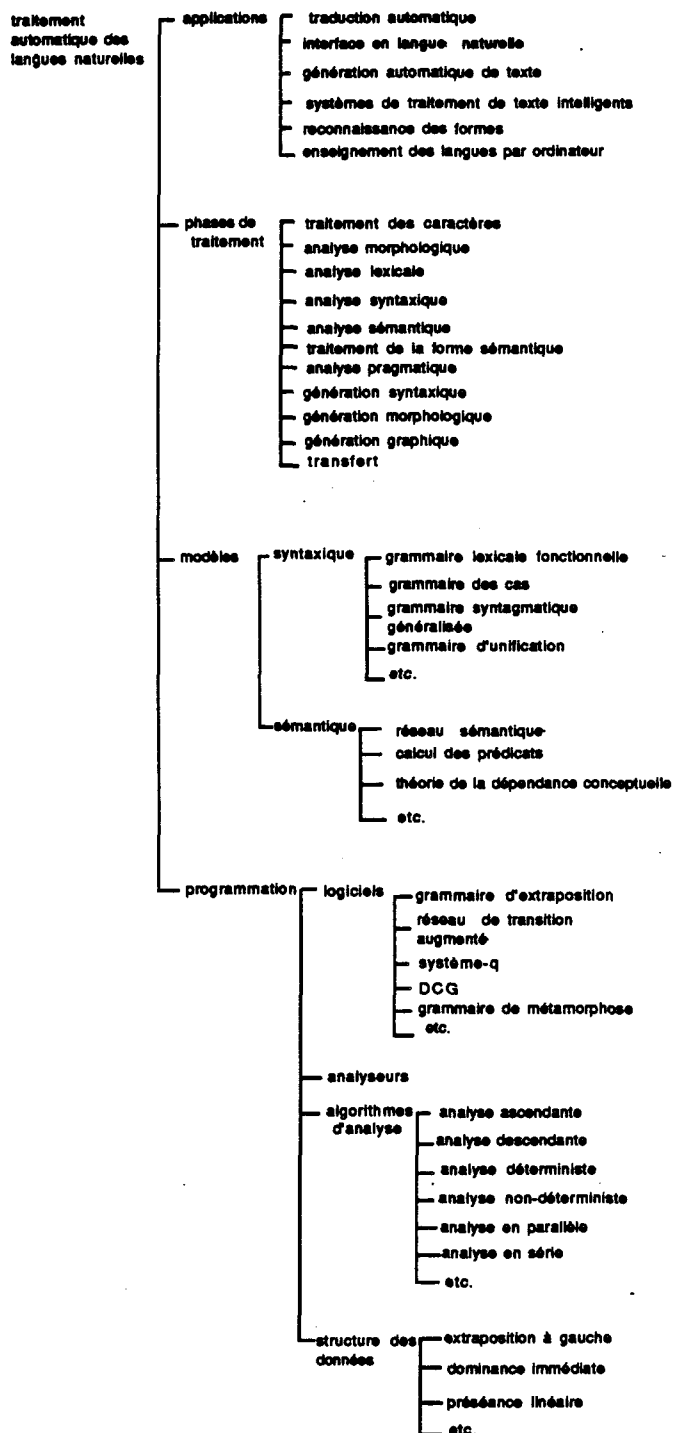
L'index français et anglais se distingue ainsi du corpus qui contient la liste des termes qui seront à l'étude en ce sens qu'il répertorie tous les termes qui figurent en entrée et en sous-entrée. Ces termes sont rangés en ordre alphabétique continu. Par exemple, le syntagme *grammaire de clauses définies* apparaît avant les syntagmes *grammaire de gap* et *grammaire de métamorphose* parce qu'en ordre alphabétique continu, on supprime les blancs entre les constituants du syntagme. Chaque terme est suivi d'une virgule et du numéro de dossier auquel il appartient. Les syntagmes ne sont pas inversés; ils suivent l'ordre naturel de la séquence lexicale.

exemple: grammaire de cases, 22
 grammaire de clauses définies, 23
 grammaire de gap, 31
 grammaire de métamorphose, 24

CHAPITRE 2

50 DOSSIERS DE TERMINOLOGIE SUR LE TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES

1. Arbre du domaine du traitement automatique des langues naturelles



2. Corpus

1. analyse ascendante
2. analyse descendante
3. analyse déterministe
4. analyse en parallèle
5. analyse en série
6. analyse lexicale
7. analyse morphologique
8. analyse non déterministe
9. analyse pragmatique
10. analyse sémantique
11. analyse syntaxique
12. analyseur syntaxique
13. calcul des prédicats
14. case
15. complémentation
16. conceptualisation
17. contrainte sémantique
18. contrainte syntaxique
19. dominance immédiate
20. extraposition à gauche
21. génération automatique de textes
22. grammaire de cases
23. grammaire de clauses définies
24. grammaire de métamorphosé
25. grammaire des cas
26. grammaire d'extraposition
27. grammaire discontinue
28. grammaire discontinue contextuelle
29. grammaire discontinue statique
30. grammaire d'unification
31. grammaire lacunaire
32. grammaire lexicale fonctionnelle
33. grammaire sémantique
34. grammaire syntagmatique généralisée
35. interface en langue naturelle
36. prédicat
37. préséance linéaire
38. primitive sémantique
39. principe de compositionnalité
40. registre
41. réseau de transition
42. réseau de transition augmenté
43. réseau de transition récursif
44. réseau sémantique
45. schéma
46. script
47. système-q
48. théorie de la dépendance conceptuelle
49. traitement automatique des langues naturelles
50. unification

<u>Entrée</u>	:	1. analyse ascendante, n. f. bottom-up parsing syn. bottom-up analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/algorithmes d'analyse
<u>Définition:</u>	:	Stratégie d'analyse des phrases qui élabore des hypothèses d'analyse à partir de données en combinant tous les constituants qu'elle rencontre.
<u>Contexte</u>	:	1. L'analyse ascendante combine aveuglément tous les constituants qu'elle trouve [...] et tente de construire l'arbre à partir de ses feuilles ou mots concrets qui apparaissent dans la chaîne (<u>Revue québécoise de linguistique</u> , vol. 14, n° 2, 1985, p. 35). 2. [...] l'analyse ascendante vérifie les mots de la chaîne d'entrée de la gauche vers la droite et construit toutes les structures possibles à gauche du mot courant à mesure que l'analyseur évolue (traduit de William B. Gevarter, <u>Artificial Intelligence, Expert Systems, Computer Vision and Natural Language Processing</u> , 1984, p. 119).
<u>Observation</u>	:	1. L'analyse ascendante cherche à partir des mots (feuilles) les règles dont le membre droit contient des mots semblables pouvant être regroupés en un seul constituant contenu dans le membre gauche. 2. L'analyse ascendante est dirigée par les données en ce sens que l'analyseur construit l'arbre syntaxique en partant des feuilles c'est-à-dire de la phrase à analyser pour atteindre la racine. La méthode consiste à partir de la phrase à appliquer à chaque pas une réduction à gauche jusqu'à ce qu'on arrive au symbole initial de la grammaire. Pour ce faire, on utilise une pile. 3. L'analyse ascendante engendre des hypothèses qui ne correspondent à rien dans la chaîne. Elle peut générer un grand nombre d'analyses incorrectes l'exactitude d'une analyse ne pouvant être déterminée une fois que toutes les analyses sont exécutées.
<u>Voir aussi</u>	:	analyse descendante, 2.; analyse non déterministe, 8.

Illustration

:

Dans cette grammaire, les premières étapes d'une analyse ascendante seraient de substituer le déterminant non terminal par *the*; l'adjectif par *new*; le nom par *program*; et le syntagme nominal par adjectif et nom.

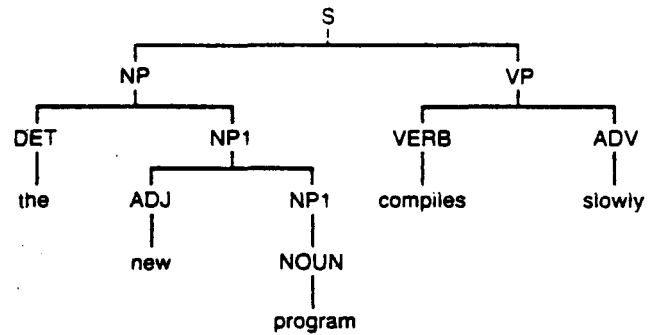
Grammar:

$S \rightarrow NP VP$
 $NP \rightarrow DET NP1$
 $NP \rightarrow NP1$
 $NP1 \rightarrow ADJ NP1$
 $NP1 \rightarrow NOUN$
 $VP \rightarrow VERB ADV$
 $NOUN \rightarrow \text{program}$
 $VERB \rightarrow \text{compiles}$
 $ADJ \rightarrow \text{new}$
 $DET \rightarrow \text{the}$
 $ADV \rightarrow \text{slowly}$

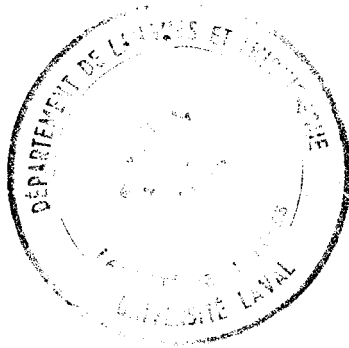
Sentence:

the new program compiles slowly

Tree:

**Analyse ascendante**

<u>Entrée</u>	:	2. analyse descendante, n. f. top-down parsing syn. top-down analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/algorithmes d'analyse
<u>Définition</u>	:	Stratégie d'analyse des phrases qui vérifie des hypothèses en les confrontant aux données et ce dès le tout début de l'analyse.
<u>Contexte</u>	:	<p>1. Une analyse syntagmatique est descendante si elle tente de construire l'arbre à partir de la racine (i.e. des hypothèses les plus abstraites, e. g. $P \rightarrow \text{comp } P$ ou $P \rightarrow \text{SN } \text{SV}$). [...] Une analyse descendante ne propose que des combinaisons "utiles" (i.e. qui peuvent aboutir à des P complètes). [...] Cependant, une analyse ascendante va engendrer des hypothèses ne correspondant à rien dans la chaîne (<u>Revue québécoise de linguistique</u>, vol. 14, n° 2, 1985, p. 35).</p> <p>2. La technique d'analyse est descendante: l'arbre syntaxique est construit de la racine vers les feuilles (Maryse Condillac, <u>Prolog: Fondements et applications</u>, 1986, p. 128).</p>
<u>Observation</u>	:	<p>1. Prenons la grammaire de clauses définies</p> <p>P. ::= GNGV GV ::= VSGV SGV ::= GN V ::= V GN ::= article nom _ commun</p> <p>L'analyse descendante consiste à construire l'arbre à partir de la racine qui est l'axiome de la grammaire ici, P. Elle cherche la première règle qui s'applique à l'axiome, c'est ici la règle $P ::= \text{GNGV}$, puis la règle $\text{GN} ::= \text{article nom _ commun}$. Lorsqu'un terminal est atteint, l'analyse vérifie que ce terminal est bien du type attendu. L'analyse se poursuit alors à partir du dernier non-terminal apparu dans l'arbre. L'analyse est réussie lorsqu'une phrase construite correspond à la phrase d'entrée.</p> <p>2. L'analyse descendante commence par l'examen des règles qui conduisent à la structure désirée (habituellement une phrase). Elle décompose ensuite chaque constituant de la structure et recherche des règles pour chacun de ces</p>



constituants. Des hypothèses sont ensuite émises sur les structures susceptibles d'être reconnues.

3. L'analyse descendante procède par retour-arrière; si toutes les règles à un niveau particulier échouent, l'analyseur retourne en arrière au niveau précédent pour essayer une autre règle. Pendant le retour-arrière, le même constituant peut être analysé plusieurs fois. Ce type d'analyse peut donc s'avérer lent.

Voir aussi : analyse ascendante, 1.; analyse non déterministe, 8.

<u>Entrée</u>	:	3. analyse déterministe, n. f. deterministic parsing syn. deterministic analysis wait-and-see parsing
<u>Domaine</u>	:	traitement automatique des langues naturelles/algorithmes d'analyse
<u>Définition</u>	:	Stratégie d'analyse des phrases qui procède sans retour-arrière ni parallélisme parce qu'elle considère qu'une phrase possède tous les éléments qui permettent de prendre une décision définitive pour l'interprétation.
<u>Contexte</u>	:	Une stratégie d'analyse syntaxique est déterministe au sens strict si, pour tout point de la chaîne d'entrée, le choix d'analyse est strictement guidé par l'état de l'analyseur en ce point et par les informations qu'il peut y obtenir (<u>Revue québécoise de linguistique</u> , vol. 14, n° 2, 1985, p. 36).
<u>Observation</u>	:	1. Une particularité de l'analyse déterministe: les spécificités lexicales de la langue analysée sont prises en considération dans la construction des règles. En effet, celles-ci doivent comporter, dans les cas d'ambiguïté lexicale de certains mots, les informations nécessaires pour prendre la bonne décision. 2. L'analyse déterministe, contrairement à l'analyse non déterministe, ne comporte pas de retour-arrière et bloque au moment même où apparaît un problème. La vitesse d'analyse s'en trouve donc améliorée. 3. Le système Parsifal mis au point en 1980 est le mieux connu des systèmes de traitement automatique des langues naturelles utilisant l'analyse déterministe. Avec ce système, Marcus a tenté de démontrer qu'il est possible d'analyser les langues naturelles de façon déterministe en exploitant au maximum les contraintes locales qu'elles contiennent. Il tente également de démontrer que l'ambiguïté est due aux types d'analyses effectuées et non au langage.
<u>Voir aussi</u>	:	analyse non déterministe, 8.

<u>Entrée</u>	:	4. analyse en parallèle, n. f. parallel parsing syn. parallel analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/ algorithmes d'analyse
<u>Définition</u>	:	Stratégie d'analyse des phrases qui vérifie plusieurs possibilités simultanément.
<u>Contexte</u>	:	Une analyse en parallèle considère toutes les hypothèses à la fois et les mène de front. Une telle approche a l'avantage d'éviter de refaire les analyses déjà faites. Elle fournit l'ensemble des solutions possibles. Cependant, comme on l'a déjà noté, le fait de devoir conserver toutes les hypothèses tout au long de l'analyse "bouffe" littéralement l'espace de travail de la machine. De plus, on peut ajouter qu'il ne semble pas que les humains fonctionnent systématiquement ainsi (et ce sont d'excellents parseurs) (<u>Revue québécoise de linguistique</u> , vol. 14, n° 2, 1985, pp. 33-34).
<u>Observation</u>	:	L'analyse en parallèle est effectuée au moyen d'un analyseur qui peut être très rapide mais qui occupe beaucoup d'espace inutilement puisqu'il mène de front toutes les hypothèses et doit donc toutes les emmagasiner.

<u>Entrée</u>	:	5. analyse en série, n. f. serial parsing syn. serial analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/ algorithmes d'analyse
<u>Définition</u>	:	Stratégie d'analyse des phrases qui fait marche arrière chaque fois qu'elle découvre une combinaison de mots inacceptable.
<u>Contexte</u>	:	Une analyse en série prend chaque hypothèse une à une et la (les) pousse jusqu'à leur limite. En cas de cul-de-sac, on fait marche-arrière jusqu'au dernier point où un choix était impossible et on essaie un autre membre de l'alternative (<u>Revue québécoise de linguistique</u> , vol. 14, n° 2, 1985, p. 34).
<u>Observation</u>	:	<ol style="list-style-type: none">1. L'analyse en série n'est efficace que si les hypothèses qu'elle contient sont bien ordonnées, les plus probables d'abord. Si la langue à analyser contient un facteur élevé d'indéterminisme, elle exigera trop de temps.2. Une analyse en série occupe peu d'espace car elle ne se perd pas en analyses inutiles mais elle est plus lente parce qu'elle doit tenir compte de plusieurs chemins et faire marche-arrière.

<u>Entrée</u>	:	6. analyse lexicale, n. f. lexical parsing syn. lexical analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Stratégie d'analyse des phrases qui confirme les hypothèses de l'analyse morphologique et qui vérifie les mots dans le lexique afin de déterminer leur catégorie lexicale et leurs traits sémantiques et syntaxiques.
<u>Contexte</u>	:	L'étape suivante est l'analyse lexicale qui consiste à déterminer à quelle catégorie lexicale appartient chaque mot (par exemple: nom ou adjectif) et à identifier les traits grammaticaux indiqués par les affixes, comme par exemple le pluriel (<u>Pour la Science</u> , n° 85, novembre 1984, p. 99).
<u>Observation</u>	:	<p>1. L'analyse lexicale tente de repérer et d'éliminer toute erreur typographique ou orthographique.</p> <p>2. L'analyse lexicale rend compte des mots d'une langue qui sont emmagasinés dans une structure de données appelée lexique. Ces mots y sont classés par catégories (terminales): noms, adjectifs, verbes, adverbes et conjonctions, prépositions, pronoms, interjections.</p> <p>3. Lors de l'analyse lexicale, certaines transformations sont faites sur les caractères afin d'obtenir une forme standard employée dans le lexique. Par exemple, toutes les lettres majuscules sont transformées en minuscules.</p> <p>4. L'analyse lexicale produit en sortie une liste de listes, chacune étant des sous-listes qui correspondent à un mot: [[c,o,n,n,a,i,s,s,e,z], [v,o,u,s], [l,e], [m,o,t], chaîne ([m, o, t])].</p>
<u>Voir aussi</u>	:	analyse morphologique, 7.

<u>Entrée</u>	:	7. analyse morphologique, n. f. morphological parsing syn. morphological analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Stratégie d'analyse des phrases qui décompose tous les mots de la chaîne selon les règles de la morphologie dérivationnelle ou flexionnelle de la langue.
<u>Contexte</u>	:	<p>1. Une analyse morphologique est faite pour tous les mots de la chaîne d'entrée qui ne figurent pas dans le dictionnaire, en particulier elle permet: a) de reconnaître les désinences des verbes et de les remplacer par une forme canonique arrivera-->arrive, arriver-->arrive; de remplacer les pluriels de substantifs par leur singulier. Ces deux processus permettent une économie appréciable d'entrées dans le dictionnaire; c) de reconnaître des concaténations spéciales de mots en un seul bloc : est-ce, existe-t-il, l'oiseau, d'avion; d) de détacher la ponctuation de la fin des mots (<u>R.A.I.R.O. Informatique / Computer Science</u>, vol. 14, n° 2, 1980, p. 141).</p> <p>2. L'analyse morphologique est effectuée pour identifier les formes de flexion régulière et l'information provenant du lexique est rattachée à chaque mot. [...] Cette phase consiste à convertir la séquence d'entrée en une représentation du sens de la phrase afin d'en permettre le traitement (traduit de Mary Dee Harris, <u>Introduction to Natural Language Processing</u>, 1983, p. 316).</p>
<u>Observation</u>	:	<p>1. Lors de l'analyse morphologique, il arrive souvent que les morphèmes (racines et suffixes) soient analysés selon leur catégorie syntaxique et sémantique.</p> <p>2. L'analyse morphologique permet d'identifier les mots de la phrase (nom, verbe, adjectif...). C'est d'ailleurs à cette étape que les principaux constituants se voient attribuer un sens élémentaire à partir du dictionnaire.</p> <p>3. L'analyse morphologique décompose les mots en radicaux et en désinences, tels que les terminaisons -s ou -ant, suivant certaines règles élémentaires: par exemple le radical de "toiser" est "toise", tandis que celui de "boiser" est "bois". Les exceptions à ces règles, telles que certaines formes du verbe "être", sont regroupées à part. D'autres règles permettent de relier la forme des affixes aux formes des mots, comme par exemple "ant" dans "en marchant" qui correspond à un gérondif et qui indique une action en cours.</p>

<u>Entrée</u>	:	8. analyse non déterministe, n. f. nondeterministic parsing syn. nondeterministic analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/algorithmes d'analyse
<u>Définition</u>	:	Stratégie d'analyse des phrases qui, lorsqu'il y a différentes possibilités d'analyse, applique successivement les différentes possibilités plutôt que de n'en retenir qu'une seule comme l'analyse déterministe.
<u>Contexte</u>	:	Elle [analyse non déterministe] consiste à envisager à chaque instant les hypothèses possibles, à tenter de déterminer la plus plausible, à la mettre en oeuvre, et à recommencer jusqu'au succès ou jusqu'à une impasse. Celle-ci est détectée par le fait que l'on parvient à un état où toutes les hypothèses permettant de continuer contredisent des choix antérieurs. En cas d'impasse, il faut revenir (backtrack) au dernier état où un autre choix aurait pu être fait, en détruisant les hypothèses construites dans l'intervalle (<u>La Recherche</u> , vol. 16, n° 170, octobre 1985, p. 1200).
<u>Observation</u>	:	1. Une analyse non déterministe peut être ascendante ou descendante. Elle peut soit imposer des contraintes fortes et n'accepter que des énoncés bien formés ou relâcher ces contraintes sur tout le déroulement de l'analyse et faire des erreurs. 2. Lorsqu'un choix incorrect est effectué en début d'analyse, l'analyse non déterministe consacre beaucoup de temps à construire des structures, parfois presque entièrement incorrectes puis, par l'effet du retour arrière, à les démolir jusqu'à ce qu'il n'en subsiste que ce qui existait à l'endroit de l'erreur avant de les reconstruire une fois que le choix erroné aura été modifié.
<u>Voir aussi</u>	:	analyse ascendante, 1.; analyse descendante, 2.; analyse déterministe, 3.

<u>Entrée</u>	:	9. analyse pragmatique, n. f. pragmatic parsing syn. pragmatic analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Stratégie d'analyse des phrases qui remplace l'énonciation d'une phrase ou d'une partie de phrase dans son contexte socio-culturel tout en tenant compte des conventions du discours.
<u>Contexte</u>	:	L'étape finale est l'analyse pragmatique qui permet d'explicitier le contexte de la phrase comme par exemple la relation entre le moment de l'énonciation et celui auquel elle se réfère. L'ordinateur peut alors tirer des inférences et formuler une réponse à la phrase qu'on lui a soumise (<u>Pour la Science</u> , n° 85, novembre 1984, p. 99).
<u>Observation</u>	:	<ol style="list-style-type: none"> 1. L'analyse pragmatique a pour tâche d'insérer complètement une phrase dans son contexte, celui-ci pouvant être un texte, une conversation, etc. 2. L'analyse pragmatique est rarement intégrée dans les programmes actuels parce que les problèmes qu'elle pose sont très complexes. Cependant, l'intérêt porté à la pragmatique dans le problème de compréhension a permis de mettre en lumière un point important. Le texte à comprendre n'est que "différentiel": l'auteur n'exprime pas une situation, il exprime la différence entre la situation qu'il vient décrire et celle qu'il sait que l'auteur connaît déjà. 3. L'analyse pragmatique est souvent imbriquée avec l'analyse sémantique dont elle cherche à modifier le résultat par la prise en considération d'éléments extérieurs au contexte.
<u>Voir aussi</u>	:	analyse sémantique, 10.

<u>Entrée</u>	:	10. analyse sémantique, n. f. semantic parsing syn. semantic analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Analyse de la structure conceptuelle de la phrase, en termes de fonctions sémantiques telles que agent, patient, instrument, but, origine ..., puis de la structure logique de la phrase exprimée dans le calcul des prédicats.
<u>Contexte</u>	:	1. Puis vient l'analyse sémantique qui consiste à représenter la phrase sous une forme dont le logiciel peut tirer des inférences (<u>Pour la Science</u> , n° 85, novembre 1984, p. 99). 2. [...] l'analyse sémantique se réduit à déterminer si oui ou non une phrase arbitraire "décrit" correctement le monde [...] (<u>Langages</u> , n° 87, septembre 1987, p. 107).
<u>Observation</u>	:	1. L'analyse sémantique traite de la signification des mots, groupes de mots et phrases. À chaque mot du lexique sont associées une ou plusieurs représentations internes ou sémantiques qui peuvent prendre des formes très diversifiées. Cette forme interne est une partie de formule logique (une constante, un prédicat, une formule). 2. L'analyse sémantique consiste à traduire la structure construite par l'analyse syntaxique en une formule logique considérée comme exprimant la signification de la question posée par l'utilisateur. À chaque règle de grammaire est associée une procédure qui détaille la construction d'une partie de la formule logique. Ces procédures, mises bout à bout dans l'ordre où les règles de grammaire ont été utilisées, constituent un programme dont l'exécution donne pour résultat la formule souhaitée. 3. La formule logique obtenue lors de l'analyse sémantique, les informaticiens choisissent le plus souvent de l'exprimer à l'aide du calcul des prédicats.
<u>Voir aussi</u>	:	calcul des prédicats, 13.; contrainte sémantique, 17.; contrainte syntaxique, 18.

<u>Entrée</u>	:	11. analyse syntaxique, n. f. syntactic parsing syn. syntactic analysis
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Stratégie d'analyse des phrases qui consiste à construire la structure de la phrase en la faisant passer à travers une grammaire française afin d'identifier la catégorie syntaxique de chaque mot et à les regrouper sous forme arborescente.
<u>Contexte</u>	:	1. L'analyse syntaxique reconnaît la structure de la phrase en identifiant les rapports entre chaque mot et en les expliquant sous forme arborescente (<u>TAUM 73</u> , 1973, p. 10). 2. L'analyse syntaxique consiste à déterminer si à une phrase donnée correspond une certaine structure, laquelle est décrite par des règles de grammaire (Bertrand Pelletier, <u>Système d'interrogation d'une banque de données en langue naturelle</u> , 1986, p. 4).
<u>Observation</u>	:	1. L'analyse syntaxique fait passer la phrase à travers une grammaire française afin de connaître sa structure. 2. Pour qu'une procédure d'analyse syntaxique ait des chances de survie, il faut qu'elle n'utilise que des ressources limitées.
<u>Voir aussi</u>	:	analyseur syntaxique, 12.

<u>Entrée</u>	:	12. analyseur syntaxique, n. m. syntactic parser syn. syntactic analyzer
<u>Domaine</u>	:	traitement automatique des langues naturelles/ analyseurs
<u>Définition</u>	:	Programme qui applique des règles grammaticales afin de déterminer la structure de la phrase sous forme d'un graphe ou d'un arbre.
<u>Contexte</u>	:	1. Un programme informatique qui construit l'arbre syntaxique des phrases est appelé un analyseur syntaxique (W. F Clocksin et C. S Mellish, <u>Programmer en Prolog</u> , 1981, p. 216). 2. D'un point de vue terminologique, il suffit de noter qu'en programmation, on appellera "syntaxique" un analyseur construisant un arbre dont les noeuds sont étiquetés par des catégories syntaxiques [...] (Alain Polguere, <u>Programmation logique des interfaces en langue naturelle</u> , 1984, p. 19).
<u>Observation</u>	:	1. Les analyseurs syntaxiques utilisent diverses méthodes pour rechercher les différentes façons de regrouper les mots. Certains procèdent du haut vers le bas: ils cherchent des phrases vraisemblables dès le début de l'analyse. D'autres procèdent du bas vers le haut et essaient différentes combinaisons locales de mots. Certains logiciels font marche arrière chaque fois qu'ils découvrent une combinaison de mots inacceptable et ils en recherchent d'autres. D'autres logiciels effectuent un traitement en parallèle afin d'étudier plusieurs possibilités simultanément. Certains utilisent des formalismes mis au point par les linguistes (comme les grammaires transformationnelles), tandis que d'autres utilisent de nouveaux formalismes spécialement conçus pour l'informatique comme les A.T.N., et donc beaucoup mieux adaptés à l'exécution des procédures d'analyse syntaxique. 2. La mise au point d'un analyseur syntaxique pose deux problèmes distincts. Le premier est la détermination précise d'un ensemble de règles, c'est-à-dire une grammaire qui spécifie l'ensemble de toutes les structures possibles des phrases d'une langue donnée. Le second problème est celui de l'analyse syntaxique car il n'est pas toujours possible de savoir quel est le rôle d'un groupe de mots dans la phrase, ou si les mots qui le composent sont réellement liés.

Voir aussi : analyse ascendante, 1.; analyse descendante, 2.; analyse déterministe, 3.; analyse non déterministe, 8.; analyse syntaxique, 11.

<u>Entrée</u>	:	<p>13. calcul des prédicats, n. m.. syn. logique des prédicats, n. f.</p> <p>predicate calculus syn. predicate logic</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Formalisme permettant de représenter la connaissance de façon très générale sans se préoccuper d'un domaine particulier en ramenant cette connaissance à une combinaison de propriétés, d'entités quantifiables, de quantificateurs et d'opérateurs logiques.
<u>Contexte</u>	:	<p>1. Le calcul des prédicat [sic] est une autre représentation qui a été souvent utilisée. [...] elle a un intérêt pratique: il existe de nombreux programmes de démonstration de théorèmes pour le calcul des prédicats. Si donc nous avons traduit une question ou un ordre dans ce langage, nous aurons automatiquement la réponse à la question ou les actions à entreprendre pour exécuter l'ordre si le démonstrateur de théorèmes arrive à prouver que l'énoncé correspondant est un théorème (Jacques Pitrat, <u>Textes, ordinateurs et compréhension</u>, 1985, p. 29).</p> <p>2. Les représentations de type logique, et tout particulièrement celle de la logique des prédicats du premier ordre, sont les premières candidates à un système de représentation des connaissances et cela pour plusieurs raisons:</p> <ul style="list-style-type: none"> i) elles ont une syntaxe rigoureusement définie, ii) elles ont une sémantique rigoureusement définie, iii) elles ont une théorie déductive rigoureusement définie, iv) elles ont des applications informatiques. <p>(<u>Langages</u>, n° 87, septembre 1987, p. 104).</p>
<u>Observation</u>	:	<p>1. Les avantages du calcul des prédicats se situent au niveau de la modularité et de la généralité. Chaque fait est valide, indépendant de ce que contient le système. La notation explicitement générale n'est pas confinée à une forme particulière de la connaissance. Le principal problème que présente cette approche est celui de l'efficacité. Tous les systèmes qui l'ont adopté se sont vu restreindre à des ensembles de faits très limités de moins de cent et souvent inférieurs à dix. La complexité d'un tel système dépend du souci de complexité qui lui est accordé.</p>

2. Le calcul des prédicats est la forme logique la plus utilisée par les informaticiens. La formule logique obtenue lors de l'analyse sémantique est exprimée à l'aide du calcul des prédicats et est confrontée aux données emmagasinées dans la machine pour l'élaboration d'une réponse.

3. Lors de l'analyse sémantique, la forme syntaxique est transformée en une forme logique exprimée en calcul des prédicats à partir de laquelle le logiciel peut appliquer des raisonnements et tirer des inférences. Le calcul des prédicats emploie des quantificateurs tels \forall ("tous") et ("il existe"). Dans le calcul des prédicats, la phrase "Des roses s'épanouiront" équivaut à l'assertion: "Il existe quelque chose qui est rose et qui s'épanouira".

4. Dans l'exemple précédent, la difficulté est de savoir si le sens de "des roses s'épanouiront" peut être représenté par une seule rose, ou s'il faut préciser deux ou plus.

Voir aussi : analyse sémantique, 10.; prédicat, 36.

<u>Entrée</u>	:	14. case, n. f. syn. slot, n. m. slot
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Composante d'une structure associée à chaque mot du dictionnaire caractérisant un attribut syntaxique ou sémantique de ce mot.
<u>Contexte</u>	:	<p>1. La notion de case est une généralisation de la notion de propriété, c'est-à-dire de couple attribut-valeur. Comme pour un attribut-valeur, il est possible d'assigner une valeur à une case (Laurence Danlos, <u>Génération automatique de textes en langue naturelle</u>, 1985, p. 152).</p> <p>2. Essentiellement un slot définit le type sémantique d'un nom, adjectif, pronom etc. (Bertrand Pelletier, <u>Système d'interrogation de banques de données en langue naturelle</u>, 1986, p. 16).</p> <p>3. Le slot est un élément-clé de cette grammaire [grammaire des cases]. Construit à l'aide de l'opérateur infixe ':', il a une structure tripartite: informations syntaxiques sur le complément, ':' variable identifiant le slot, ':' traits sémantiques (Alain Polguère, <u>Programmation logique des interfaces en langue naturelle</u>, 1984, p. 20).</p>
<u>Observation</u>	:	<p>1. Le terme <u>slot</u> est fréquemment utilisé dans la communauté scientifique et le terme <u>case</u> est très peu employé. Laurence Danlos utilise le terme <u>case</u>.</p> <p>2. Voici un exemple de cases:</p> <p style="padding-left: 40px;">pomme [+N, -Hum, +Com, -Ani] homme [=N, +Hum, -Com, +Ani] mange [+Ani] [+V] [+Com]</p> <p>où <u>pomme</u> est un nom à caractère humain, comestible et non animé. De même, <u>homme</u> est un nom à caractère humain, non comestible mais animé. Finalement, <u>mange</u> est un verbe dont le premier actant (sujet logique) doit être animé et dont le complément doit être comestible.</p> <p>3. Chaque case comporte une règle de case qui peut, d'une part, vérifier le type de composante contenue dans la case et, d'autre part, mener des actions.</p>

4. Les unités de base de la connaissance sont des entités (personne, objets, action, état, etc.) qui sont représentées sous forme de molécules et qui sont caractérisées par un ensemble de cases. Par exemple, une personne est une molécule caractérisée par les cases NOM, PRENOM, NATIONALITE, etc. Un attentat est une molécule comportant les cases ACTEUR, CIBLE, ACTE, RESULTAT, DATE et GEO.

Voir aussi

:

complémentation, 15.; contrainte sémantique, 17.;
contrainte syntaxique, 18.; grammaire de cases, 22.

<u>Entrée</u>	:	15. complémentation, n. f. complémentation
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Représentation qui permet de définir les compléments possibles du verbe, du nom et de l'adjectif à partir des contraintes sémantiques et syntaxiques contenues dans le lexique.
<u>Contexte</u>	:	Par complémentation, nous entendons la détermination du sujet logique d'un verbe, le rattachement des compléments du verbe (objet direct, circonstanciel, ...) et du nom (relative, adjectif, ...), etc. [...] Les critères de complémentation (du nom, du verbe, etc.) peuvent se diviser en deux types: syntaxique et sémantique (Bertrand Pelletier, <u>Système d'interrogation de banque de données en langue naturelle</u> , 1986, p. 16).
<u>Observation</u>	:	1. À chaque mot du dictionnaire correspond une liste de traits. Par exemple,

pomme [+ N, - Hum, + Com, - Ani]
homme [+ N, + Hum, - Com, - Ani]
mange [+ Ani] [+V] [+Com]

Ces traits servent de contraintes syntaxiques et sémantiques: pomme est un nom à caractère non humain, comestible, mais inanimé. De même, homme est un nom à caractère humain, non comestible, mais animé. Finalement mange est un verbe dont le premier actant (sujet logique) doit être animé et dont le complément doit être comestible.

2. Dans le cas de la complémentation droite d'un verbe, une composante syntaxique y est ajoutée. Il s'agit d'une complémentation syntaxico-sémantique. Voici quelques entrées du dictionnaire de Mc Cord:

verbe (give, take, (S, W, X, Z, C), C, X :: prsn :: fac,
[iobj :: S :: prsn :: stdnt, obj :: Z :: grade, pobj
(in) :: W :: crs]).

noun (teacher, fac (X,Y), nil, X :: prsn :: fac, [npobj (of
:: Y :: subject]).

noun (student, stdnt, (S, _, _, _), nil, S :: prsn :: stdnt, [
]).

noun (course, crs(X, Y, _, _), nil, X :: crs, [nobj(in) :: Y
:: subject]).

Ces clauses indiquent notamment que le verbe give nécessite comme premier actant une entité dont la case est X :: prsn :: fac, c'est-à-dire de type sémantique prsn

(personne) faisant partie du sous-type fac (faculté). De plus, ce verbe prend comme complément d'objet indirect une entité de type prsn et de sous-type stdnt(étudiant), ainsi (notamment) qu'un complément prépositionnel, introduit par «in», de type crs (cours). De même, l'entité student est un nom de type prsn, de sous-type stdnt, et n'accepte aucun complément ([]).

3. La validation et le rattachement de la complémentarité se font, d'une part, à l'aide de règles de grammaire DCG et, d'autre part, par l'intermédiaire du prédicat filler. Les prédicats apparaissant comme second argument des clauses ci-dessus servent à la construction de la formule logique correspondant à la requête soumise.

4. Dans ce système, les cases sont des termes utilisés pour contrôler la complémentation de verbes, de noms et d'adjectifs. Ils sont associés au sens des mots du lexique et déterminent les compléments possibles pour un mot donné. Le sens d'un mot implique une prédication comme donner (X, Y, Z) dont les cases
obj.: Y , ; obj.: Z
sont associées à certains arguments. Le sujet logique est traité de façon particulière et aucune case lui est associée.

5. La complémentation est une technique mise au point par Michael McCord en 1982.

Voir aussi

:

case, 14.; contrainte sémantique, 17.; contrainte syntaxique, 18.; grammaire de cases, 22.

<u>Entrée</u>	:	16. conceptualisation, n. f. conceptualization																		
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique																		
<u>Définition</u>	:	Représentation sémantique à l'intérieur de laquelle toute phrase est composée d'une action et d'un acteur, d'un objet et d'une direction ainsi que d'un contenant ou d'un instrument.																		
<u>Contexte</u>	:	Nous pouvons dire approximativement qu'une conceptualisation correspond au sens d'une clause. Une conceptualisation comporte une action à laquelle sont liés divers objets ou sens-structure. Cette représentation comporte une douzaine d'actions différentes. L'une est MTRANS qui indique un transfert d'information. L'auteur du transfert est relié à l'action par <=> (Jacques Pitrat, <u>Textes, ordinateurs et compréhension</u> , 1985, p. 4).																		
<u>Observation</u>	:	<p>1. La conceptualisation est une structure interne qui représente l'information contenue dans une phrase en dépendance conceptuelle. À l'intérieur de ces structures, il y a différents rôles à remplir: acteur, action, objet, direction, instrument et état.</p> <p>Les concepts qui peuvent jouer ces différents rôles sont:</p> <p>PP: objet du monde réel, ACT: actions du monde réel, PA: attributs d'objets, AA: attributs d'actions.</p> <p>2. Une conceptualisation peut comporter des temps:</p> <table><tr><td>(null)</td><td>présent</td></tr><tr><td>P</td><td>passé</td></tr><tr><td>f</td><td>futur</td></tr><tr><td>/</td><td>négation</td></tr><tr><td>ts</td><td>début de transition</td></tr><tr><td>tf</td><td>fin de transition</td></tr><tr><td>c</td><td>conditionnel</td></tr><tr><td>k</td><td>continu</td></tr><tr><td>?</td><td>question</td></tr></table>	(null)	présent	P	passé	f	futur	/	négation	ts	début de transition	tf	fin de transition	c	conditionnel	k	continu	?	question
(null)	présent																			
P	passé																			
f	futur																			
/	négation																			
ts	début de transition																			
tf	fin de transition																			
c	conditionnel																			
k	continu																			
?	question																			
<u>Voir aussi</u>	:	théorie de la dépendance conceptuelle, 48.																		

<u>Entrée</u>	:	18. contrainte syntaxique, n. f. syntactic constraint
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Condition grammaticale ou d'ordonnancement à laquelle une structure linguistique doit répondre.
<u>Contexte</u>	:	Les contraintes syntaxiques s'expriment en termes de rapports entre mots. Parmi ces rapports, ceux qui concernent l'accord des mots entre eux occupent une place prépondérante. On entend par accord la convention établie relativement à une ou plusieurs composantes morphologiques (de genre, de nombre, de personne, ...) entre deux ou plusieurs mots appartenant à une même phrase. De ces mots, l'un plus fort (par exemple le nom), impose à un ou plusieurs autres mots (par exemple un nom ou un adjectif) la nécessité de prendre la même position que lui dans les différentes catégories morphologiques (Maryse Condillac, <u>Prolog: fondements et applications</u> , 1986, p. 121).
<u>Observation</u>	:	Au niveau de chaque règle d'analyse, les contraintes syntaxiques apposent aux mots certains traits syntaxiques qui permettront à deux mots de correspondre entre eux. Ces traits syntaxiques sont: - pour les noms et les adjectifs: le genre et le nombre; - pour les verbes conjugués à un temps simple: le nombre. Les trois personnes du singulier sont regroupées en "singulier" et les trois personnes du pluriel en "pluriel"; - pour les verbes conjugués à un temps composé: le genre et le nombre.
<u>Voir aussi</u>	:	case, 14.; complémentation, 15.; contrainte sémantique, 17.

<u>Entrée</u>	:	19. dominance immédiate, n. f. abrév. DI, n. f. immediate dominance abrév. ID
<u>Domaine</u>	:	traitement automatique des langues naturelles/structure des données
<u>Définition</u>	:	Mécanisme qui prévoit que dans une règle de réécriture du type $A \rightarrow B_1 \dots B_n$, le membre gauche A domine immédiatement chacun des membres droits B_i et B_n .
<u>Contexte</u>	:	<p>Les règles de structure syntagmatique peuvent être condensées par le mécanisme de dominance immédiate et de précedence linéaire. Si nous avons les règles suivantes:</p> $A \rightarrow B, C, D.$ $A \rightarrow B, D, C.$ $A \rightarrow D, B, C.$ <p>Il peut être intéressant de les regrouper en une seule règle qui énoncerait que A domine immédiatement B, C, D, ces trois constituants étant dans un ordre quelconque à l'exception de B qui doit toujours apparaître avant C. Ceci est noté:</p> $A \rightarrow B; C; C; D; B < C.$ <p>La relation \rightarrow spécifie la dominance et le symbole ; est un séparateur de symboles qui n'impose aucun ordre des fils. L'inégalité < impose que B apparaisse avant C (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u>, 1989, p. 38).</p>
<u>Observation</u>	:	<p>1. Jean-Yves Morin dans son article intitulé <i>Théorie syntaxique et théorie du parsage</i> utilise l'abréviation <u>DI</u> pour <u>dominance immédiate</u>.</p> <p>2. Dans une règle de réécriture de la forme $P \rightarrow SN \dots SV$, le constituant P domine immédiatement les constituants $SN \dots SV$. En d'autres mots, les constructions $SN \dots SV$ sont des sous-parties immédiates de P.</p> <p>3. En dominance immédiate, le membre droit des règles de dominance immédiate est un ensemble de symboles terminaux et non terminaux. Les items répétitifs sont donc rejetés.</p> <p>4. La dominance immédiate a été mise au point par Stewart Shieber en 1984.</p>
<u>Voir aussi</u>	:	grammaire syntagmatique généralisée, 34.; préséance linéaire, 37.
<u>Illustration</u>	:	voir grammaire syntagmatique généralisée, 34.

<u>Entrée</u>	:	20. extraposition à gauche, n. f. left extraposition
<u>Domaine</u>		traitement automatique des langues naturelles/structure des données
<u>Définition:</u>		Structure de phrase dans laquelle un constituant est déplacé à gauche de la phrase comme pour le pronom relatif ou le pronom interrogatif dans la proposition relative ou dans la phrase interrogative, par exemple.
<u>Contexte</u>	:	<p>1. L'extraposition à gauche apparaît lorsqu'un sous-constituant, placé à gauche de celui qui est incomplet dans la phrase, représente d'une certaine façon l'élément manquant (Maryse Condillac, <u>Prolog: fondements et applications</u>, 1986, p. 125).</p> <p>2. Le concept d'extraposition à gauche joue un rôle important, de façon directe ou indirecte, dans bien des descriptions formelles de propositions relatives ou interrogatives. Il existe plusieurs contraintes globales dont les contraintes d'îlots relatives à ce concept qui ont été introduites pour restreindre les situations où l'extraposition à gauche peut être appliquée (traduit de <u>American Journal of Computational Linguistics</u>, vol. 7, n° 4, octobre-décembre 1981, p. 245).</p>
<u>Observation</u>	:	<p>1. L'extraposition à gauche a été utilisée par les grammairiens comme moyen pour décrire les phrases interrogatives et les propositions relatives en langue naturelle.</p> <p>2. Dans l'exemple: L'homme que Jean a rencontré est un peintre, il y a une proposition relative composée d'un pronom relatif suivi de la phrase <<Jean a rencontré>> à laquelle il manque un syntagme nominal (dans ce cas-ci, c'est le complément d'objet direct qui est manquant). Ce syntagme nominal est représenté dans la phrase par le pronom relatif <<que>>. On peut alors réécrire cette phrase: L'homme que; Jean a rencontré X; est un peintre où X (appelé trace) représente le syntagme nominal absent, <<que>> un marqueur et l'indice représente le lien entre ces deux entités. Le pronom relatif est lui-même rattaché au nom qui le précède dans la principale. On peut dire que le constituant qui a été remplacé par la marque X a été extraposé à gauche et qu'il est, dans cette nouvelle position, représenté par un marqueur.</p>
<u>Voir aussi</u>	:	grammaire d'extraposition, 26.

<u>Entrée</u>	:	<p>21. génération automatique de textes, n. f. abrév. génération automatique, n. f. génération de texte, n. f. génération, n. f.</p> <p>text generation</p>
<u>Domaine</u>	:	traitement des langues naturelles/applications
<u>Définition</u>	:	Processus de production d'une phrase ou d'un texte à partir d'une structure sémantique abstraite (conceptuelle ou logique).
<u>Contexte</u>	:	<p>1. Historiquement, la génération automatique [...] ne s'est développée que depuis moins d'une dizaine d'années. Mais la génération automatique est un domaine en plein essor, le besoin de produire des textes élaborés se faisant de plus en plus pressant: interrogation d'une banque de données, résultats d'un système expert, bureautique, etc. [...] Tout système qui produit un texte à partir d'une représentation formelle est du domaine de la génération automatique (<u>Revue québécoise de linguistique</u>, vol. 14, n° 2, 1985, p. 65).</p> <p>2. La génération part de représentations informatiques internes pour parvenir à des sorties linguistiques. Ses applications s'étendent de la synthèse de la parole à l'interrogation de bases de données (réponse "en clair"), voire à l'aide à la rédaction (<u>Langages</u>, n° 87, septembre 1987, p. 9).</p> <p>3. Une communication passe par les étapes suivantes: l'émetteur doit d'abord réunir un ensemble d'informations à priori éparses dans son cerveau pour en construire une série de messages ayant une structure particulière. C'est ce que nous appelons l'élaboration des messages. Il s'agit ensuite de synthétiser une expression écrite de ces messages en respectant les règles de grammaire et de morphologie. C'est tout ce groupe d'opérations que nous désignons par génération (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u>, 1989, p. 102).</p>
<u>Observation</u>	:	<p>1. Laurence Danlos utilise le terme <u>génération automatique de textes</u> et l'abréviation <u>génération automatique</u> dans son volume intitulé <i>Génération automatique de textes en langue naturelle</i> et Patrick Saint-Dizier et al. utilisent <u>génération</u> et <u>génération de texte</u>.</p> <p>2. La génération automatique de textes accepte comme entrée un arbre syntagmatique, une formule logique, un sous-graphe d'un réseau sémantique, une expression dans un langage d'interrogation de banque de données, une</p>

suite de déductions logiques ou même des données brutes telles que des cotes de la bourse ou des vitesses de vent.

3. Les premiers travaux en génération automatique de textes avaient pour but de tester une grammaire (Yngue 1961 et Friedman 1969).

4. Le programme de Friedman était destiné à servir d'aide aux linguistes pour l'étude des relations transformationnelles de la grammaire générative de Chomsky. Il ne s'agissait pas de transmettre un message mais de produire des phrases au hasard pour vérifier des règles de grammaire.

Illustration

:

voir traitement automatique des langues naturelles, 49.

<u>Entrée</u>	:	22. grammaire de cases, n. f., terme prop. syn. grammaire de slots, n. f. abrég. GS, n. f. slot grammar
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle syntaxique
<u>Définition</u>	:	Formalisme qui décrit les compléments de certains constituants (nom, adjectif et verbe) et qui fait état des relations de dépendance que ces constituants entretiennent avec leur dépendants (exemple: frapper: objet direct [concret]).
<u>Contexte</u>	:	<p>1. <u>Le passif dans les grammaires de slots.</u> Dans la GS de McCord, le passif est analysé selon une optique transformationnelle un peu particulière. Les transformations s'opèrent dans la liste des cases compléments du verbe. Lorsqu'un passif est identifié (auxiliaire être et désinence participiale du verbe), on met en jeu le prédicat <i>voix</i> qui effectue plusieurs opérations:</p> <ol style="list-style-type: none"> 1. chercher dans la liste des compléments possibles du verbe si l'un d'entre eux a la structure sémantique du sujet syntaxique du verbe; 2. vérifier que c'est dans ce cas un complément d'objet; 3. augmenter la liste des cases d'un complément d'agent ayant la valeur sémantique du sujet, agent du verbe (Alain Polguere, <u>Programmation logique des interfaces en langue naturelle</u>, 1984, p. 8). <p>2. Dans ce formalisme [grammaire de cases], chaque constituant possède un ensemble de fonctions possibles (appelées cases disponibles) qui sont remplies en cours d'analyse. Plutôt que de procéder de gauche à droite comme un réseau de transition augmenté ordinaire, une grammaire de cases contient une suite d'états qui, à partir de la tête d'une construction, travaillent dans les deux directions. Donc, par exemple, lorsqu'on rencontre un nom on fait appel au registre SN qui spécifie les mots et les syntagmes dont il peut être suivi à gauche (adjectifs, déterminants, etc.) et précédé à droite (syntagmes prépositionnels, propositions relatives, etc.) (traduit de Terry Winograd, <u>Language as a cognitive process</u>, 1985, p. 311).</p>
<u>Observation</u>	:	1. Dans la communauté scientifique le terme <u>grammaire de slots</u> est utilisé plus souvent que <u>grammaire de cases</u> qui est l'équivalent proposé ici par notre spécialiste. Dans le document d'Alain Polguere on rencontre l'abrég. <u>GS</u> .

2. La grammaire de cases n'utilise qu'un seul trait sémantique par case.

3. Elle contient un analyseur basé sur des cases et des règles qui procède de bas en haut et conserve une liste de cases disponibles pour chaque phrase construite. Il a également recours à des états comme le font les réseaux de transition augmentés. Mais l'utilisation des états dans les grammaires de cases est bien plus restreinte que dans ces derniers.

4. Les grammaires de cases ont été élaborées par M. C. McCord en 1980.

Voir aussi

:

case, 14.; complémentation, 15.; réseau de transition augmenté, 42.

<u>Entrée</u>	:	<p>23. grammaire de clauses définies, n. f. abrég. DCG, n. f. var. grammaire à clauses définies, n. f. grammaire DCG, n. f.</p> <p>definite clause grammar abrég. DCG</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Symbolisme permettant de formuler des règles de réécriture du type "indépendantes du contexte" et dont les règles sont traduites directement dans une forme exécutable par Prolog.
<u>Contexte</u>	:	<p>1. Nous présenterons dans cette section une description succincte du formalisme des grammaires de clauses définies (definite clause grammars, ou DCG) [...] En fait, la possibilité d'avoir comme symboles de grammaire des termes complexes plutôt que de simples atomes confère aux DCG suffisamment de puissance pour décrire les langages de type 0 (systèmes de réécriture non-limités). Bien qu'il ne soit pas encore établi si les langues naturelles entrent dans cette catégorie, le formalisme des DCG équivaut au modèle le plus puissant connu soit la machine de Turing (Bertrand Pelletier, <u>Système d'interrogation de banques de données en langue naturelle</u>, 1986, pp.6 et 9).</p> <p>2. Les grammaires à clauses définies ont été introduites par F. Pereira et D. Warren (Pereira 1980). Une règle des DCG est de la forme</p> $X ::= b$ <p>(Maryse Condillac, <u>Prolog: Fondements et applications</u>, 1986, p. 124).</p>
<u>Observation</u>	:	<p>1. Alain Polguere et Bertrand Pelletier utilisent l'abréviation <u>DCG</u> et le terme <u>grammaire de clauses définies</u> tandis que Maryse Condillac et Patrick Saint-Dizier et al. utilisent le terme <u>grammaire à clauses définies</u> et <u>grammaire DCG</u>.</p> <p>2. Les grammaires de clauses définies sont un cas particulier des grammaires de métamorphose. L'intérêt particulier de ces grammaires par rapport aux grammaires de métamorphose est leur facilité d'implantation et leur grande lisibilité sans perte significative de puissance expressive. Dans les grammaires de clauses définies, le contexte est obligatoirement exprimé par des arguments alors que dans les grammaires de métamorphose, on peut utiliser des symboles en partie gauche de règle.</p> <p>ex. $\infty \rightarrow b_1 \dots n$</p>

3. Dans une grammaire de clauses définies, les parties gauches sont séparées des parties droites par (->).

Voir aussi : grammaire de métamorphose, 24.

Illustration : Voici une petite grammaire de clauses définies, à gauche de sa traduction en Prolog, qui permet de reconnaître des phrases telles que:

Mon chien regarde ton chat.
Le chat poursuit le chien.

phrase -->	syntagme_nominal, syntagme_verbal.	phrase (S, S1) :- syntagme_nominal (S, S2), syntagme_verbal (S2, S1).
syntagme_nominal -->	déterminant, nom_commun.	syntagme_nominal (S, S1) :- déterminant (S, S2), nom_commun(S2, S1).
syntagme_verbal -->	verbe, syntagme_nominal.	syntagme_verbal(S, S1) :- verbe (S, S2), syntagme_nominal (S2, S1).
déterminant --> [le].		déterminant ([le S], S).
déterminant --> [mon].		déterminant ([mon S], S).
déterminant --> [ton].		déterminant ([ton S], S).
nom_commun -> [chat].		nom_commun ([chat S], S).
nom_commun -> [chien].		nom_commun ([chien S], S).
verbe/ -> [poursuit].		verbe ([poursuit S], S).
		verbe ([regarde S], S).

Grammaire de clauses définies

<u>Entrée</u>	:	24. grammaire de métamorphose, n. f. abrég. MG, n. f. metamorphosis grammar abrég. MG
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Symbolisme permettant de formuler des règles de réécriture du type "non contraintes" (de puissance équivalente aux machines de Turing) et dont les règles sont traduites directement dans une forme exécutable par Prolog.
<u>Contexte</u>	:	Les grammaires de métamorphose permettent d'écrire, en restant à un niveau plus conceptuel, des programmes Prolog et d'introduire des éléments indépendants du contexte par l'intermédiaire du symbole <i>a</i> en partie gauche de règle (Maryse Condillac, <u>Prolog: fondements et applications</u> , 1986, p.123).
<u>Observation</u>	:	<p>1. Maryse Condillac utilise l'abréviation <u>MG</u> pour <u>grammaire de métamorphose</u>.</p> <p>2. Il s'agit d'une axiomatisation en logique de premier ordre de la façon de traiter l'associativité de la concaténation lorsqu'on travaille sur des chaînes, afin de disposer en Prolog des facilités que donnent les systèmes-q, et d'obtenir aussi un outil très puissant pour tout traitement syntaxique et sémantique des langages.</p> <p>3. La grammaire de métamorphose permet d'exprimer l'extraposition à gauche de façon différente. Elle implique l'utilisation de règles dont le membre gauche est un non-terminal suivi d'une suite de symboles terminaux fictifs qui n'apparaissent pas dans le vocabulaire d'entrée.</p> <p>4. Une règle de grammaire de métamorphose est de la forme:</p> $X a :: b$ <p>où X est un non-terminal, a et b sont des suites de terminaux et de non-terminaux pouvant contenir des prédicats Prolog.</p> <p>5. Les grammaires de métamorphose permettent de décrire la formation de relatives en utilisant l'aspect indépendant du contexte qu'elles autorisent:</p> <p>sujet :: + t_sujet (trace du sujet)</p> <p>relative :: = pronom P</p> <p>pronom t_sujet :: = qui</p> <p>6. Les grammaires de métamorphose ont été mises au point par Alain Colmerauer en 1975.</p>
<u>Voir aussi</u>	:	extraposition à gauche, 20.; grammaire de clauses définies, 23.; système-q, 47.

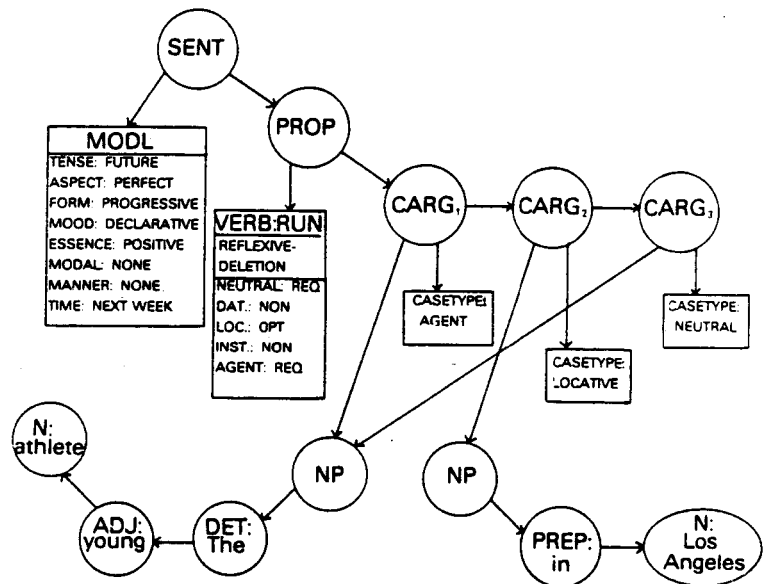
<u>Entrée</u>	:	25. grammaire des cas, n. f. syn. grammaire casuelle, n. f. case grammar
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle syntaxique
<u>Définition</u>	:	Approche sémantique qui considère le verbe comme le constituant principal de la phrase et qui étudie les relations des différents syntagmes nominaux avec le verbe (leur cas) (D'après Sylvia Pavel, Intelligence artificielle, <u>Cahiers de terminologie</u> , CT-31, 1987, p. 12).
<u>Contexte</u>	:	La grammaire des cas est une forme de grammaire transformationnelle [...] qui s'appuie sur le fait que la structure profonde d'une phrase simple est composée d'un verbe et d'un ou plusieurs syntagmes verbaux qui entretiennent une relation particulière avec le verbe (traduit de William B. Gevarter, <u>Artificial Intelligence, Expert Systems, Computer Vision and Natural Language Processing</u> , 1984, p. 114).
<u>Observation</u>	:	<p>1. Sylvia Pavel propose les termes <u>grammaire des cas</u> et <u>grammaire casuelle</u> dans son vocabulaire de l'intelligence artificielle</p> <p>2. La grammaire des cas est une extension des grammaires transformationnelles. Elle décrit la langue d'une perspective plus sémantique que ces dernières. Elle consiste à élargir la théorie de la grammaire générative transformationnelle en fournissant, outre une description des relations structurales existant à l'intérieur d'une phrase, des notions sur les relations fonctionnelles entre les différents syntagmes de la phrase.</p> <p>3. Dans la grammaire des cas, le verbe de la phrase est le gouverneur principal de la structure; le verbe principal de la proposition est le point autour duquel les autres syntagmes viennent se greffer. Les auxiliaires fournissent plus de détails sur la modalité.</p> <p>4. Prenons la grammaire des cas suivante: P-> +P M->temps, aspects (tous les modes) P-> V = C₁ ... C_n V-> courir, casser, mettre ... tous les verbes du vocabulaire C_i-> k + SN K-> (PREP) qu'on interprète comme suit: Une phrase P est constituée de la modalité M et de la proposition P à l'intérieur de laquelle la modalité consiste</p>

en une série de termes qui décrivent les aspects d'une phrase comme un tout. La proposition P se définit comme un verbe V et quelques cas où chacun représente un Kasus et un syntagme nominal. Le Kasus qui peut être nul est la préposition qui introduit le syntagme nominal et qui définit sa relation avec le verbe (COI ou COD).

5. La grammaire des cas a été mise au point par Charles J. Fillmore en 1967.

Voir aussi : grammaire sémantique, 33.; théorie de la dépendance conceptuelle, 48.

Illustration : Voici une grammaire des cas pour la phrase: "The young athlete will be running in Los Angeles next week."



Grammaire des cas

<u>Entrée</u>	:	<p>26. grammaire d'extraposition, n. f. abrég. GE, n. f. XG, n. f. grammaire XG, n. f.</p> <p>extraposition grammar abrég. XG</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Formalisme conçu pour le traitement des mouvements de constituants vers la gauche.
<u>Contexte</u>	:	<p>1. Dans le cadre des grammaires d'extraposition, Pereira (1982) propose que les quantificateurs nominaux (il ne considère que ce cas particulier) soient rattachés à une forme normale au SN dont la tête est la plus à droite (donc minimale et proximale) (<u>Revue québécoise de linguistique</u>, vol. 14, n° 2, 1985, pp. 24-25).</p> <p>2. Les grammaires d'extraposition [...] sont une extension des DCG [grammaires de clauses définies] permettant de rendre compte du déplacement et de l'élosion des constituants (Alain Polguere, <u>Programmation logique des interfaces en langue naturelle</u>, 1984, p. 9).</p> <p>3. Tout comme les DCG, les grammaires d'extraposition permettent la construction d'arbres syntagmatiques par l'intermédiaire des arguments. Cependant, à l'analyse d'une phrase selon une grammaire XG correspond un graphe d'analyse. (Bertrand Pelletier, <u>Système d'interrogation de banques de données en langue naturelle</u>, 1986, p. 9).</p> <p>4. La philosophie d'une GE est que certains éléments de la langue vont être considérés comme des marqueurs indiquant qu'une extraposition (extraction gauche) d'un élément a lieu et précisant la nature de l'élément déplacé (Alain Polguere, <u>Programmation logique des interfaces en langue naturelle</u>, 1984, p. 10).</p>
<u>Observation</u>	:	<p>1. Bertrand Pelletier utilise également l'abréviation anglaise <u>XG</u>.</p> <p>2. Les grammaires d'extraposition permettent de décrire certains phénomènes linguistiques comme les phrases interrogatives et les propositions relatives de façon claire et concise.</p> <p>3. Les grammaires d'extraposition peuvent contenir des règles indépendantes du contexte ou encore des règles de grammaire de clauses définies. On y retrouve des règles</p>

dont le membre gauche contient le symbole "..." qui permet de déplacer un constituant sur un intervalle de mots de longueur indéterminée.

4. Les grammaires d'extraposition ont des règles de la forme

$$S_1 S_2 X S_{k-1} S_k :: = R$$

où les S_i et R sont des suites de terminaux et de non-terminaux et X représente une suite arbitraire de symboles grammaticaux.

5. Dans la grammaire d'extraposition, le déplacement des constituants est permis par l'emploi d'une pile d'extraposition sur laquelle est placée la trace du constituant manquant à droite. Lors de l'analyse, lorsqu'un constituant est exigé dans une règle de grammaire, s'il ne se trouve pas dans la phrase à l'endroit attendu, l'analyseur vérifie si ce constituant ne figurerait pas sur le dessus de cette pile.

6. Considérons la structure de surface suivante:

L'homme que_i [_{ph} Pierre rencontra_i] était un politicien.

Dans cet exemple, t symbolise la trace du constituant manquant (l'homme), le pronom relatif *que* indique que la suite de la phrase contient une trace représentant un groupe nominal absent; le lien entre les deux est effectué par l'indice commun i . Par exemple, la règle

rel_marker ... trace --> rel_pronoun

indique qu'une suite quelconque de mots peut se retrouver entre le rel_marker et la trace. Ce déplacement permet de traiter une relative comme une phrase complète (à l'aide de la règle de grammaire définissant une phrase), mais à laquelle il manque un constituant sujet ou objet:

L'homme [Pierre rencontra l'homme] était un politicien.

7. Bien qu'elle décrit l'extraposition à gauche de façon économique, la grammaire d'extraposition est inflexible car les lacunes (gap) qui apparaissent dans le membre droit doivent être réécrites en ordre consécutif à la fin du membre droit et doivent également obéir à la contrainte nesting qui veut que deux lacunes soient entièrement indépendantes ou que l'une puisse contenir l'autre en entier.

8. Le formalisme des grammaires d'extraposition a été mis au point par Fernando Pereira en 1981.

Voir aussi : extraposition à gauche, 20.

Illustration :

```

sentence --> noun_phrase, verb_phrase.

noun_phrase --> determiner, noun, relative.
noun_phrase --> trace.

relative --> [].
relative --> open, rel_marker, sentence, close.

rel_marker ... trace --> rel_pronoun.

open ... close --> [].

noun --> [cat].
noun --> [mouse].
noun --> [fish].
verb --> [chased].
verb --> [likes].
verb --> [squeaks].
determiner --> [the].
rel_pronoun --> [that].

```

Grammaire d'extraposition pour les relatives

<u>Entrée</u>	:	27. grammaire discontinue, n. f. discontinuous grammar abrév. DG
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Formalisme qui permet le déplacement d'une suite inconnue de symboles à l'intérieur d'une règle.
<u>Contexte</u>	:	<p>1. [...] les grammaires discontinues peuvent être restreintes par un processus de filtrage par arbre dans le but de réduire leur puissance (Pascale Sébillot et Patrick Saint-Dizier, Un processus de filtrage par arbres syntaxiques dans les grammaires discontinues, <u>Proceedings Journées Prolog-CNET</u>, 1986, p. 471).</p> <p>2. Les grammaires discontinues sont une généralisation des grammaires d'extraposition parce qu'elles permettent l'extraposition à droite aussi bien que l'extraposition à gauche [...] (traduit de Veronica Dahl et Patrick Saint Dizier, <u>Constrained Discontinuous Grammars - A Linguistically Motivated Tool for Processing Language</u>, 1986, p. 2).</p> <p>3. Les grammaires discontinues permettent une description plus souple des phénomènes de déplacement que leur prédécesseur, les grammaires d'extraposition (Pereira 1981) (traduit de Patrick Saint-Dizier, Contextual Discontinuous Grammars, <u>Natural Language and Logic Programming 11</u>, 1988, p. 29).</p>
<u>Observation</u>	:	<p>1. Les grammaires discontinues possèdent des propriétés héritées des grammaires logiques. On peut citer:</p> <ul style="list-style-type: none"> -capacité d'une machine de Turing; -uniformité de formalismes pour les différents niveaux dans des systèmes de langue naturelle; -possibilités descriptives de très haut niveau (métrarègles); -facilité d'installation; -possibilité d'évolution des traits morphologiques constituants. <p>2. Elles possèdent également des propriétés qui leur sont propres:</p> <ul style="list-style-type: none"> -pouvoir expressif; -modularité. <p>3. Les règles des grammaires discontinues peuvent être considérées comme des métrarègles.</p> <p>4. La famille des grammaires logiques discontinues qui, à l'origine, étaient appelées grammaires lacunaires ont été</p>

prises au point par Veronica Dahl et H. Abramson en 1984.

Voir aussi : extraposition à gauche, 20.; grammaire discontinue contextuelle, 28.; grammaire discontinue statique, 29.; grammaire lacunaire, 31.

Illustration : Soit le langage $L(G) = \{a^n b^m c^n d^m \mid m, n \in W\}$

La grammaire le générant est:

$S \rightarrow as \ bs \ cs \ ds.$
 $as \ gap(G) \ cs \rightarrow [a] \ as \ gap(G) \ [c] \ cs.$
 $as \ gap(G) \ cs \rightarrow gap(G).$
 $bs \ gap(G) \ ds \rightarrow [b] \ bs \ gap(G) \ [d] \ ds.$
 $bs \ gap(G) \ ds \rightarrow gap(G).$

Grammaire discontinue

<u>Entrée</u>	:	28. grammaire discontinue contextuelle, n. f., terme prop. var. grammaire Contextuelle Discontinue, n. f. contextual discontinuous grammar abrég. CDG
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Formalisme permettant d'exprimer une relation de dépendance entre règles selon laquelle l'application d'une première règle entraîne l'application subséquente des autres règles qui lui sont reliées par la règle de grammaire discontinue contextuelle.
<u>Contexte</u>	:	Les gapping grammars [grammaires lacunaires] (Dahl et Abramson 1984) puis les grammaires Contextuelles Discontinues [grammaires discontinues contextuelles] (Saint-Dizier 88) ont été définies dans le but d'augmenter la puissance d'expression des grammaires d'extraposition en autorisant des mouvements plus variés. Le formalisme de base [grammaire lacunaire] a été augmenté de mécanismes de filtrage de la forme de l'arbre syntaxique en cours de construction qui vont bloquer l'application de certaines transformations (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u> , 1989, p. 46).
<u>Observation</u>	:	<p>1. Le terme <u>grammaire discontinue contextuelle</u> a été formé à partir de la base <u>grammaire discontinue</u> dont il constitue une extension.</p> <p>2. Une règle discontinue contextuelle a la forme d'un ensemble de règles indépendantes du contexte: $\{ a \rightarrow a_1, b \rightarrow b_1, \dots, n \rightarrow n_1 \}$ <p>Dans cette règle, a, b, ..., n sont symboles non terminaux et a_1, b_1, \dots, n_1 sont des suites de symboles non terminaux et terminaux. Un ensemble contenant une règle simple: $\{ a \rightarrow a_1 \}$ équivalait à une règle indépendante du contexte.</p> <p>3. L'ensemble des notations utilisées dans une grammaire discontinue contextuelle indique clairement que les règles de type indépendante du contexte peuvent être appliquées dans n'importe quel ordre (mais toutes les règles doivent être utilisées) lors de la construction de l'arbre syntaxique. Cependant, des restrictions de préséance linéaire peuvent être formulées pour assurer un certain ordre partiel.</p> <p>4. Les grammaires discontinues contextuelles ne sont rattachées à aucune théorie linguistique. Leur stratégie consiste plutôt à adopter certains aspects significatifs des</p> </p>

théories linguistiques afin d'accroître leur représentation linguistique et de rendre leurs règles plus faciles à utiliser par les rédacteurs de grammaires.

5. Les grammaires discontinues contextuelles ont été mises au point par Patrick Saint-Dizier en 1988 et sont inspirées des grammaires discontinues statiques qui elles-mêmes constituent une classe restrictive des grammaires discontinues.

Voir aussi : grammaire d'extraposition, 26.; grammaire discontinue, 27.; grammaire discontinue statique, 29.; grammaire lacunaire, 31.; préséance linéaire, 37.

Illustration : Voici un bref exemple de grammaire discontinue contextuelle qui reconnaît le langage:

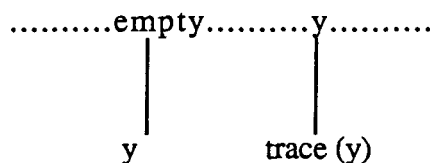
$$L(G) = \{a^n b^m d^m ; n, m \in \mathbb{N}\}$$

qui s'écrit de la façon suivante:

$s \rightarrow as, bs, cs, ds.$
 $as \rightarrow [a], as / cs \rightarrow [c], cs.$
 $bs \rightarrow [b], bs / ds \rightarrow [d], ds.$
 $as \rightarrow [a].$
 $bs \rightarrow [b].$
 $cs \rightarrow [c].$
 $ds \rightarrow [d].$

Grammaire discontinue contextuelle

<u>Entrée</u>	:	29. grammaire discontinue statique, n. f., terme prop. static discontinuous grammar abrég. SDG
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Formalisme qui, dès le tout début de la phase de traitement, élimine la génération de phrases incorrectes à l'aide d'un filtre.
<u>Contexte</u>	:	La grammaire discontinue statique utilise des graphes d'analyse plutôt que des arbres d'analyse comme le fait la grammaire discontinue pour structurer les données.
<u>Observation</u>	:	<p>1. Le terme <u>grammaire discontinue statique</u> a été formé à partir de la base <u>grammaire discontinue</u> dont il constitue une extension.</p> <p>2. Dans la grammaire discontinue statique, la structure d'arbre est préservée pendant que la puissance de type 0 est coupée. Ceci est rendu possible en immobilisant les skips et en permettant aux constituants qui les entourent d'être déplacés.</p> <p>3. En grammaire discontinue statique, une règle-type s'écrit comme suit: empty, skip, y --> y, skip, trace (y) pour obtenir le déplacement d'une catégorie y vers une position vide. Les phrases incorrectes sont éliminées par l'application d'un filtre à une étape ultérieure. Il est à noter que le déplacement est décrit de façon statique en réécrivant "empty" en "y" et "y" en trace (y)". Lorsque la sous-chaîne espacée ne bouge plus, deux branches peuvent décrire l'expansion de la règle, chacune étant suivie d'un seul parent:</p>



4. La grammaire discontinue statique est issue de récents travaux sur l'utilisation des grammaires discontinues pour générer le langage naturel à l'intérieur de la théorie gouvernement et liage. Ils ont donné l'idée de conserver la puissance de type 0 des grammaires discontinues tout en conservant la simplicité des arbres d'analyse indépendants du contexte; plutôt que de permettre aux espaces d'être repositionnés, les règles doivent demeurer en place jusqu'à ce que les constituants qui les entourent soient réécrits ou déplacés.

5. La grammaire discontinue statique a été mise au point par Veronica Dahl en 1986.

Voir aussi : grammaire discontinue, 27.; grammaire discontinue contextuelle, 28.

<u>Entrée</u>	:	30. grammaire d'unification, n. f. unification grammar
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Formalisme dont les règles doivent satisfaire à des contraintes d'unification pour être appliquées.
<u>Contexte</u>	:	<p>Indiquons seulement qu'une grammaire d'unification se distingue d'une grammaire générative [Chomsky 1957] entre autres par le fait que le même statut est attribué aux fonctions thématiques et aux fonctions ou catégories syntaxiques. De plus, une grammaire d'unification permet d'indiquer le focus d'une phrase. Cette information est utilisée pour certains choix syntaxiques [Mckeown 1983]. Ainsi, le choix entre la construction active ou passive d'un verbe repose sur la règle suivante:</p> <p>Règle 2: si le focus est sur l'AGENT, construire la phrase à l'actif; si le focus est sur l'OBJET, construire la phrase au passif.</p> <p>Cette règle s'appuie sur le fait qu'il est généralement admis que le focus d'une phrase doit être en position sujet.</p> <p>(1) Max a abandonné la maison [focus sur Max] (2) La maison a été abandonnée par Max [focus sur la maison]</p> <p>(Laurence Danlos, <u>Génération automatique de textes en langue naturelle</u>, 1985, pp. 67-68).</p>
<u>Observation</u>	:	<ol style="list-style-type: none"> 1. Les grammaires d'unification adjoignent une description fonctionnelle à chaque mot et à chaque phrase. Une description fonctionnelle est une liste de noms appelés attributs auxquels on a associé une valeur. 2. Dans une grammaire d'unification chaque règle possède un membre droit et un membre gauche séparés par une flèche. Le membre gauche est une description fonctionnelle tandis que le membre droit est une expression booléenne représentant soit: <ol style="list-style-type: none"> 1. des descriptions fonctionnelles; 2. une expression identifiée par le mot Composant suivi d'une suite d'un ou plusieurs chemins; 3. une expression d'ordonnancement identifiée par le mot Ordre suivi d'une suite composée de chemins et de zéro ou plusieurs situations de... 3. Dans une grammaire d'unification, les chemins jouent à peu près le même rôle que les variables en programmation logique: relier des ensembles de valeurs unifiées. 4. Dans ce formalisme, il est possible de travailler avec des descriptions linguistiques sans tenir compte de la

combinaison exacte de chaque expression possible. De plus, il est nécessaire d'inclure dans une expression les parties qui peuvent limiter ses possibilités d'unification avec d'autres expressions.

5. Dans une grammaire d'unification, aucune description fonctionnelle n'est décrite en entier et il est toujours possible de la compléter avec des détails supplémentaires. Ainsi, il est possible de compléter une description fonctionnelle construite par unification à l'aide de la composante syntaxique d'une grammaire, de l'information sémantique ou par unification.

6. La grammaire d'unification a été mise au point par Martin Kay en 1979.

Voir aussi : unification, 50.

Illustration :

1. $\{\} \rightarrow (\{ \text{Head} = \text{NONE} \} \text{ and } \{ \text{Cat} = \text{noun} \} \quad \{ \text{Cat} = \text{verb} \}) \text{ or } (\sim \{ \text{Head} = \text{NONE} \} \text{ and } \{ \text{Cat} = \text{NONE} \})$
2. $\{ \text{Head} = \{ \text{Cat} = \text{Verb} \} \} \rightarrow \text{Order: } \langle \text{Head} \rangle \dots$
3. $\sim \{ \text{Head} = \{ \text{Cat} = \text{Verb} \} \} \rightarrow \text{Order: } \dots \langle \text{Head} \rangle$
4. $\{ \text{Head} = \{ \text{Head} = \{ \text{Cat} = \text{Verb} \} \} \} \rightarrow \{ \text{Head} = \{ \text{Head} = \{ \text{Subj} = \{ \text{Head} = \{ \text{Art} = \{ \text{Case} = \text{Nom} \} \} \} \} \} \} \text{ and } \text{Constituent: } \langle \text{Head Head Subj} \rangle$
5. $\left[\text{Head} = \left[\begin{array}{l} \text{Cat} = \text{Verb} \\ \text{Obj} = \{ \text{Cat} = \text{NP} \} \end{array} \right] \right] \rightarrow \{ \text{Head} = \{ \text{Obj} = \{ \text{Head} = \{ \text{Art} = \{ \text{Case} = \text{Acc} \} \} \} \} \} \text{ and } \text{Constituent: } \langle \text{Head Obj} \rangle$
6. $\{ \text{Head} = \{ \text{Art} = \{ \text{Cat} = \text{Det} \} \} \} \rightarrow \text{Constituent: } \langle \text{Det} \rangle$

Grammaire d'unification simple

<u>Entrée</u>	:	<p>31. grammaire lacunaire, n., f., terme prop. syn. gapping grammar, n. f. abrév. GG, n. f. syn. grammaire de gap</p> <p>gapping grammar abrév. GG</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Formalisme conçu pour effectuer des déplacements de constituants ou pour rendre compte de dépendances lointaines entre constituants de la phrase.
<u>Contexte</u>	:	<p>1. Les gapping grammars utilisent un symbole non-terminal gap (X) pour référer dans une phrase une sous-chaîne X dont la structure et la composition ne sont pas d'un intérêt immédiat, bien que se trouvant entre des objets intéressants pour la règle en cours de description (Maryse Condillac, <u>Prolog: fondements et applications</u>, 1986, p. 127).</p> <p>2. Les gapping grammars (Dahl et Abramson 1984) puis les grammaires Contextuelles Discontinues (Saint-Dizier 88) ont été définies dans le but d'augmenter la puissance d'expression des grammaires d'extraposition en autorisant des mouvements plus variés (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u>, 1989, p. 46).</p>
<u>Observation</u>	:	<p>1. Maryse Condillac dans son ouvrage intitulé <i>Prolog: fondements et applications</i> utilise le terme anglais <u>gapping grammar</u> et l'abréviation <u>GG</u> alors que Sylvia Pavel dans sa terminologie de l'intelligence artificielle propose <u>grammaire discontinue</u> comme équivalent français qui ne couvre pas tout à fait la même notion. Bertrand Pelletier utilise pour sa part le terme <u>grammaire de gap</u>. Le terme <u>grammaire lacunaire</u> semble indiqué à la place des formes partiellement ou totalement empruntées qu'on retrouve dans la littérature.</p> <p>2. Les grammaires lacunaires sont à l'origine des grammaires discontinues, bien que moins performantes.</p> <p>3. Les règles d'une grammaire lacunaire sont considérées comme des métarègles qui représentent un ensemble (possiblement infini) de règles de grammaire ordinaires. Elles permettent d'indiquer l'endroit où les sous-chaînes indéfinies ou intermédiaires peuvent être laissées en attente dans le membre gauche et réordonnées par l'application de règles permettant l'analyse ultérieure de ces sous-chaînes par d'autres règles.</p>

4. La grammaire lacunaire est un formalisme puissant qui peut être utilisé pour décrire de façon élégante et concise des phrases complexes écrites en langue naturelle aussi bien que des langages formels.

5. Dans une grammaire lacunaire, l'extraposition à gauche de l'adjectif dans:

"Cette jolie demeure agréable..." qui devient "Agréable, cette jolie demeure..." est représentée par la règle:

Det Lacune (X) Adj. -> Adj [,] Det Lacune (X).

La structure de base apparaît dans le membre gauche de la règle tandis que la structure avec le déplacement apparaît dans le membre droit.

6. La grammaire lacunaire a été mise au point par Veronica Dahl et H. Abramson en 1984. C'est une généralisation des grammaires de métamorphose, de clauses définies et des grammaires d'extraposition.

Voir aussi

: extraposition à gauche, 20.; grammaire discontinue, 27.

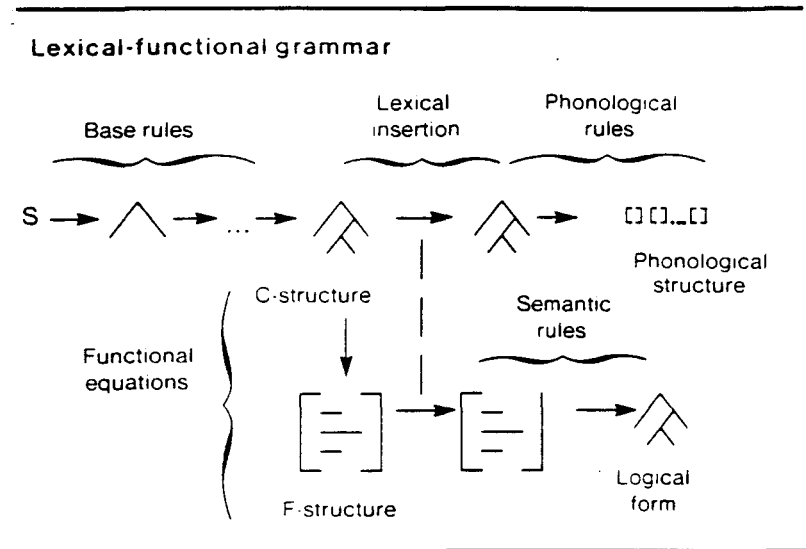
<u>Entrée</u>	:	32. grammaire lexicale fonctionnelle, n. f. abrég. GLF, n. f. lexical-functional grammar abrég. LFG
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle syntaxique
<u>Définition</u>	:	Formalisme qui établit une structure fonctionnelle pour chaque phrase dans laquelle les fonctions grammaticales (sujet, objet...) sont étroitement associées aux mots ou aux groupes de mots qui remplissent ces fonctions.
<u>Contexte</u>	:	<p>1. Dans des théories comme la théorie gouvernement-liage (GL) ou la grammaire lexicale fonctionnelle (GLF) la structure syntagmatique (S-structure dans GL et C-structure dans GLF) joue un rôle central moins important que dans la grammaire syntagmatique généralisée (GSG), dans la mesure où l'on retrouve dans ces théories (GL et GLF) plusieurs autres types (ou niveaux) de représentation (D-structure pour la théorie GL, F-structure pour la théorie GLF) (<u>Revue québécoise de linguistique</u>, vol. 14, n° 2, 1985, p. 14).</p> <p>2. En effet, d'autres formalismes tels que les grammaires lexicales fonctionnelles [Bresnan 82] ou les grammaires généralisées à structure syntagmatique [Gazdar 82] ont aussi été implantées en PROLOG, mais en considérant celui-ci comme un simple langage de programmation (Maryse Condillac, <u>Prolog: fondements et applications</u>, 1986, p. 123).</p> <p>3. Récemment, une nouvelle théorie grammaticale, la grammaire lexicale fonctionnelle, a été conçue dans le but évident de répondre à la double demande d'analyse et d'apprentissage (traduit de <u>American Journal of Computational Linguistics</u>, vol. 8, n° 3-4, juillet-décembre 1982, p. 98).</p>
<u>Observation</u>	:	<p>1. La théorie de la grammaire lexicale fonctionnelle est basée sur les relations grammaticales c'est-à-dire qu'elle a recours aux fonctions (sujet, objet...) pour décrire les opérations syntaxiques plutôt que de les définir de façon structurale comme le fait la théorie transformationnelle.</p> <p>2. La grammaire lexicale fonctionnelle ne contient aucune règle transformationnelle. Les règles syntagmatiques et les règles lexicales en particulier sont les pivots de cette théorie.</p> <p>3. La grammaire lexicale fonctionnelle ne comporte pas de transformation. Le travail anciennement imputé aux</p>

transformations comme le passif est assuré par l'information emmagasinée dans les entrées lexicales jumelées aux items lexicaux.

4. La grammaire lexicale fonctionnelle a été mise au point par R. Kaplan et J. Bresnan en 1982.

Illustration

:



Grammaire lexicale fonctionnelle

<u>Entrée</u>	:	33. grammaire sémantique, n. f. semantic grammar
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Formalisme doté d'un lexique et d'un ensemble de règles de réécriture qui utilise des classes sémantiques pour déterminer les catégories lexicales.
<u>Contexte</u>	:	<p>1. Les grammaires sémantiques prennent en considération la structure des questions sans s'intéresser explicitement aux catégories syntaxiques classiques telles que noms, verbes... Les catégories intervenant dans les règles sont plutôt fortement orientées vers le domaine considéré. L'intervention sémantique de la grammaire rend l'analyse des questions fortement déterministe et partant, très efficace (<u>R.A.I.R.O. Informatique / Computer Science</u>, vol. 14, n° 2, 1980, p. 137).</p> <p>2. Une grammaire sémantique ressemble à une grammaire syntaxique à l'exception que les catégories de mots (noms, verbes) sont remplacées par des classes sémantiques (bateau, propriétés du bateau) (traduit de <u>Byte</u>, vol. 12, n° 14, décembre 1987, p. 227).</p>
<u>Observation</u>	:	<p>1. Cette approche a l'avantage de donner un accès direct aux composantes sémantiques d'une phrase, de simplifier et de raccourcir le traitement de façon substantielle.</p> <p>2. Dans des domaines pratiques portant sur des domaines limités, il est souvent utile d'utiliser des composantes sémantiques à la place des constituants syntaxiques comme par exemple les syntagmes nominaux et prépositionnels. Il est cependant difficile de transférer les règles de réécriture d'un domaine d'application à un autre.</p> <p>3. La grammaire sémantique comporte trois aspects: l'analyse du langage naturel, la recherche de réponses dans la base et la génération de réponses.</p> <p>4. La grammaire sémantique peut accepter des réponses incorrectes. En effet, elle peut toujours refléter la construction grammaticale de la phrase sans cependant être pointilleuse vis-à-vis des problèmes tels que l'accord du genre et du nombre ou la présence des constituants.</p> <p>5. Une des applications les plus marquantes de la grammaire sémantique, le système LADDER, a été élaborée par Gary G. Hendrix pour le compte de la marine américaine en 1978.</p>

Illustration

:

Rewrite rules:

 $S \rightarrow \text{what is SHIP-PROPERTY of SHIP?}$ $\text{SHIP-PROPERTY} \rightarrow \text{the SHIP-PROP} \mid \text{SHIP-PROP}$ $\text{SHIP-PROP} \rightarrow \text{speed} \mid \text{length} \mid \text{draft} \mid \text{beam} \mid \text{type}$ $\text{SHIP} \rightarrow \text{SHIP-NAME} \mid \text{the fastest SHIP2} \mid \dots$ $\text{SHIP-NAME} \rightarrow \text{Kennedy} \mid \text{Kitty Hawk} \mid \dots$

Sample input:

WHAT IS THE LENGTH OF THE KENNEDY?

**Règles de réécriture utilisées par la grammaire
sémantique du système LADDER**

<u>Entrée</u>	:	<p>34. grammaire syntagmatique généralisée, n. f. abrég. G.S.G., n. f. syn. grammaire de structure de phrase généralisée, n. f.</p> <p>generalized phrase structure grammar abrég. GPSG</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle syntaxique
<u>Définition</u>	:	Grammaire syntagmatique augmentée de métarègles et d'une sémantique compositionnelle qui décompose chaque phrase de la chaîne d'entrée en constituants et en syntagmes selon des règles de dominance immédiate et de préséance linéaire.
<u>Contexte</u>	:	<p>En grammaire syntagmatique généralisée, la structure syntagmatique (ou S-structure) constitue le seul niveau de représentation syntaxique. [...] Comme [la] GSG est une théorie monostratale (i.e. à un seul niveau de représentation syntaxique), le passage y constitue une application quasi-fonctionnelle (à cause de l'existence de chaînes ambiguës). Le problème de passage peut donc y être vu simplement comme celui de la définition d'une (quasi-) fonction de transformation (transduction) d'une chaîne c_1 (élément d'un ensemble L) en un arbre a_1, élément d'un ensemble $A(L)$ étiqueté par des catégories complexes (généralisées) représentant sa structure syntagmatique (enrichie):</p> <p>Passage: $L \Rightarrow A(L)$ $c_1 \Rightarrow a_1$</p> <p>(Revue québécoise de linguistique, vol. 14, n° 2, 1985, p. 14).</p>
<u>Observation</u>	:	<p>1. Sylvia Pavel utilise le terme <u>grammaire de structure de phrase généralisée</u> dans son vocabulaire intitulé <i>Intelligence Artificielle</i>. Jean-Yves Morin emploie le terme <u>grammaire syntagmatique généralisée</u> et l'abréviation <u>GSG</u>.</p> <p>2. Les grammaires syntagmatiques généralisées servent à décrire des phénomènes linguistiques comme la pronominalisation, l'interrogation, la relative, les formes active et passive.</p> <p>3. Elle est une généralisation des grammaires indépendantes du contexte. Elle est augmentée de métarègles qui génèrent de nouvelles règles syntagmatiques à partir des règles existantes.</p>

4. Une règle de grammaire syntagmatique généralisée a la forme:

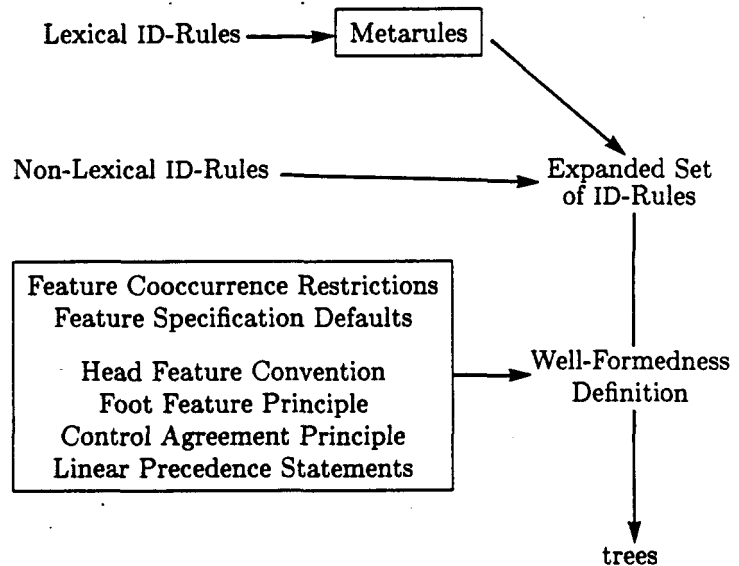
$A \rightarrow B, D, C$ (dominance immédiate)

$B < C, B < D$ (précedence linéaire)

5. Les grammaires syntagmatiques généralisées ont été introduites par G. Gazdar et al. en 1980.

Voir aussi : dominance immédiate, 19.; préséance linéaire, 37.

Illustration



Grammaire syntagmatique généralisée

<u>Entrée</u>	:	35. interface en langue naturelle, n. f. natural language interface abrév. NLI'S
<u>Domaine</u>	:	traitement automatique des langues naturelles/applications
<u>Définition</u>	:	Système qui permet le passage d'un énoncé en langue naturelle vers une commande directement interprétable en langage spécialisé.
<u>Contexte</u>	:	Construire un [sic] interface en langue naturelle (LN), c'est programmer le passage d'un énoncé en LN vers une commande directement interprétable (Alain Polguere, <u>Programmation logique des interfaces en langue naturelle</u> , 1984, p. 3).
<u>Observation</u>	:	<p>1. L'interface en langue naturelle comporte trois opérations (traduction, analyse et évaluation) qui permettent de dégager la structure propre de l'énoncé, de lui associer une représentation jugée adéquate en fonction de la tâche à effectuer et, enfin, d'évaluer cette représentation c'est-à-dire exécuter la commande ou la requête.</p> <p>2. On dit d'une interface qu'elle est transparente lorsqu'elle réagit vite et correctement et qu'elle tolère des fautes mineures et qu'elle est instructive lorsqu'elle est incapable de reconnaître une requête.</p> <p>3. Le principal avantage des interfaces en langue naturelle est leur facilité d'utilisation car l'utilisateur communique avec l'ordinateur au moyen de questions plutôt qu'avec un langage spécialisé. L'utilisateur peut ainsi apprendre progressivement les contraintes imposées par le système et si l'ordinateur ne comprend pas sa question, il peut la reformuler jusqu'à ce que l'ordinateur l'accepte.</p> <p>4. Le programme SHRDLU mis au point par Terry Winograd en 1972 est l'un des premiers exemples d'interface.</p>
<u>Illustration</u>	:	<div style="display: flex; justify-content: space-between; align-items: flex-start;"> <div style="text-align: right;"> <p>ÉNONCE LN</p> <p>STRUCTURE</p> <p>FORMULE DE REPRÉSENTATION- DU SENS</p> </div> <div style="text-align: center;"> <p>-----></p> <p><----- -----></p> <p>-----></p> </div> <div style="text-align: left;"> <p>ANALYSE</p> <p>TRADUCTION</p> <p>ÉVALUATION et EXÉCUTION</p> </div> </div> <p style="text-align: center;">Structure interne d'une interface en langue naturelle</p>

<u>Entrée</u>	:	36. prédicat, n. m. predicate
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Sens élémentaire attribué aux principales composantes de la phrase.
<u>Contexte</u>	:	<p>1. Un même prédicat [...] doit être interprété de façon différente selon le type sémantique de son sujet: si le sujet renvoie à un objet individualisable et déterminé, alors il doit être interprétable comme une "appartenance à une certaine classe"; si par contre, le sujet renvoie à du générique, alors il est interprétable comme "l'inclusion dans une certaine classe" (<u>Langages</u>, n° 87, septembre 1987, p. 65).</p> <p>2. Les prédicats sont tous les termes que l'on ne considère que rattachés à des éléments, tandis que les éléments, eux, ont un statut autonome, c'est-à-dire peuvent faire l'objet pour eux-même, d'une recherche documentaire (<u>Linguisticae Investigationes</u>, vol. 4, n° 1, 1980, p. 45).</p> <p>3. Un prédicat est une fonction qui prend l'une des valeurs: vrai ou faux. Le nombre de ses arguments dépend de sa nature. Les prédicats peuvent être liés par les opérations logiques: N (négation), ^ (conjonction), V (disjonction) ... Les arguments d'un prédicat peuvent être des variables qui sont quantifiées par V (quel que soit) ou (il existe). $(\forall X) P(x)$ indique que le prédicat P est toujours vrai, quelle que soit la valeur donnée à la variable x. $(\exists X) P(X)$ signifie qu'il existe au moins un [sic] valeur de x telle que P(x) soit vrai (Jacques Pitrat, <u>Textes, ordinateurs et compréhension</u>, 1985, p. 29).</p>
<u>Observation</u>	:	Les prédicats sont sémantiquement interprétés comme dénotant des relations entre des individus. Ces relations peuvent toutefois être prises comme des éléments primitifs.
<u>Voir aussi</u>	:	calcul des prédicats, 13.

<u>Entrée</u>	:	<p>37. préséance linéaire, n. f. abrév. PL, n. f. syn. précédence linéaire, n. f.</p> <p>linear precedence abrév. LP</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/structure des données
<u>Définition</u>	:	Relation binaire d'ordonnancement entre les membres droits des règles de réécriture dans une grammaire syntagmatique généralisée.
<u>Contexte</u>	:	<p>Les règles de structure syntagmatique peuvent être condensées par le mécanisme de dominance immédiate et de précédence linéaire. Si nous avons les règles suivantes:</p> <p style="padding-left: 40px;">A --> B, C, D. A --> B, D, C. A --> D, B, C.</p> <p>Il peut être intéressant de les regrouper en une seule règle qui énoncerait que A domine immédiatement B, C, D, ces trois constituants étant dans un ordre quelconque à l'exception de B qui doit toujours apparaître avant C. Ceci est noté:</p> <p style="padding-left: 40px;">A --> B; C; C; D; B < C.</p> <p>La relation --> spécifie la dominance et le symbole ; est un séparateur de symboles qui n'impose aucun ordre des fils. L'inégalité < impose que B apparaisse avant C (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u>, 1989, p. 38).</p>
<u>Observation</u>	:	<p>1. Patrick Saint-Dizier et coll. utilisent le terme <u>précedence linéaire</u> et Jean-Yves Morin l'abréviation <u>PL</u>.</p> <p>2. Dans une règle de réécriture de type P --> SN ... SV, le constituant SN précède linéairement le constituant SV.</p> <p>3. Une règle de préséance linéaire a la forme V < S ce qui signifie que le verbe doit précéder le complément S. La règle de préséance linéaire s'applique dans la grammaire au complet à chaque règle qui comporte les catégories V et S.</p> <p>4. La préséance linéaire indique un ordonnancement unique du membre droit des règles contenues dans une grammaire.</p> <p>5. La préséance linéaire est un formalisme mis au point par Stewart Shieber en 1984.</p>
<u>Voir aussi</u>	:	dominance immédiate, 19.; grammaire syntagmatique généralisée, 34.
<u>Illustration</u>	:	voir grammaire syntagmatique généralisée, 34.

<u>Entrée</u>	:	38. primitive sémantique, n. f. abrév. P. S., n. f. semantic primitive
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Mécanisme de réduction qui permet la représentation sémantique d'une langue naturelle par le biais d'un algorithme de traduction.
<u>Contexte</u>	:	<p>1. Les secondes [les primitives sémantiques] ne sont pas relatives à une langue déterminée, mais au langage: ce sont des entités langagières, qu'on le situe dans le métalangage linguistique ou dans l'appareil cognitif. Voire, pour certains auteurs, les primitives ne sont plus relatives au langage et deviennent des catégories (au sens philosophique du terme) purement conceptuelles: ce point de vue n'est pas rare en I.A., où l'on recherche des modèles valides aussi bien pour le traitement du langage que pour celui de la vision ou de la représentation des événements (<u>Langages</u>, n° 87, septembre 1987, p. 82).</p> <p>2. Une représentation à base de P.S. [primitives sémantiques] peut être vue comme un langage permettant de décrire les éléments d'un univers sémantique en utilisant un nombre fini de significations élémentaires, elles-mêmes indéfinissables: les P.S. (<u>Langages</u>, n° 87, septembre 1987, p. 48).</p>
<u>Observation</u>	:	<p>1. Les primitives sémantiques, en dépendance conceptuelle, sont les entités les plus élémentaires pour décrire le monde.</p> <p>2. Le nombre de primitives sémantiques doit être inférieur au nombre de mots du vocabulaire du langage afin d'éviter la redondance: les inférences qui sont réalisées à partir d'un concept ne mettent généralement en jeu qu'un sous-ensemble de propriétés de celui-ci. Si ce sous-ensemble est commun à plusieurs autres concepts, on peut alors rattacher les inférences, non plus, au concept lui-même dans sa totalité, mais au nouveau concept caractérisé par ce sous-ensemble de propriétés.</p> <p>3. Utiliser une décomposition en primitive sémantique, c'est faire apparaître les niveaux intermédiaires de cette hiérarchie. Ainsi, par exemple, "courir" et "marcher" partagent les inférences liées à l'action "se déplacer" et leur expression en termes de cette dernière action permet de ne conserver qu'un seul exemplaire des inférences qui lui sont rattachées.</p>

4. Les primitives sémantiques doivent permettre d'exprimer toutes les distinctions sémantiques utilisées. Il doit y avoir, dans l'application, équivalence entre ce que l'on représente et ce qui est représenté.

5. Une primitive sémantique ne peut être définie à l'aide d'autres primitives (qu'il y ait ou non cyclicité).

6. La définition de formes canoniques dans les représentations est un moyen d'éliminer la redondance. Cercone souligne que dans les primitives des dépendances conceptuelles de Schank, "ingérer" n'est en rien plus primitif que "manger" ou "boire". Si l'on souhaite factoriser les inférences qui s'y rattachent, sans pour cela utiliser "ingérer" dans la représentation de "manger" et de "boire", il suffit d'explicitier une règle du genre "si manger alors ingérer" et une règle similaire concernant "boire". Les inférences liées à "ingérer" pourront alors ne figurer qu'une seule fois dans le système.

Voir aussi

:

théorie de la dépendance conceptuelle, 48.

<u>Entrée</u>	:	39. principe de compositionnalité, n. f. compositionality principle syn. Frege's principle
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Processus d'interprétation sémantique des constituants selon lequel on peut faire correspondre une opération sémantique à chaque opération syntaxique.
<u>Contexte</u>	:	Des règles de composition sémantique sont associées aux structures syntaxiques pour décrire comment la représentation interne d'une structure se déduit des représentations internes des sous-structures qui la composent (principe de compositionnalité). Lorsque cette déduction n'aboutit pas à une solution unique, des éléments extérieurs sont alors mis à contribution. Ces éléments sont, en particulier, la position d'autres mots dans la phrase, le contexte local à l'énonciation ou des règles de raisonnement de bon sens (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u> , 1989, p. 8).
<u>Observation</u>	:	1. Le principe de compositionnalité constitue une façon de limiter le nombre de règles sémantiques car il recourt à des définitions récursives au niveau sémantique et syntaxique. Ce principe s'appuie sur le fait que le sens d'un tout joue un rôle sur la signification des parties et sur la façon dont elles sont réunies ce qui signifie que les règles sémantiques sont construites de concert avec les règles syntaxiques: à chaque expression de base correspond une interprétation et à chaque règle syntaxique est associée une règle sémantique. Cette dernière va assigner, à chaque expression complexe formée par une règle syntaxique, une interprétation qui est déterminée par les énoncés simples qui servent d'entrée aux règles syntaxiques.

<u>Entrée</u>	:	40. registre, n. m. register
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Variable des réseaux de transition augmentés contenant des informations globales sur la phrase analysée.
<u>Contexte</u>	:	L'intérêt de ces registres est double: nous pouvons d'abord y stocker des informations que nous pourrions tester ensuite. Par exemple, nous pourrions nous rappeler que nous avons passé par un certain arc A en liant cet arc à une action qui va donner la valeur "vrai" à un registre R créé à cet effet. [...] Le deuxième intérêt des registres est de stocker des morceaux de la représentation du sens que des actions ultérieures sauront utilisées pour construire des parties plus importantes à partir des éléments partiels déjà rassemblés (Jacques Pitrat, <u>Textes, ordinateurs et compréhension</u> , 1985, p. 87).
<u>Observation</u>	:	<p>1. Les registres ressemblent aux variables contenues dans un langage de programmation car ils sont identifiés par un nom et ils emmagasinent de l'information.</p> <p>2. Les syntagmes majeurs de la phrase analysée sont mis en attente dans les registres jusqu'à ce que la structure de la phrase soit déterminée. Par exemple, lorsqu'on est en présence d'un verbe, il est emmagasiné dans le registre V; de la même façon, tous les constituants d'un syntagme nominal sont gardés en mémoire dans le registre SN.</p>
<u>Voir aussi</u>	:	réseau de transition augmenté, 42.

<u>Entrée</u>	:	41. réseau de transition, n. m. transition network syn. transition net
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Ensemble fini d'états (appelés aussi noeuds) reliés par des arcs sur lesquels se trouvent inscrites des conditions portant sur des éléments de la séquence [phrase] analysée, éléments dont la lecture est une condition à la transition d'un état à l'autre dans le réseau (Pierre Plante, <u>Le Deredec, logiciel pour l'analyse de contenu des textes</u> , 1985, p. 12).
<u>Contexte</u>	:	En même temps qu'il analyse une phrase, le réseau de transition peut amasser de l'information sur la formation des mots et remplir des cases dans un schéma associé à chaque mot. Il peut donc déterminer si les syntagmes nominaux sont singuliers ou pluriels même si le nom se rapporte à une personne et le genre nécessaire pour obtenir une structure profonde (traduit de William B. Gevarter, <u>Artificial Intelligence, Expert Systems, Computer Vision and Natural Language Processing</u> , 1985, p. 116).
<u>Observation</u>	:	<p>1. La définition de Pierre Plante servait à décrire le terme <u>réseau de transition augmenté</u> mais elle définit mieux le <u>terme réseau de transition</u>.</p> <p>2. Il existe plusieurs types de réseaux de transition dont la complexité et l'efficacité varient. Par exemple, la machine à état fini est un dispositif théorique qui débute dans un état particulier et qui change d'état lorsque interviennent des conditions spécifiques. Il est fini dans le sens que l'état suivant peut être déterminé à partir de l'état courant et des conditions qui peuvent engendrer une transition peu importe où en est le processus. Le nombre d'états est restreint et donc considéré comme fini.</p> <p>3. Le réseau le plus élémentaire, le graphe de transition à état fini est composé d'un groupe de noeuds reliés par des arcs dirigés. Chaque noeud représente un état faisant partie d'une machine à état fini et les arcs représentent les transitions d'un état à un autre. Chaque transition correspond à un mot de la chaîne.</p> <p>4. Le réseau de transition équivaut à une grammaire de type 3 ou grammaire ordinaire et comporte des règles. Il</p>

est utilisé pour générer ou reconnaître des chaînes de mots.

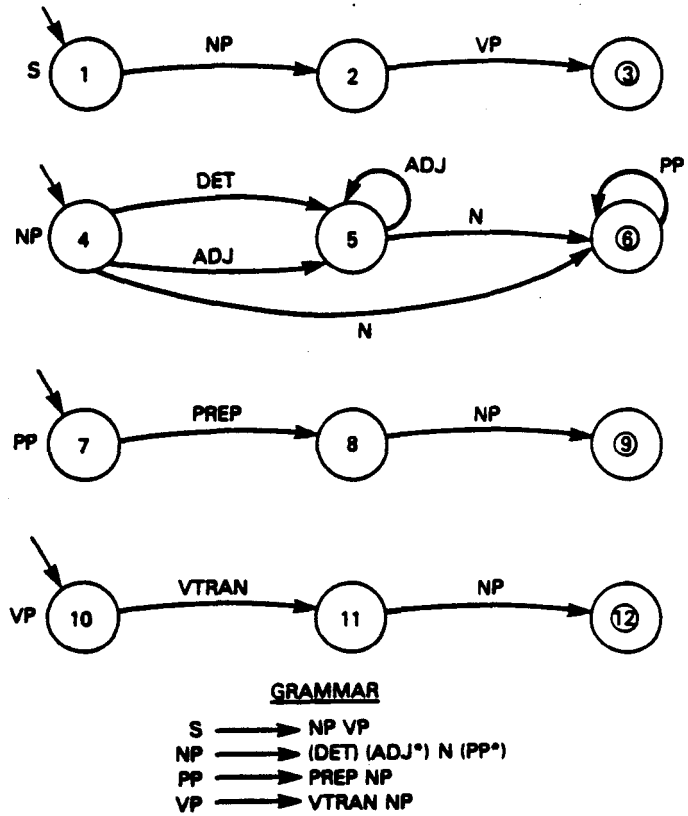
Voir aussi

:

réseau de transition augmenté, 42.; réseau de transition récursif, 43.

Illustration

:



Réseau de transition simple

<u>Entrée</u>	:	<p>42. réseau de transition augmenté, n. m. abrév. A.T.N., n. m. syn. réseau de transition enrichi, n. m.</p> <p>augmented transition network abrév. ATN</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/phases de traitement
<u>Définition</u>	:	Réseau de transition de la puissance d'une grammaire transformationnelle qui permet d'accumuler de l'information dans des registres et de formuler des conditions en tenant compte des informations contenues dans ces registres.
<u>Contexte</u>	:	<p>1. [...] Les réseaux de transition augmentés (A.T.N.) expriment la structure des phrases et des groupes de mots sous forme d'une suite explicite de changements d'états à suivre par la machine (<u>Pour la Science</u>, n° 85, novembre 1984, p. 99).</p> <p>2. [...] Ce système [réseau de transition augmenté] peut être décrit comme une généralisation des grammaires context-free pour atteindre la puissance d'une grammaire transformationnelle (<u>TAUM 1973</u>, p. 10).</p> <p>3. Aux environs de 1970, W. A. Woods a développé une procédure d'analyse grammaticale automatique fondée sur la notion de "réseau de transition enrichi" ("Augmented Transition Network" ou A.T.N.). Au départ, un A.T.N. est un réseau de transition bien connu en théorie des automates; il s'agit d'un graphe dont les noeuds correspondent à des états d'un automate et les arcs à des changements d'états (<u>Linguisticae Investigationes</u>, vol. 4, n° 1, 1980, p. 48).</p>
<u>Observation</u>	:	<p>1. Laurence Danlos dans son ouvrage intitulé <i>Génération automatique de textes en langues naturelles</i> et Jacques-Henri Jayez dans son article intitulé <i>Un survol des recherches sur le traitement automatique du langage naturel</i> utilisent le terme <u>réseau de transition enrichi</u> plutôt que <u>réseau de transition augmenté</u>.</p> <p>2. Les principaux avantages des réseaux de transition augmentés sont les suivants:</p> <ol style="list-style-type: none"> 1) la simplicité, 2) le pouvoir génératif, 3) l'efficacité de la représentation, 4) l'efficacité des opérations, 5) la flexibilité au niveau de l'expérimentation.

3. Les réseaux de transition augmentés sont une extension des réseaux de transition rékursifs. Ils comportent trois caractéristiques additionnelles: des registres qui peuvent emmagasiner des conditions ou de l'information globale, sans tenir compte du sous-réseau particulier qui est traité; des conditions qui permettent aux arcs d'être sélectionnés si le registre répond à certaines conditions; et des actions qui permettent aux arcs de modifier la structure des données.

4. Les réseaux de transition augmentés ne se limitent pas seulement à accepter ou à rejeter chaque mot ou chaque phrase mais procèdent plutôt à la construction de la structure de la phrase à l'intérieur de l'arbre d'analyse de façon que chaque mot entré soit assigné à un élément du réseau.

5. Les syntagmes majeurs de la phrase sont mis en attente dans les registres jusqu'à ce que la structure de la phrase soit déterminée. Par exemple, lorsqu'on est en présence d'un verbe, il est emmagasiné dans le registre V; de la même façon, tous les constituants d'un syntagme nominal sont gardés en mémoire dans le registre SN.

6. Dans un réseau de transition augmenté, les arcs peuvent être étiquetés non seulement par des mots, des classes de mots et des non-terminaux mais aussi par des tests arbitraires qui dépendent de l'état des registres globaux.

7. Le réseau de transition augmenté est un formalisme conçu par William Woods en 1970 à l'intérieur duquel chaque arc est assorti d'un ensemble de conditions et d'actions que l'automate doit vérifier et accomplir au cours de chaque changement d'état ou transition.

8. À l'origine, les réseaux de transition augmentés ont été conçus pour remplacer les transformations comme méthode d'analyse.

Voir aussi

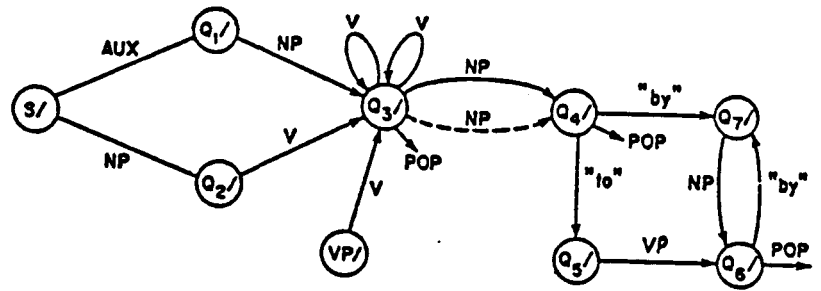
:

registre, 40.; réseau de transition, 41.; réseau de transition rékursif, 43.

Illustration

:

Voici un réseau de transition augmenté tiré de Woods (1970):

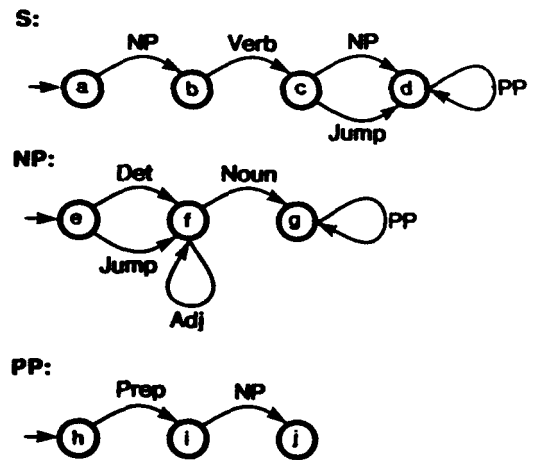


Réseau de transition augmenté simple

<u>Entrée</u>	:	43. réseau de transition récursif, n. m. abrév. RTN, n. m. recursive transition network abrév. RTN
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Réseau de transition dans lequel il est possible, lors de l'exécution d'un sous-réseau A, de faire appel à nouveau à ce même sous-réseau.
<u>Contexte</u>	:	1. Réseau issu des automates d'états finis qui peut prendre en considération l'aspect récursif de certaines décisions. En puissance d'expression, le RTN est équivalent aux grammaires à contexte libre de l'informatique classique (Sylvia Pavel, Intelligence Artificielle, <u>Les cahiers de terminologie</u> , CT-31, 1987, p. 84). 2. Les réseaux de transition récursifs (RTN par la suite, pour «Recursive Transition Network») sont un moyen de spécifier une grammaire. [...] (Alain Bonnet, <u>L'intelligence artificielle: promesses et réalités</u> , 1984, p. 53).
<u>Observation</u>	:	1. Un réseau de transition récursif est composé d'un ensemble de réseaux étiquetés et équivaut à un réseau de transition simple sauf que chaque réseau porte une étiquette et que chaque arbre est étiqueté par un mot, une catégorie lexicale ou une catégorie syntaxique qui elle-même étiquette un réseau dans la grammaire. 2. Le réseau de transition récursif contient un état de départ et un ou plusieurs états finaux. Tous les états et les arcs qui les relient sont identifiés. Il vérifie les arcs de chacun des noeuds ou états; si un arc est un symbole terminal (catégorie lexicale), le contrôle se déplace vers le noeud situé à la tête de l'arc et le processus recommence. Si l'arc est un nom d'état, (comme un syntagme nominal), il représente un réseau de transition récursif complet. 3. Les réseaux de transition récursifs contiennent également une nouvelle sorte d'arc, l'arc jump (identifié par l'étiquette JUMP) qui peut être emprunté même s'il ne peut être combiné à un mot de l'entrée.
<u>Voir aussi</u>	:	réseau de transition, 41.

Illustration**Context-free grammar**

$S \rightarrow NP VP$
 $NP \rightarrow Det NP2$
 $NP \rightarrow NP2$
 $NP2 \rightarrow Noun$
 $NP2 \rightarrow Adj NP2$
 $NP2 \rightarrow NP2 PP$
 $PP \rightarrow Prep NP$
 $VP \rightarrow Verb$
 $VP \rightarrow Verb NP$
 $VP \rightarrow VP PP$

Recursive transition networks

Réseau de transition récursif et grammaire indépendante du contexte.

<u>Entrée</u>		44. réseau sémantique, n. m. abrév. R.S., n. m. semantic network syn. semantic net
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Ensemble de noeuds représentant des concepts exprimés par des mots en langue naturelle et reliés entre eux par des arcs représentant les relations sémantiques que ces concepts entretiennent entre eux.
<u>Contexte</u>	:	1. Un réseau sémantique est une représentation du sens sous la forme de concepts reliés entre eux par des graphes étiquetés par les concepts et les arcs reliant les noeuds sont étiquetés par des relations sémantiques reliant les concepts (<u>Méthodologie Informatique Philosophie</u> , n° 2, 1985, p. 66). 2. Le rôle des réseaux sémantiques est de représenter le sens d'une phrase sans véhiculer quelque information syntaxique que ce soit: c'est ce qui leur mérite leur appellation. Il ne faut pas les confondre avec les réseaux sémantiques du même nom qu'on retrouve ailleurs en intelligence artificielle. Ce sont des objets linguistiques; ils permettent d'exprimer une phrase dans un langage pictural, peu importe que cette dernière soit ambiguë ou incohérente (Michel Boyer, <u>Génération de phrases à partir de réseaux sémantiques</u> , mai 1984, pp. 6-7).
<u>Observation</u>	:	1. Introduits par des psychologues, les réseaux sémantiques correspondent à un besoin de classification de concepts et d'établissement d'un modèle du "monde" (une situation, un acte, un discours), à l'aide de ces concepts et de liaisons installées entre eux. En tant que tels, les réseaux sémantiques ne sont plus guère utilisés qu'en analyse du langage naturel mais ils servent par contre de support à d'autres formalismes: les systèmes de classification utilisés dans certains systèmes experts font appel à eux. 2. Graphiquement, un réseau sémantique apparaît comme un ensemble de points et d'arcs orientés dans lequel: a) un point représente un concept qu'on repère par le mot qui le désigne habituellement ainsi les concepts d'"animal", de "personne" ou de "général" sont représentés par: ANIMAL.....HOMME.....GENERAL Tous les concepts sont généralement définis dans un lexique.

b) un arc peut:

soit représenter une relation (sémantique) R entre deux concepts X et Y et constituer un triplet:

XRY

-soit représenter une relation sémantique entre deux triplets et constituer dès lors un nouveau triplet, susceptible d'entrer à son tour dans d'autres relations et ainsi de suite.

Ainsi une phrase comme:

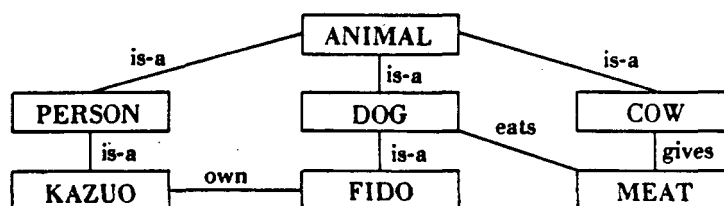
Le vaillant général commande la grande armée

MOD COMMANDE MOD
(GENERAL-->VAILLANT)----->(ARMEE-->GRANDE)

3. Les réseaux sémantiques ont été élaborés par Quillian en 1968 dont le but principal était de simuler la compréhension humaine en concevant un formalisme, fondé sur le sens des mots, représentant la structure de la mémoire sémantique de l'être humain.

Illustration

Voici un exemple emprunté à Terry Winograd qui permet d'engendrer des phrases comme: quelqu'un nourrit son chien de viande de boeuf, Kasuo possède Fido, ou certains animaux sont des vaches, etc.



Réseau sémantique

<u>Entrée</u>	:	45. schéma, n. m. syn. frame, n. m. cadre, n. m. prototype, n. m. frame
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Représentation du sens contenant les composantes essentielles du sens d'un mot y compris des valeurs par défaut.
<u>Contexte</u>	:	<p>1. La notion de prototype ou de schéma (frame en anglais) répond au désir de permettre la classification d'objets par la description d'objets typiques plutôt que par l'énumération des caractéristiques strictement partagées par tous les objets de chaque classe. Il semble que l'homme organise lui-même ses connaissances en classification de prototypes, laissant la possibilité à des objets de se distinguer par certaines propriétés du prototype de leur classe. Un prototype définit donc une classe d'objet de manière très souple; il comporte une partie fixe correspondant à tout ce qui est commun à tous les éléments (instances) de la classe et une partie variable, obligatoire ou facultative, différente pour chaque instance (<u>La Recherche</u>, vol. 16, n° 170, octobre 1985, p. 1247).</p> <p>2. Les schémas reposent sur la nécessité de raisonner par analogie en imposant des stéréotypes à la réalité. Ce mode de raisonnement, qui permet d'arriver à des conclusions sur la base de similitudes partielles, s'oppose à la déduction logique où tout est vrai ou faux (ou incertain) (Laurence Danlos, <u>Génération automatique de textes en langues naturelles</u>, 1985, p. 60).</p> <p>3. Un schéma-type est défini par: (i) une dénomination d'un "objet"; (ii) une liste d'attributs (slots) qui sont autant de propriétés caractérisant cet objet. (<u>Langages</u>, n° 87, septembre 1987, p. 89).</p>
<u>Observation</u>	:	<p>1. Sylvia Pavel propose dans son vocabulaire sur l'intelligence artificielle le terme <u>cadre</u> tandis que Jean-Pierre Desclés et François Rastier proposent le terme <u>schéma</u> et Hervé Gallaire propose le terme <u>prototype</u>.</p> <p>2. Un schéma structure les données en deux niveaux: les données qui sont toujours réalisées dans la situation en jeu et les données qui sont spécifiques à la situation.</p>

3. Un schéma possède des valeurs qui, par défaut, peuvent être précisées ou non.

4. Un schéma ressemble un peu aux noeuds des réseaux sémantiques, ou à un acteur indépendant. Il ne correspond pas à un fait isolé comme dans une représentation logique mais bien à un ensemble d'informations appartenant à un même concept.

5. Voici un ensemble de schémas pour les concepts "donner" et "payer":

Donner est une action
acteur: personne
bénéficiaire: personne
objet: objet physique

Payer est un don

objet: argent
raison: dette

On peut donc conclure que payer est une forme de don.

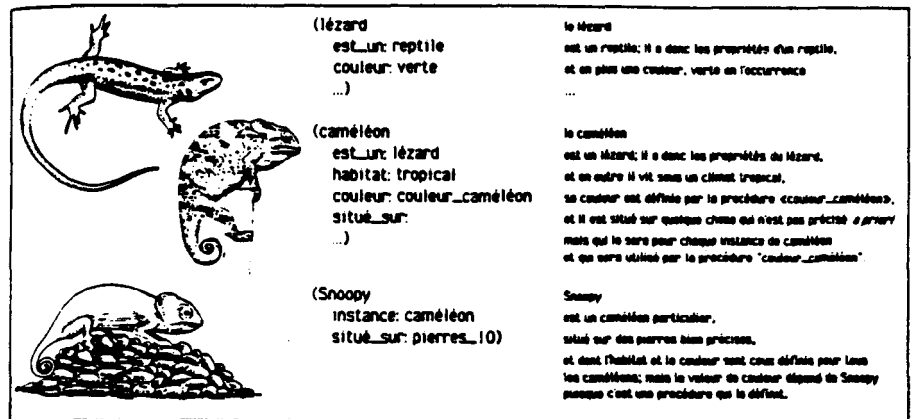
6. Le concept de schéma a été introduit par Marvin Minsky en 1975 qui le considère comme une approche pour résoudre des problèmes découlant de l'élaboration de théories sur la vision, la mémoire, le raisonnement logique et la compréhension du langage.

Voir aussi

: case, 14.; réseau sémantique, 44.; script, 46.

Illustration

:



**Schéma pour les sous-classes de reptiles :
lézards et caméléons**

<u>Entrée</u>	:	46. script, n. m. syn. scénario, n. m. script
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Série de processus typiques propres à un domaine sémantique (par exemple, le scénario du restaurant au sein du domaine // alimentation//).
<u>Contexte</u>	:	<p>1. Les scripts sont des séquences d'événements correspondant au déroulement typique d'une action parfaitement banale, déroulement que l'auteur d'un texte n'a aucune raison d'explicitier (ainsi, l'auteur est fondé à admettre que chacun de ses lecteurs sait que "se rendre à son bureau" se décompose fréquemment en: a) sortir de son logement, b) rejoindre à pied un moyen de transport, c) l'utiliser, d) recommencer s'il y a lieu les pas b) c), e) rejoindre à pied son bureau (<u>La Recherche</u>, vol. 16, n° 170, octobre 1985, pp. 1204-1205).</p> <p>2. Un script est une sorte de schéma utilisé pour décrire les structures textuelles (<u>Langages</u>, n° 87, septembre 1987, p. 95).</p>
<u>Observation</u>	:	<p>1. Daniel Kayser et François Rastier emploient le terme <u>script</u> et Sylvia Pavel propose le terme <u>scénario</u> en guise de synonyme.</p> <p>2. Le système des scripts est l'approche la plus en vogue pour exprimer la pragmatique.</p> <p>3. Le script permet d'emmagasiner les moments marquants d'un événement et fournit la possibilité de prévoir des actions liées de loin avec cet événement tout en organisant des actions qui lui sont spécifiques. Il est composé de plusieurs parties: les structures qui permettent de créer le cadre de l'action, les conditions de départ du script, les résultats du script une fois complété, les scènes qui ont eu lieu pendant le script.</p> <p>4. La notion de MOP (Memory Organization Packet), mise au point par Schank (Yale), a généralisé celle de script en permettant de mettre en commun des éléments appartenant à des scripts différents: voyager en métro, en avion ou en train correspondent à des scripts distincts, car le déroulement des épisodes est différent; mais il y a dans chacun de ces scripts des informations sur les titres de transport, les correspondances, les retards, etc. Il serait inefficace de répéter les éléments communs à chacun des</p>

scripts et de ne pas tirer parti d'une structure partielle commune.

5. Ce formalisme conçu pour la structure du récit a été mis au point par Roger Schank et Robert Abelson en 1977 qui soutiennent que la mémoire humaine ne consiste pas en propositions emmagasinées en fonction des mots qu'elles contiennent mais qu'elle est orientée sur le concept plutôt que sur les mots.

6. Par exemple, si on demande à quelqu'un, "Qui était ton amie en 1968?", il ne fera pas une rétrospective de toutes ses amies pour déterminer celle qu'il fréquentait en 1968; en se rappelant ces événements, il va établir quels participants y prenaient place et sélectionnera alors un participant qui était une amie.

Voir aussi : conceptualisation, 16.; schéma, 45.

Illustration : Un script pour un restaurant peut contenir certains faits élémentaires comme:

S-client

Conditions d'entrée:

S a faim.

S a de l'argent

Résultats:

S a moins d'argent.

S n'a pas faim

S est satisfait (facultatif)

Script du restaurant

<u>Entrée</u>	:	47. système-q, n. m. var. système-Q, n. m. Q-system
<u>Domaine</u>	:	traitement automatique des langues naturelles/logiciels
<u>Définition</u>	:	Formalisme de puissance équivalente aux grammaires formelles de type 0 qui permet d'effectuer des transformations sur des graphes de chaînes.
<u>Contexte</u>	:	<p>1. Un système-q n'est rien d'autre qu'un ensemble de règles permettant de transformer un tel [sic] graphe en un graphe orienté (Alain Colmerauer, <u>Les systèmes-q ou un formalisme pour analyser ou synthétiser des phrases sur ordinateur</u>, 1970, p. 4).</p> <p>2. Ainsi les systèmes-Q par exemple opèrent sur des Q-graphes dont chaque branche est étiquetée par une arborescence simplement étiquetée (<u>COLING 84</u>, 1984, p. 12).</p>
<u>Observation</u>	:	<p>1. Le terme <u>système-q</u> s'écrit avec un q minuscule dans le document de Alain Colmerauer alors que partout ailleurs il s'écrit avec un Q majuscule.</p> <p>2. Le système-q a été mis au point par Alain Colmerauer dans le cadre du projet de Traduction automatique de l'Université de Montréal.</p> <p>3. Les systèmes-q sont réversibles ce qui signifie qu'un même système-q peut être utilisé pour décrire une transformation et la transformation inverse comme, par exemple l'analyse et la synthèse des phrases. Leur réversibilité permet également de retrouver les données à partir des résultats.</p> <p>4. Plusieurs systèmes-q peuvent être enchaînés les uns à la suite des autres. Chacun prend comme données les résultats du précédent.</p> <p>5. Les membres gauches et les membres droits sont structurés exactement de la même façon.</p> <p>6. Le même système-q, mais inversé permet d'analyser ses paraphrases et de retrouver la phrase normalisée qui leur correspond.</p> <p>7. Un système-q peut aussi être simplifié. Les règles prennent alors la forme $a_1+a_2+...+a_p=b_1+b_2+...+b_q$.</p>

$a_1+a_2+\dots+a$ correspond au membre gauche et $b_1+b_2+\dots+b_q$ correspond au membre droit.

8. Le système-q simplifié s'applique sur un graphe de chaînes donné et le transforme en un nouveau graphe de chaînes. La transformation se fait en deux parties soit par l'addition d'un certain nombre de flèches au graphe, et par la suppression de certaines flèches jugées inutiles.

<u>Entrée</u>	:	<p>48. théorie de la dépendance conceptuelle, n. f. abrég. TDC syn. dépendance conceptuelle, n. f. abrég. D. C., n. f.</p> <p>conceptual dependancy theory abrég. CDT syn. conceptual dependancy</p>
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Représentation sémantique qui consiste à évoquer un concept pour que de lui-même en évoque d'autres qui dépendent de lui.
<u>Contexte</u>	:	<p>1. La théorie de la dépendance conceptuelle a été développée par Robert Schank afin de pouvoir représenter les phrases simples du langage par des schémas montrant bien les relations sémantiques entre les différents concepts intervenants (Alain Bonnet, <u>L'intelligence artificielle: promesses et réalité</u>, 1984, p. 41).</p> <p>2. L'unité de base [de la dépendance conceptuelle] est la conceptualisation qui comprend une action et un certain nombre de liens qui interviennent dans cette action. Un objet peut être un objet physique: Pierre, un chien, ou une autre conceptualisation (Jacques Pitrat, <u>Textes, ordinateurs et compréhension</u>, 1985, p. 33).</p> <p>3. Fondamentalement, la D. C. est "language free", soit indépendante de tout langage utilisé pour le dialogue homme-machine; elle ne fait pas appel aux mots du langage courant, mais seulement aux concepts traités par la pensée; ces concepts sont représentés dans ce qui suit par des "mots" en majuscules, entourés par des astérisques (ex.: *ATRANS*) (<u>Linguisticae Investigationes</u>, vol. 4, n° 1, 1984, p. 56).</p> <p>4. Conceptual dependancy theory *CDT; théorie de la dépendance conceptuelle * TDC (Sylvia Pavel, Intelligence artificielle, <u>Les cahiers de terminologie</u>, CT-31, 1987, p. 18).</p>
<u>Observation</u>	:	<p>1. La théorie de la dépendance conceptuelle correspond à une représentation par graphes dans lesquels:</p> <p>- les noeuds peuvent être de quatre types différents:</p>

PP: objet du monde réel,
 ACT: actions du monde réel,
 PA: attributs d'objets,
 AA: attributs d'actions,

qui, grossièrement, correspondent dans le langage courant, aux concepts recouverts par les noms, les verbes, les adjectifs et les adverbes respectivement.

- les flèches servent à exprimer les relations dans lesquelles les concepts dépendent les uns des autres; il y a 16 relations de dépendance parmi lesquelles on peut citer:

2. Avec la théorie de la dépendance conceptuelle, Schank propose une approche purement sémantique combinant un nombre très restreint de primitives sémantiques ces dernières étant les composantes les plus élémentaires utilisées pour décrire le monde.

3. La théorie de la dépendance conceptuelle ressemble à la grammaire des cas car la représentation du sens d'une phrase repose sur son action et inclut un certain nombre de cases qui relient les autres parties de la phrase à cette action. Cependant dans la dépendance conceptuelle, l'action n'est pas représentée par le verbe mais plutôt par l'interrelation d'un ensemble d'actions primitives qui représentent un concept inclus dans le sens du verbe. En d'autres termes, l'action est représentée par les concepts contenus dans le verbe plutôt que par les items lexicaux, en l'occurrence le mot.

4. En dépendance conceptuelle, deux phrases avec la même signification ont la même représentation interne que les mots soient disposés dans le même ordre ou non. Dans les deux phrases

Jacques casse la fenêtre.
 La fenêtre est cassée par Jean

l'action est représentée par les concepts qui composent le verbe casser. Jean est la personne qui exécute l'action, et la fenêtre est l'objet de l'action. La représentation serait donc la même pour les deux actions.

5. Dans la théorie de la dépendance conceptuelle de Schank, *ATRANS* est un concept très général réalisable en donner, rendre, acheter, etc. Pour Schank, il y a 11 actions de base de ce type:
 FAIRE, *BOUGER*, *INGERER*, etc..

Voir aussi

:

conceptualisation, 16.; primitive sémantique, 38.

Illustration

: Voici Margie, un exemple de représentation utilisant la théorie de la dépendance conceptuelle, mise au point par Schank en 1974:

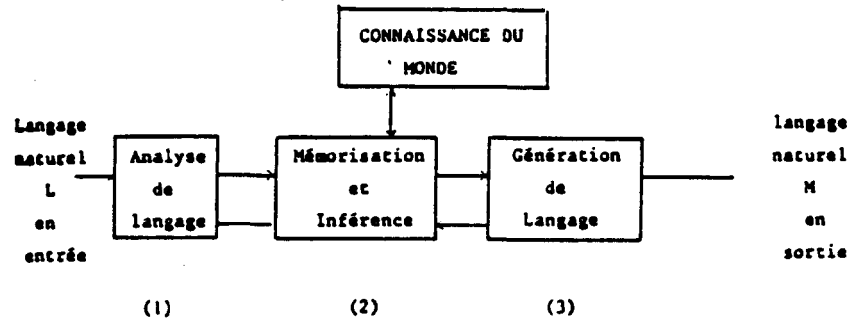
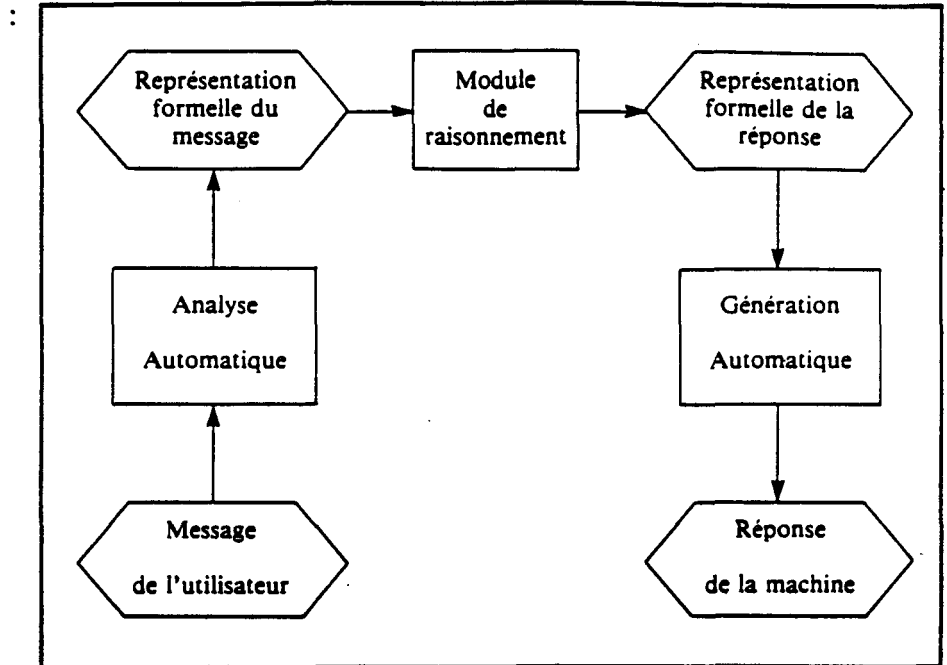


Schéma général de Margie

<u>Entrée</u>	:	<p>49. traitement automatique des langues naturelles, n. m. abrég. traitement des langues naturelles, n. m. abrég. TLN, n. m. syn. traitement automatique du langage naturel, n. m. abrég. traitement du langage naturel, n. m.</p> <p>natural language processing syn.: NLP</p>
<u>Domaine</u>	:	intelligence artificielle/traitement automatique des langues naturelles
<u>Définition</u>	:	Capacité d'un ordinateur à traiter le même langage que les humains utilisent dans leur discours normal.
<u>Contexte</u>	:	<p>1. Rappelons que les différentes applications informatiques liées au traitement des langues naturelles font appel aux ressources de la linguistique formalisée [...] (<u>Terminogramme</u>, n° 46, janvier 1988, p. 11).</p> <p>2. Il existe plusieurs niveaux dans le traitement du langage naturel: la phonétique, la morphologie, la lexicologie, la syntaxe, la sémantique et la pragmatique (Annie Gal, Guy Lapalme et Patrick Saint-Dizier, <u>Prolog pour l'analyse automatique du langage naturel</u>, 1989, p. 7).</p> <p>3. Le traitement du langage naturel [...] a été développé par A. Bonnet avec le programme BAOBAB [...] (<u>Technique et science informatiques</u>, vol. 1, n° 2, p. 32).</p>
<u>Observation</u>	:	<p>1. Le terme <u>traitement automatique des langues naturelles</u> n'apparaît jamais en entier dans la littérature. Ce sont plutôt les formes réduites qui sont utilisées. Jacques-Henri Jayez utilise à la fois <u>traitement automatique du langage naturel</u> et <u>traitement automatique des langues naturelles</u> dans son article intitulé <i>Un survol de recherches sur le traitement automatique du langage naturel</i>.</p> <p>2. En traitement automatique des langues naturelles, l'analyse consiste habituellement à combiner les symboles d'une phrase en un groupe de mots qui peut être remplacé par un symbole plus général. Ce nouveau symbole peut à son tour être combiné en un autre groupe, et ainsi de suite, jusqu'à ce qu'une structure correcte apparaisse.</p> <p>3 Le traitement automatique des langues naturelles allie l'aspect pratique de l'intelligence artificielle à l'aspect théorique de la linguistique.</p> <p>4. Les premiers travaux en traitement automatique des langues naturelles remontent au décryptage des codes secrets (cryptographie) lors de la Seconde Guerre</p>

Mondiale qui avaient montré que certaines manipulations purement statistiques pouvaient faire apparaître des propriétés aidant à comprendre des messages transmis en langue naturelle.

Illustration



Dialogue entre l'homme et la machine

<u>Entrée</u>	:	50. unification, n. f. unification
<u>Domaine</u>	:	traitement automatique des langues naturelles/modèle sémantique
<u>Définition</u>	:	Opération qui ramène deux structures différentes en un unificateur commun en complétant l'une ou l'autre des structures.
<u>Contexte</u>	:	<p>1. L'unification est l'opération de base de Prolog. C'est elle qui permet de tester, à l'aide de substitutions si deux arbres ou deux sous-arbres sont identiques. L'unification permet aussi d'accéder aux différents sous-constituants d'un terme, par l'emploi d'inconnues qui vont s'unifier avec une partie d'un terme (Patrick Saint-Dizier, <u>Initiation à la programmation en Prolog</u>, 1987, p. 38).</p> <p>2. L'unification est l'action de substituer des termes à des variables qui, lorsque appliquées à deux expressions, les transforme en une seule et même expression (traduit de Veronica Dahl, Hiding Complexity from the Casual Writer of Parsers, <u>Natural Language Understanding and Logic Programming</u>, 1985, p. 9).</p>
<u>Observation</u>	:	<p>1. L'unification compare deux expressions et détermine si elles décrivent la même chose. Si non, elle l'indique tout simplement et si c'est le cas, elle construit une nouvelle expression, en général plus spécifique que les expressions originales parce qu'elle regroupe tous les détails contenus dans ces deux expressions.</p> <p>2. Voici un exemple d'unification:</p> <div style="text-align: center;"> SN SV trait: humain : : concret trait: : humain </div> <p>L'unification de ces deux syntagmes devient possible en complétant les traits de SV. Donc SV devient:</p> <div style="text-align: center;">SV: => humain : : concret</div>
<u>Voir aussi</u>	:	grammaire d'unification, 29.

4. INDEX FRANÇAIS

- A -

analyse ascendante, 1
 analyse descendante, 2
 analyse déterministe, 3
 analyse en parallèle, 4
 analyse en série, 5
 analyse lexicale, 6
 analyse morphologique, 7
 analyse non déterministe, 8
 analyse pragmatique, 9
 analyse sémantique, 10
 analyse syntaxique, 11
 analyseur syntaxique, 12
 A.T.N., 42

- C -

cadre, 45
 calcul des prédicats, 13
 case, 14
 complémentation, 15
 conceptualisation, 16
 contrainte sémantique, 17
 contrainte syntaxique, 18

- D -

D.C., 48
 DCG, 24
 dépendance conceptuelle, 48
 DÍ, 19
 dominance immédiate, 19

- E -

extraposition à gauche, 20

- F -

frame, 45

- G -

gapping grammar, 31
 GE, 26
 génération, 21
 génération automatique, 21
 génération automatique de textes, 21

génération de texte, 21
 GG, 31
 GLF, 32
 grammaire à clauses définies, 23
 grammaire casuelle, 25
 grammaire Contextuelle discontinue, 28
 grammaire DCG, 23
 grammaire de cases, 22
 grammaire de clauses définies, 23
 grammaire de gap, 31
 grammaire de métamorphose, 24
 grammaire des cas, 25
 grammaire de slots, 22
 grammaire de structure de phrase généralisée, 34
 grammaire d'extraposition, 26
 grammaire discontinue, 27
 grammaire discontinue contextuelle, 28
 grammaire discontinue statique, 29
 grammaire d'unification, 30
 grammaire lacunaire, 31
 grammaire lexicale fonctionnelle, 32
 grammaire sémantique, 33
 grammaire syntagmatique généralisée, 34
 grammaire XG, 26
 GS, 31
 GSG, 34

– I –

interface en langue naturelle, 35

– L –

logique des prédicats, 13

– M –

MG, 26

– P –

PL, 37
 précédence linéaire, 37
 prédicat, 36
 préséance linéaire, 37
 primitive sémantique, 38
 principe de compositionnalité, 39
 prototype, 45
 P.S., 38

- R -

registre, 40
réseau de transition, 41
réseau de transition augmenté, 42
réseau de transition enrichi, 42
réseau de transition récursif, 43
réseau sémantique, 44
R.S., 44
RTN, 43

- S -

scénario, 46
schéma, 45
script, 46
slot, 14
système - q, 47
système - Q, 47

- T -

TDC, 48
théorie de la dépendance conceptuelle, 48
TLN, 49
traitement automatique des langues naturelles, 49
traitement automatique du langage naturel, 49
traitement des langues naturelles, 49
traitement du langage naturel, 49

- U -

unification, 50

- X -

XG, 26

5. INDEX ANGLAIS

– A –

ATN, 42
augmented transition network, 42

– B –

bottom-up analysis, 1
bottom-up parsing, 1

– C –

case grammar, 26
CD, 48
CDG, 22
CDT, 48
complementation, 15
compositionality principle, 39
conceptual dependancy, 48
conceptual dependancy theory, 48
conceptualization, 16
contextual discontinuous grammar, 22

– D –

DCG, 24
definite clause grammar, 24
deterministic analysis, 3
deterministic parsing, 3
discontinuous grammar, 28

– E –

extraposition grammar, 27

– F –

frame, 45
Frege's principle, 39

– G –

gapping grammar, 30
generalized phrase structure grammar, 34
GG, 30
GPSG, 34

– I –

ID, 19
 immediate dominance, 19

– L –

left extraposition, 20
 lexical analysis, 6
 lexical functional grammar, 31
 LFG, 31
 linear precedence, 37
 LP, 37

– M –

metamorphosis grammar, 25
 MG, 26
 morphological parsing, 7

– N –

natural language interface, 35
 natural language processing, 49
 NLI, 35
 NLP, 49
 nondeterministic analysis, 8
 nondeterministic parsing, 8

– P –

parallel analysis, 4
 parallel parsing, 4
 pragmatic analysis, 9
 pragmatic parsing, 9
 predicate, 36
 predicate calculus, 13

– Q –

Q-system, 47

– R –

recursive transition network, 43
 register, 40
 RTN, 43

- S -

script, 46
SDG, 33
semantic analysis, 10
semantic constraint, 17
semantic grammar, 32
semantic net, 44
semantic network, 44
semantic parsing, 10
semantic primitive, 38
serial analysis, 5
serial parsing, 5
slot, 14
slot grammar, 23
static discontinuous grammar, 33
syntactic analysis, 11
syntactic analyzer, 12
syntactic constraint, 18
syntactic parser, 12
syntactic parsing, 11

- T -

text generation, 21
top-down analysis, 2
top-down parsing, 2
transition network, 41

- U -

unification, 50
unification grammar, 29

- W -

wait-and-see parsing, 3

- X -

XG, 26

CHAPITRE 3

MODES DE FORMATION NÉOLOGIQUE EN TRAITEMENT AUTOMATIQUE DES LANGUES NATURELLES

Introduction

Le vocabulaire du traitement automatique des langues naturelles, à l'instar de celui de toute langue de spécialité, est composé de néologismes lexicaux ou termes nouveaux, en majorité des syntagmes, qui combinent plusieurs unités lexicales selon les règles morphosyntaxiques de la langue afin de dénommer les concepts nouvellement introduits en français. Le syntagme comme mode de formation, sous des allures bien françaises, a subi bien souvent l'influence de l'anglais puisque le concept est déjà dénommé dans cette langue; il ne suffit que de traduire les mots et à emprunter la forme en respectant la structure syntaxique du français. Mieux connue sous le nom de calque, cette pratique est en grande partie responsable, entre autres, de la présence massive de syntagmes dans les langues de spécialité.

La partie A de ce chapitre, consacrée à l'étude du syntagme, brosse un tableau du syntagme comme mode de formation néologique privilégié dans les langues de spécialité, puis des principaux types de syntagmes rencontrés dans le corpus et enfin, des particularités de ce corpus par rapport aux modèles de base proposés par Guilbert (Guilbert 1970).

La partie B traite du calque, de son incidence sur le vocabulaire du traitement automatique des langues naturelles, des principaux types rencontrés et des particularités propres au présent corpus.

A. Syntagme

«Les lexies complexes sont donc une nouveauté formelle de la langue actuelle, répondant à des besoins nouveaux de dénomination, pour les réalités d'une civilisation dans l'orbite anglo-américain (Meunier-Crespo 1987 : 275).

En langue de spécialité, «la référence à un objet ou à une notion peut être obtenue par l'intermédiaire d'une lexie simple ou d'une lexie complexe dont la tendance est l'univocité de la signification» (Boulanger 1979 : 17). La lexie complexe ou syntagme lexicalisé, pour reprendre l'appellation de Rey-Debove (1971), est la juxtaposition de plusieurs lexies simples qui désignent un concept unique et permanent à l'intérieur d'un champ sémantique bien spécifique et qui ne peuvent être déterminées de façon individuelle. Le syntagme lexicalisé se comporte dans la phrase de la même façon que la lexie simple. On dit qu'il est lexicalisé bien que, non encore intégré au dictionnaire, il jouisse d'«une autonomie lexicale suffisamment reconnue par une communauté linguistique pour fonctionner en tant qu'unité monoréférentielle indépendante dans la communication» (ibid). Il se distingue ainsi du syntagme libre (ou accidentel) qui, lui, est la réunion fortuite d'unités lexicales à des fins exclusivement discursives.

Fortement inspiré de l'anglais, le vocabulaire spécialisé des langues anglo-germaniques faisant un usage abondant de la combinatoire syntagmatique, le syntagme lexicalisé en langue de spécialité est principalement introduit par le recours massif au calque comme mode de formation; «on emprunte à la langue étrangère le syntagme, mais on traduit littéralement les éléments» (Vinay et Darbelnet cités dans Santoyo 1987 : 241). Cependant, malgré l'influence de l'anglais, le syntagme terminologique bien que différent du syntagme saussurien a ceci de commun avec lui: il répond aux trois caractères de binarité, linéarité et

de solidarité soulevés par Bonnard (1972 : 5899). On dit du syntagme qu'il est linéaire lorsque son expansion s'effectue de la gauche vers la droite, qu'il est binaire lorsqu'il obéit à la structure déterminé --> déterminant et qu'il est solidaire parce qu'«il s'agit toujours d'unités plus vastes, composées elles-mêmes d'unités plus restreintes, les unes et les autres étant dans un rapport de solidarité réciproque» (Saussure cité dans Bonnard 1972 : 5901).

Une étude statistique menée par Roger Goffin (1979 ; 171-182) sur le vocabulaire technique allemand révèle que 85 pour cent du vocabulaire en formation est composé de syntagmes simples ou complexes variant de deux à six éléments. Transposé à la langue générale, ce pourcentage passe de 85 à 12, 5 p. cent d'après les résultats obtenus par Boulanger (1989 : 521).

En français, devant les besoins incessants de nouvelles dénominations des vocabulaires spécialisés, le recours à des lexies simples agissant comme bases et additionnées de certaines déterminations apparaît comme un procédé des plus efficaces pour désigner des concepts nouveaux. L'expansion du syntagme est effectuée à raison d'un déterminant à la fois allant de l'implicite vers l'explicite, le terme qui suit apportant une délimitation du contenu du terme qui précède. C'est d'ailleurs ce même déterminant qui permet, dans le cas de séries formées sur une base commune, d'illustrer certaines oppositions. Par exemple, la base *analyse* donne lieu aux syntagmes

analyse ascendante

analyse descendante

analyse déterministe

analyse non déterministe

où *analyse ascendante* s'oppose à *analyse descendante* et *analyse déterministe* s'oppose à *analyse non déterministe*.

1. Découpage du syntagme lexicalisé

En traitement automatique des langues naturelles comme dans toute langue de spécialité la combinatoire syntagmatique est un procédé amplement utilisé. «Les unités lexicales complexes ne sont en définitive que des segments d'énoncés extraits de leur contexte et confirmés dans leur cohésion par la répétition dans le discours et par l'expression d'une unité conceptuelle» (Guilbert 1965 : 272). Il appartient au terminologue de déterminer si les syntagmes repérés sont le fruit d'une association fortuite ou s'ils servent à dénommer un concept unique et permanent. Pour ce faire, il a recours à différents critères qui, bien que n'étant pas exclusifs, constituent des moyens valables de déterminer la nature d'un syntagme et s'il doit faire l'objet d'une étude plus poussée. Nous pensons ici aux critères proposés par Goffin (1979 : 159-168) plus particulièrement aux critères quantitatifs et taxinomiques.

Dans le cas qui nous intéresse, nous avons dépouillé la documentation française disponible sur notre sujet pour repérer des syntagmes lexicalisés. Il est monnaie courante d'y rencontrer des syntagmes français suivis entre parenthèses de leur équivalent et parfois même de leur sigle anglais. La présence d'un syntagme et de son équivalent anglais a donc servi de critère pour déterminer ce que Guilbert appelle «la référence au caractère unique et constant du signifié» (Guilbert 1970 : 121). Par exemple, le fait que le syntagme *réseau de transition augmenté* soit suivi la plupart du temps de son équivalent anglais *augmented transition network* et de *ATN* dans la documentation dépouillée a donné un indice sur le degré de lexicalisation de ce syntagme. Mais lorsque l'équivalent anglais n'était pas aussi disponible, nous avons consulté les renvois bibliographiques afin de vérifier s'il était possible de retracer un quelconque document de présentation du concept rédigé en anglais par le concepteur lui-même. C'est donc en établissant «l'équivalent de langue à langue» (Vinay 1979 : 96) que nous avons pu, dans de nombreux cas, déceler la présence d'un syntagme lexicalisé.

Le critère taxinomique qui consiste à replacer un syntagme dans son domaine d'énonciation a également été fort utile pour déterminer la nature d'un syntagme et il a servi à établir la différence avec d'autres référents voisins (par exemple, *réseau de transition* est à la source de la structuration générique/spécifique: *réseau de transition augmenté* et *réseau de transition récursif*).

En règle générale, un syntagme est consacré par l'usage. Le critère quantitatif, soit la fréquence d'apparition d'un syntagme dans le discours, bien que ne constituant pas un critère de rejet, est venu confirmer la présence d'un syntagme et son degré de lexicalisation. Toutefois, puisqu'il n'est utilisé que dans des situations de communication très restreintes, il n'est pas toujours possible de vérifier le degré de lexicalisation d'un syntagme de cette façon.

Le contexte d'utilisation a également contribué à déterminer si nous étions en présence d'un syntagme. Car, en effet, «il peut fournir des indices sur la longueur du syntagme; il peut aider à circonscrire le syntagme par opposition à la lexie simple ou par opposition à d'autres lexies apparaissant dans le même entourage contextuel» (Boulanger 1979 : 181).

Par ailleurs, la prise en considération de certains signes typographiques tels la ponctuation, les guillemets, le soulignement, les caractères gras, l'italique, etc., a permis dans certains cas de déceler la présence d'un syntagme puisque les concepts nouveaux sont souvent introduits de cette façon.

En dernier lieu, une fois toutes ces étapes franchies, nous avons eu recours à un spécialiste pour déterminer si oui ou non le groupe de mots analysé correspondait bien à un référent réel et unique.

C'est donc l'ensemble de ces critères qui a permis de délimiter les syntagmes de notre nomenclature.

2. Pourcentage des résultats

Notre corpus qui compte au total 92 termes est composé de 59 syntagmes (64 %) et de 33 lexies simples (36 %) comme le démontre le tableau suivant:

Tableau 1: Lexies simples et syntagmes

	nombre	pourcentage
syntagmes	59	64 %
lexies	33	36 %
	92	100 %

Les résultats de l'étude ne viennent pas corroborer les résultats de Goffin mais tendent clairement à démontrer, en dépit d'un pourcentage inférieur (85 p. cent chez Goffin contre 64 p. cent), la prédominance du syntagme comme mode de formation du vocabulaire. Nous imputons ces résultats à deux facteurs. Tout d'abord, les cinquante notions couvertes ne sont pas exhaustives; elle ne représentent qu'une partie de la terminologie relative au sous-domaine du traitement automatique des langues naturelles. Ensuite, certaines de ces notions ont été choisies en rapport avec d'autres notions de contenu sémantique plus vaste (*unification* dans *grammaire d'unification*, *prédicat* dans *calcul des prédicats*). Pour ces raisons, il apparaît donc hasardeux de généraliser ces résultats à l'ensemble du sous-domaine.

Parmi les 59 syntagmes dénombrés, on compte 42 syntagmes simples (71 %) et 17 syntagmes complexes (29 %).

Tableau 2: Syntagmes simples et syntagmes complexes

syntagmes	nombre	pourcentage
simples	42	71 %
complexes	17	29 %
	59	100 %

Le syntagme simple est constitué d'une base formée d'une lexie simple et d'une expansion simple à une détermination tandis que le syntagme complexe est formé d'une base composée d'un syntagme lexicalisé et d'une expansion multiple à une ou plusieurs déterminations. Dans notre corpus, on ne rencontre que des cas d'expansion multiple à une détermination. Voici comment se fait le découpage d'un syntagme complexe:

grammaire lexicale fonctionnelle

|-----| |-----|

dé --> dā + dā

|-----|

nom + adj.

|----|

dé + dā

Le déterminé *grammaire lexicale* à lui seul est un syntagme lexicalisé auquel on a adjoind le déterminant adjectival *fonctionnelle*. Il en est de même pour les syntagmes

complexes *grammaire syntagmatique généralisée, réseau de transition augmenté* et *réseau de transition récursif*.

Les 59 syntagmes répertoriés sont formés de deux à six constituants. Au total, 25 syntagmes (27 %) sont formés de deux constituants, 18 (20 %) de trois constituants, 10 (11 %) de quatre constituants, trois (3 %) de cinq constituants et finalement trois (3 %) sont formés de six constituants. Le tableau 3 offre une vue d'ensemble des résultats obtenus.

Tableau 3: Structure syntaxique des syntagmes lexicalisés en traitement automatique des langues naturelles

constituants	nombre	pourcentage
deux	25	27 %
trois	18	20 %
quatre	10	11 %
cinq	3	3 %
six	3	3 %
	59	64 %

Deux constituants

analyse ascendante
analyse descendante
analyse déterministe
analyse lexicale
analyse morphologique
analyse pragmatique
analyse sémantique
analyse syntaxique
analyseur syntaxique
contrainte sémantique
contrainte syntaxique
dépendance conceptuelle
dominance immédiate
gapping grammar
génération automatique
grammaire casuelle
grammaire DCG
grammaire discontinue
grammaire lacunaire
grammaire sémantique
grammaire XG
précédence linéaire
préséance linéaire
primitive sémantique
réseau sémantique

Quatre constituants

calcul des (de les) prédicats
génération automatique de textes
grammaire à clauses définies
grammaire de clauses définie
grammaire des (de les) cas
interface en langue naturelle
logique des (de les) prédicats
réseau de transition augmenté
réseau de transition enrichi
réseau de transition récursif

Trois constituants

analyse en parallèle
analyse en série
analyse non déterministe
extraposition à gauche
génération de textes
grammaire Contextuelle Discontinue
grammaire de cases
grammaire de gap
grammaire de métamorphose
grammaire de slot
grammaire d'extraposition
grammaire discontinue contextuelle
grammaire discontinue statique
grammaire d'unification
grammaire lexicale fonctionnelle
grammaire syntagmatique généralisée
principe de compositionnalité
réseau de transition

Cinq constituants

théorie de la dépendance conceptuelle
traitement des (de les) langues naturelles
traitement du (de le) langage naturel

Six constituants

grammaire de structure de phrase généralisée
traitement automatique des (de les) langues naturelles
traitement automatique du (de le) langage naturel

À la lumière de ces résultats, les syntagmes lexicalisés formés de deux constituants obéissant au modèle **nom + adjectif** sont les plus productifs suivis en ordre de fréquence des syntagmes formés de trois, de quatre, de cinq et de six constituants. Les résultats obtenus par Goffin (1979) faisaient état d'un plus grand nombre de syntagmes à trois constituants répondant au modèle **nom + joncteur + nom** que de syntagmes à deux

constituants. À eux seuls, les modèles **nom + adjectif** et **nom + joncteur + nom** totalisent 57 p. cent des syntagmes tandis que chez Goffin ils atteignaient 80 p. cent. Par ailleurs, le présent corpus comporte 16 syntagmes à quatre constituants et plus, ce qui signifie un pourcentage de 19 p. cent contre 5 p. cent dans l'étude de Goffin.

2.1 Typologie des syntagmes

Le syntagme dans le corpus obéit à la structure fondamentale déterminé --> déterminant proposée par Guilbert (1970 : 117) à l'exception des modèles empruntés intégralement à l'anglais. La structure syntaxique du syntagme et son expansion dépendent de la nature du déterminé et du rôle fonctionnel du déterminant.

Voici les six types de syntagmes rencontrés dans le corpus en ordre de fréquence. Les modèles de syntagmes complexes ont été ramenés à la structure binaire.

Tableau 4 : Typologie des syntagmes: modèles binaires

modèle	nombre	pourcentage
nom + adjectif	32	54 %
nom + joncteur + nom	15	25 %
nom + joncteur + prédéterminant + nom	8	14 %
nom + joncteur ø + nom	2	3 %
nom + adverbe + adjectif	1	2 %
adjectif + nom	1	2 %
	59	100 %

2.1.1 nom + adjectif

analyse ascendante
 analyse descendante
 analyse déterministe
 analyse lexicale
 analyse morphologique
 analyse pragmatique
 analyse sémantique
 analyse syntaxique
 analyseur syntaxique
 contrainte sémantique
 contrainte syntaxique
 dépendance conceptuelle
 dominance immédiate
 génération automatique
 grammaire casuelle
 grammaire Contextuelle Discontinue
 grammaire de structure de phrase généralisée
 grammaire discontinue
 grammaire discontinue contextuelle
 grammaire discontinue statique
 grammaire lacunaire
 grammaire lexicale fonctionnelle
 grammaire sémantique
 grammaire syntagmatique généralisée
 précédence linéaire
 préséance linéaire
 primitive sémantique
 réseau de transition augmenté
 réseau de transition enrichi
 réseau de transition récursif
 réseau sémantique

2.1.2 nom + joncteur + nom

analyse en parallèle
 analyse en série
 extraposition à gauche
 génération automatique de textes
 génération de textes
 grammaire à clauses définies
 grammaire de cases
 grammaire de clauses définies
 grammaire de gap
 grammaire de métamorphose
 grammaire d'extraposition
 grammaire de slot
 grammaire d'unification
 interface en langue naturelle
 principe de compositionnalité
 réseau de transition

2.1.3 nom + joncteur + prédéterminant + nom

calcul des (de les) prédicats
 grammaire des (de les) cas
 logique des (de les) prédicats
 traitement automatique des (de les) langues naturelles
 traitement automatique du (de le) langage naturel
 traitement du (de le) langage naturel
 théorie de la dépendance conceptuelle

2.1.4 nom + joncteur \emptyset + nom

grammaire DCG
 grammaire XG

2.1.5 nom + adverbe + adjectif

analyse non déterministe

2.1.6 adjectif + nom

gapping grammar

3. Analyse

Comme nous l'avons vu précédemment, la tendance qui se dégage du corpus est une nette prédilection pour les syntagmes adjectivaux et les syntagmes prépositionnels. Le modèle **nom + adjectif** est de loin le plus fréquent avec ses 31 occurrences. «L'adjectif apparaît, en effet, comme l'instrument le plus apte à fournir les réalisations particulières et différenciées d'unités lexicales préexistantes» (Guilbert 1965 : 270). Outre des syntagmes simples, cette combinaison produit aussi des syntagmes complexes. Que le premier élément soit une lexie simple (*grammaire*) ou une lexie complexe (*grammaire discontinue*), il occupe la fonction de déterminé (ou base). Le second élément, l'adjectif dans le cas qui nous intéresse, joue le rôle de déterminant.

Certaines bases comme *analyse* et *grammaire*, dotées d'une vocation de génération à cause de l'étendue de la notion qu'elles couvrent, donnent lieu à des séries d'unités terminologiques répondant à ce modèle. L'adjectif, en tant que second élément du syntagme, permet non seulement de constituer des unités de signification nouvelles dans un champ sémantique nouveau mais permet aussi de distinguer, et parfois même d'opposer, les unités lexicales issues d'une même base par leur trait le plus significatif. Dans le cas des syntagmes complexes, il permet de restreindre le contenu sémantique de la base comme c'est le cas pour *grammaire discontinue* qui devient *grammaire discontinue contextuelle* et *grammaire discontinue statique* par l'adjonction d'une seconde détermination qui vient ajouter un élément distinctif aux syntagmes ainsi formés. Dans les syntagmes complexes *grammaire discontinue contextuelle* et *grammaire discontinue statique*, le syntagme lexicalisé *grammaire discontinue* est secondairement déterminé par les adjectifs *contextuelle* et *statique*, car en vertu du principe de linéarité de la phrase, c'est le terme directement à la droite de la base qui est le déterminant principal; «dans l'ordre de déroulement du discours, le terme qui se trouve graphiquement à droite du terme de base correspond à une détermination plus essentielle du

point de vue du contenu sémantique de l'ensemble signifié» (Guilbert 1965 : 272-273). La base (*grammaire discontinue*), elle, est porteuse de la substance essentielle de ce syntagme. C'est d'ailleurs elle qui revêt la fonction grammaticale de substantif régissant ainsi l'accord en genre et en nombre du déterminant et du pré-déterminant (s'il y a lieu). Même si nous sommes en présence de syntagmes complexes, la structure binaire déterminé --> déterminant est sauvegardée.

Il arrive que l'adjectif ait une fonction autre que celle qu'on lui connaît. C'est le cas de l'adjectif de relation qui résulte de la transposition syntaxique d'un substantif en adjectif: *casuelle* --> *cas*. Il y a donc possibilité d'alternance entre le modèle **nom + adjectif** et le modèle **nom + joncteur + nom** qu'on rencontre dans *grammaire casuelle* et *grammaire des cas*. D'un point de vue sémantique, les adjectifs de relation se comportent de la même façon par rapport à la base que le déterminant introduit par la préposition *de* en ce sens que le déterminé et le déterminant sont considérés comme deux notions individuelles.

Le modèle **nom + joncteur + nom** est le second en importance avec ses seize occurrences. Au lieu d'être établi par juxtaposition comme avec l'adjectif, le rapport syntagmatique est obtenu par liaison par le biais d'une préposition. Les expansions de la base, qu'elles soient introduites par la préposition *de*, *à* ou *en*, jouent le même rôle que l'adjectif mais leur degré de cohésion varie en fonction de la préposition utilisée. La préposition *de* assure le plus fort degré de cohésion et de lexicalisation comparable à celui de l'adjectif parce que, en liant la base à son déterminant nominal, elle confère au nom la même fonction que l'adjectif. C'est d'ailleurs celle qu'on rencontre le plus souvent avec onze occurrences, suivie de *en* avec trois occurrences et de *à* avec deux occurrences.

La préposition *de* renferme une idée de moyen comme dans *grammaire de métamorphose*, *grammaire d'unification*, *réseau de transition* et de désignation

comme dans *grammaire de clauses définies, principe de compositionnalité*. La préposition *en*, de son côté, comporte l'idée de manière comme dans *analyse en parallèle, analyse en série* et l'idée de matière comme dans *interface en langue naturelle*. Enfin la préposition *à* exprime tantôt une idée de direction *extraposition à gauche*, tantôt la matière constituante *grammaire à clauses définies*.

Les syntagmes complexes *grammaire de clauses définies* et *interface en langue naturelle*, basés sur le modèle *nom + joncteur + nom*, et *théorie de la dépendance conceptuelle, traitement (automatique) des langues naturelles* et *traitement automatique du langage naturel* basés sur le modèle *nom + joncteur + prédéterminant + nom* ont pour déterminants les syntagmes lexicalisés *dépendance conceptuelle, langue (s) naturelle (s), langage naturel*, et *clauses définies* qui jouent le rôle de déterminants uniques par rapport à la base de la même façon qu'un syntagme à une détermination. Nous sommes ici en présence d'une expansion à droite d'une base simple à l'aide d'un syntagme lexicalisé.

Dans *théorie de la dépendance conceptuelle*, la présence de l'article défini confère au déterminant une valeur spécifique qui peut réduire les chances de lexicalisation du syntagme car la base et le déterminant conservent leur pleine valeur alors qu'en son absence la fonction du déterminant se rapproche de celle de l'adjectif. Le syntagme *théorie de la dépendance conceptuelle* deviendra tout simplement *dépendance conceptuelle*. Comme nous l'avons vu précédemment, les chances de lexicalisation de ce syntagme étaient faibles à cause de la présence de la préposition et de l'article défini.

L'emploi du pluriel comme dans *calcul des prédicats, grammaire des cas* ou *traitement (automatique) des langues naturelles* augmente les chances de lexicalisation du syntagme à cause de sa valeur de généralisation.

Peu à peu, et ce malgré qu'ils répondent à la structure binaire déterminé --> déterminant, certains syntagmes subissent une réduction importante passant d'un syntagme complexe à une lexie simple comme dans *génération automatique de textes* qui successivement deviendra *génération automatique* puis *génération*. La progression de ce syntagme s'effectue sur une base binaire de même que sa régression comme le démontre la décomposition suivante:

génération

1 |-----> génération automatique

2 |-----> génération automatique de textes

Notre corpus comprend deux syntagmes *grammaire DCG* et *grammaire XG* qui répondent au modèle **nom + joncteur \emptyset + nom** où le sigle pris isolément joue le rôle de substantif. Bien que ce modèle soit tout à fait acceptable en français, il n'en demeure pas moins que le sigle, dans ce cas-ci emprunté à l'anglais, ne constitue pas une détermination courante en français. Nous sommes ici en présence d'une utilisation abusive du sigle puisque sa raison d'être est la réduction et non la formation d'un syntagme. Par ailleurs les sigles *DCG* (*définite clause grammar*) et *XG* (*extraposition grammar*) renferment déjà la notion de grammaire et de dire *grammaire DCG* ou *grammaire XG* est une forme de redondance car *grammaire* y est répétée deux fois. C'est donc une formation à rejeter en français.

La base *primitive* dans le syntagme *primitive sémantique* est employée comme substantif alors qu'à l'origine c'est un adjectif féminin. Il s'est produit ici une substantivation de l'adjectif *primitive* sous l'influence de l'anglais *primitive*. Non seulement avons-nous assisté à un changement de catégorie lexicale mais à un changement

de genre car *semantic primitive* aurait dû se lire «primitif sémantique». Il y a également eu déstructuration syntaxique abusive.

Dans le syntagme *gapping grammar*, emprunté intégralement à l'anglais, la structure étrangère déterminé --> déterminant est conservée. Bien qu'il soit classifié dans la catégorie **adjectif + nom**, il ne correspond pas tout à fait à ce modèle puisque *gapping* relève davantage du gérondif que de l'adjectif.

3. Conclusion

À la lumière des résultats obtenus dans la présente partie, il semble que l'anglais ait une incidence marquée sur le groupement syntagmatique comme mode de formation du vocabulaire du traitement automatique des langues naturelles. Car, en effet, sous le syntagme se cache l'absence d'une appellation pour désigner un concept déjà dénommé en anglais. C'est devant ce besoin de dénommer qu'on a recours au syntagme souvent introduit par la voie du calque dont il sera question dans la prochaine partie. On peut toutefois affirmer que la conversion du syntagme de l'anglais au français s'effectue conformément à l'ordre déterminé - déterminant du français ce qui en fait un mode de formation efficace.

En dépit du manque d'uniformité de la formation des vocabulaires spécialisés, on peut cependant remarquer que certaines bases ont la propriété d'engendrer des séries comme c'est le cas pour les termes *analyse* et *grammaire* empruntés à la linguistique. Autre fait à noter, la longueur des syntagmes ne semble pas non plus être limitée.

Il faut également préciser que, même si le corpus comporte des syntagmes formés de deux à six constituants, ce sont les syntagmes à deux et à trois constituants qui sont les plus fréquents. Ces derniers sont souvent abrégés par la voie du sigle tandis que les premiers

demeurent tels quels en grande partie à cause de leur plus grande cohésion attribuable à l'absence de préposition.

Les syntagmes formés de quatre constituants et plus, pour leur part, font souvent l'objet d'une réduction importante allant d'une à plusieurs unités principalement à cause du manque de cohésion qu'occasionne la présence de prépositions et de déterminants et de la valeur plus descriptive qu'ils comportent. Il est même arrivé d'observer la réduction d'un syntagme complexe en une unité simple.

Mais avant tout, il ne faut pas oublier que le traitement automatique des langues naturelles est une discipline qui rejoint un très grand nombre d'utilisateurs et que c'est souvent eux qui consacrent l'usage d'un terme plutôt que d'un autre.

B. Calque

Phénomène relativement récent de même que sa dénomination explicite, le calque constitue en ce moment «l'un des principaux moteurs de l'actuel changement lexical» (Santoyo 1987 : 244) dont sont témoins les vocabulaires scientifiques et techniques. Le traitement automatique des langues naturelles, dans la même lignée que l'informatique et l'intelligence artificielle dont il constitue un sous-domaine d'application, n'échappe pas à la règle. C'est un phénomène observable en anglais mais qui, selon Santoyo, se répercute sur toutes les langues occidentales «qui sont en passe de devenir des langues traduites» (1987 : 243).

L'uniformisation des modes de vie, eux-mêmes calqués sur l'«American way of life», l'internationalisation des sciences et des techniques, la rapidité des communications sont autant de facteurs d'ordre plus sociologique que linguistique qui contribuent à l'anglicisation progressive du français. Pour répondre aux besoins pressants de communication, les «calques (comme sida) facilitent la communication internationale dans le domaine scientifique et technique dans la mesure où des structures lexicales peuvent être aisément reconnues par les spécialistes» (Chansou cité dans Santoyo 1987 : 242-243).

C'est pour rester le plus près du modèle de départ que le transfert d'une langue à l'autre s'effectue de façon partielle; on copie la structure du modèle anglais tout en utilisant les mots du vocabulaire de la langue d'arrivée. C'est d'ailleurs pourquoi on situe le calque à mi-chemin «entre les moyens externes (emprunts) et les moyens internes (la formation des mots nouveaux par dérivation, composition, etc.)» (Goldis 1976 : 100). Il est devenu la valeur sûre des modes de formation; sous des allures bien françaises, il permet de garder un point de référence avec l'anglais si bien qu'il est très difficile à repérer à l'oeil nu sans tenir compte de l'équivalent anglais. Malheureusement les équivalents français s'avèrent trop souvent maladroits et se rapprochent davantage de la description et de la définition perdant ainsi le caractère synthétique du terme anglais qui est dans bien des cas plus accrocheur.

Par souci de concision et d'efficacité, les langues occidentales ont plus ou moins le choix de se raccrocher au calque si elles veulent survivre à l'invasion de la terminologie américaine.

L'anglais, pour ne pas dire l'américain, est une langue très productive à bien des niveaux (diplomatique, économique, politique, scientifique, technique, etc.). En somme, il a pris la place qu'occupait auparavant le français sur le plan international. Ce dernier, au fil des ans, a vu son statut passer de celui de langue «prêteuse» à celui de langue «emprunteuse» et doit maintenant faire face au raz-de-marée de l'américain sous peine d'y être submergé.

La suprématie d'une langue sur les autres était prévisible et c'est pour la contrer qu'on avait assisté à la création de l'espéranto qui se voulait la langue des communications internationales. L'échec de cette tentative est en partie causé par l'absence d'âme de cette langue sans culture. L'anglais joue aujourd'hui le rôle que devait jouer l'espéranto et il est en train de révolutionner le mode de formation du vocabulaire des langues occidentales qui en sont réduites à adopter une stratégie de «récupération linguistique» par la voie du calque. Toutefois, il demeure de loin préférable à l'emprunt intégral surtout s'il respecte la structure morphologique et syntaxique de la langue emprunteuse car il est dorénavant appelé à contribuer à l'enrichissement de la langue au même titre que la création et la dérivation. «Le calque est appelé à donner une nouvelle impulsion aux études lexicographiques; dans 50 ans on ne concevra plus de dictionnaire étymologique qui ne tiendra pas compte 1) de la filiation sémantique des sens, 2) de l'étude systématique des mots introduits par la traduction» (Bally cité dans Meunier-Crespo 1987 : 277).

Chronologiquement et ce, pour pallier l'absence momentanée d'une unité conceptuelle dans la langue d'arrivée, «l'emprunt précède souvent le calque, la nécessité de dénommer pressant davantage que la mise au point d'un vocable de la langue» (Boulanger 1987 : 327). L'emprunt intégral se définit comme une unité lexicale étrangère empruntée

textuellement par la langue emprunteuse sans avoir subi de substitution de forme au préalable. Ce non-interventionnisme est en partie imputable au parallélisme apparent des moyens d'expression des langues surtout si ces langues ont pour dénominateur commun d'utiliser la composition et la dérivation comme mode de formation. C'est principalement là que réside la différence entre l'emprunt et le calque: «créé dans le processus de traduction, le calque est le résultat du contact entre les deux langues» (Goldis 1976 : 101). On assiste ici à un dédoublement du calque qui est à la fois un procédé et un résultat. Le calque procédé est «la substitution d'un signifiant de texture étrangère par un signifiant ancien ou nouveau pour désigner soit un signifié totalement étranger, soit un signifié déjà opérationnel en français, soit un signifié lui-même nouveau en français. Quant au calque produit c'est l'unité lexicale issue de la substitution» (Boulanger 1987 : 328).

1. Pourcentage des résultats

Sur un total de 92 termes français, notre corpus compte 60 calques complets ou partiels et dix emprunts intégraux. De ce nombre, on compte quatre propositions de remplacement: *grammaire discontinue contextuelle*, *grammaire de cases*, *grammaire discontinue statique* et *grammaire lacunaire*. On appelle proposition de remplacement tout terme proposé par le terminologue qui vise à remplacer un terme qui lui apparaît plus ou moins satisfaisant du point de vue de la langue. Les cinquante-six termes restants, puisés à même la documentation dépouillée, sont des calques issus des milieux spécialisés.

À des fins de classification, nous avons retenu cinq types de calques comme en rend compte le tableau suivant:

Tableau 5: Typologie du calque

type	nombre	pourcentage
1. traductions littérales	27	45 %
2. adaptations syntaxiques	15	25 %
3. adaptations morphosyntaxiques	11	18, 4 %
4. adaptations morphologiques	2	3, 3 %
5. adaptations partielles	5	8, 3 %
	60	100%

1.1 Traductions littérales

<i>bottom-up parsing</i>	<i>analyse ascendante</i>
<i>CDT</i>	<i>TDC</i>
<i>complementation</i>	<i>complémentation</i>
<i>conceptual dependancy</i>	<i>dépendance conceptuelle</i>
<i>conceptualisation</i>	<i>conceptualisation</i>
<i>contextual discontinuous grammar</i>	<i>grammaire Contextuelle Discontinue</i> <i>grammaire discontinue contextuelle</i>
<i>deterministic analysis</i>	<i>analyse déterministe</i>
<i>discontinuous grammar</i>	<i>grammaire discontinue</i>
<i>generalized phrase structure grammar</i>	<i>grammaire syntagmatique généralisée</i>
<i>ID</i>	<i>DI</i>
<i>immédiate dominance</i>	<i>dominance immédiate</i>
<i>lexical functional grammar</i>	<i>grammaire lexicale fonctionnelle</i>
<i>LFG</i>	<i>GLF</i>
<i>linear precedence</i>	<i>préséance linéaire</i> <i>précédence linéaire</i>
<i>LP</i>	<i>PL</i>
<i>predicate</i>	<i>prédicat</i>
<i>register</i>	<i>registre</i>
<i>semantic grammar</i>	<i>grammaire sémantique</i>
<i>semantic network</i>	<i>réseau sémantique</i>
<i>script</i>	<i>scénario</i>
<i>static discontinous grammar</i>	<i>grammaire discontinue statique</i>
<i>top-down parsing</i>	<i>analyse descendante</i>
<i>unification</i>	<i>unification</i>
<i>XG</i>	<i>GE</i>

1.2 Adaptations syntaxiques

<i>augmented transition network</i>	<i>réseau de transition augmenté</i> <i>réseau de transition enrichi</i>
<i>compositionality principle</i>	<i>principe de compositionnalité</i>
<i>conceptual dependancy theory</i>	<i>théorie de la dépendance conceptuelle</i>
<i>extraposition grammar</i>	<i>grammaire d'extraposition</i>
<i>generalized phrase structure grammar</i>	<i>grammaire de structure de phrase généralisée</i>
<i>left extraposition</i>	<i>extraposition à gauche</i>
<i>metarmorphosis grammar</i>	<i>grammaire de métamorphose</i>
<i>natural language processing</i>	<i>traitement (automatique) du langage naturel</i>
<i>nondeterministic analysis</i>	<i>analyse non déterministe</i>
<i>parallel analysis</i>	<i>analyse en parallèle</i>
<i>recursive transition network</i>	<i>réseau de transition récursif</i>
<i>serial analysis</i>	<i>analyse en série</i>
<i>transition network</i>	<i>réseau de transition</i>
<i>unification grammar</i>	<i>grammaire d'unification</i>

1.3 Adaptations morphosyntaxiques

<i>case grammar</i>	<i>grammaire des cas</i>
<i>definite clause grammar</i>	<i>grammaire à clauses définies</i> <i>grammaire de clauses définies</i>
<i>logic predicate</i>	<i>logique des prédicats</i>
<i>natural language processing</i>	<i>traitement (automatique) des langues naturelles</i>
<i>natural language interface</i>	<i>interface en langue naturelle</i>
<i>predicate calculus</i>	<i>calcul des prédicats</i>
<i>slot grammar</i>	<i>grammaire de cases</i>
<i>text generation</i>	<i>génération (automatique) de textes</i>

1.4 Adaptations morphologiques

case grammar

grammaire casuelle

gapping grammar

grammaire lacunaire

1.5 Adaptations partielles

definite clause grammar (DCG)

grammaire DCG

extraposition grammar (XG)

grammaire XG

gapping grammar

grammaire de gap

semantic primitive

primitive sémantique

slot grammar

grammaire de slots

GS

2. Analyse

Le calque est constitué de lexies simples, de lexies complexes et apparaît même sous forme de sigles. En voici quelques exemples:

predicate

prédicat

predicate calculus

calcul des prédicats

XG

GE

Dans notre corpus, tous les calques rencontrés, à l'exception des emprunts intégraux (comme *gapping grammar*), répondent au modèle déterminé --> déterminant propre au français si bien qu'ils peuvent passer inaperçus tellement ils ressemblent à des compositions indigènes. Tel n'est pas le cas et c'est seulement par un examen approfondi des origines anglaises qu'on peut parvenir à les repérer contrairement aux emprunts intégraux. Des termes comme *prédicat*, *registre*, *unification* sont des exemples de lexies simples de facture bien française lorsqu'ils sont employés dans la langue générale mais qui, pris dans le sous-domaine du traitement automatique des langues naturelles,

relèvent davantage de la néologie de forme (calque) que de la néologie sémantique surtout s'ils ont pour équivalent anglais les termes *predicate*, *register*, *unification*.

Pour effectuer la classification des calques contenus dans notre corpus, nous avons utilisé la classification dont Jean-Claude Boulanger s'est servi pour interpréter les résultats d'une étude parallèle à celle-ci portant sur le sous-domaine de l'intelligence artificielle (Boulanger 1987 : 328-329).

Parmi les types de calques rencontrés, la traduction littérale ou mot à mot d'un syntagme ou d'un sigle d'une langue à l'autre est très productive de néologismes lexicaux de type:

<i>linear precedence</i>	<i>précédence linéaire</i> <i>préséance linéaire</i>
<i>immediate dominance</i>	<i>dominance immédiate</i>
<i>X G</i>	<i>GE</i>

Le terme *précédence* a été calqué de toutes pièces sur le modèle anglais car il n'existe pas en français; il constitue un néologisme formel en français. Le terme *préséance* est l'équivalent français proposé qui paraît plus acceptable.

L'adaptation syntaxique, elle aussi très prolifique et la seconde en importance dans notre corpus, est une traduction qui respecte, dans la séquence dé --> dā, les rapports logiques qui unissent ces éléments au moyen de prépositions. Par exemple, dans

<i>transition network</i>	<i>réseau de transition</i>
<i>serial analysis</i>	<i>analyse en série</i>

qui répondent au modèle **nom + prép. + nom**, il y a addition des rapports de jonction *de* et *en* nécessaires en français mais implicites en anglais.

Vient ensuite l'adaptation morphosyntaxique qui, pour sa part, tient compte des règles d'agencement lexical et des modèles français. Dans

natural language interface

interface en langue naturelle

predicate calculus

calcul des prédicats

formés sur le modèle **nom + prép. + prédéterminant + nom**, la transposition s'observe par la marque d'intégration morphologique du substantif (genre ou nombre). La notion de genre de *langue naturelle* et le rapport de jonction *en* entre le déterminé *interface* et le déterminant *langue naturelle* sont absents du modèle anglais. Il en est de même pour *calcul des prédicats* où la notion de nombre cette fois-ci est rendue dans l'équivalent français alors qu'elle est absente en anglais.

L'adaptation morphologique est la substitution de la forme étrangère par son équivalent indigène. Le «calqueur» modifie et adapte le terme de départ comme

case grammar

grammaire casuelle

gapping grammar

grammaire lacunaire

Le syntagme *case grammar* a également comme équivalent *grammaire des cas*. Dans l'exemple ci-dessus, le terme *cas* dérivé adjectivement produit l'adjectif de relation *casuel*, d'où *grammaire casuelle*.

Enfin, l'adaptation partielle conserve un élément étranger anglais dans sa composition comme c'est le cas pour les syntagmes

extraposition grammar

grammaire XG

gapping grammar

grammaire de gap

sémantic primitive

primitive sémantique

à l'intérieur desquels les termes anglais *gap* et *primitive* ainsi que le sigle anglais *XG* ont été empruntés intégralement à l'anglais.

Par ailleurs, dans certains cas, on assiste à une réduction du syntagme par degré où en général ce sont les déterminants qui disparaissent pour ne conserver que les éléments les plus significatifs du syntagme. Le terme anglais *text generation* a pour équivalent *génération automatique de textes* qui avec l'usage et par économie va devenir *génération de textes* puis par réduction *génération* tout court. C'est également le cas de *théorie de la dépendance conceptuelle* qui va devenir *dépendance conceptuelle* et de *traitement automatique des langues naturelles* qui va devenir *traitement des langues naturelles*.

L'emprunt du sigle étranger correspondant à un concept très connu des spécialistes va parfois refouler la nécessité de créer un sigle équivalent dans la langue d'arrivée. C'est le cas de

(angl.) ATN (franç.) *réseau de transition augmenté* --> ATN

(angl.) RTN (franç.) *réseau de transition récursif* --> RTN

qui sont utilisés massivement à la place de leur équivalent français, une fois que ce dernier a été spécifié dans le texte.

Les exemples comme *grammaire DCG* et *grammaire XG* formés sur le modèle **nom + joncteur ø + nom** n'existent pas en anglais bien qu'ils fassent usage du sigle anglais. Ce dernier joue ici le rôle de substantif. Cependant, le sigle est redondant car il reprend le terme **grammaire** contenu dans DCG et XG. On pourrait le qualifier de faux calque en ce sens qu'il copie un modèle soi-disant anglais qui n'en est pas un en réalité. C'est aussi une surcaractérisation par siglaison (Dubuc 1985 : 111).

Certaines unités lexicales sont issues d'une création nouvelle calquée littéralement sur l'anglais comme c'est le cas du terme *compositionnalité* dans *principe de compositionnalité*.

3. Conclusion

N'en déplaise à bon nombre de puristes, le calque est de loin le mode de formation qui sera amené à prendre le plus d'expansion dans le futur. C'est un moyen simple et rentable d'enrichir le vocabulaire des langues de spécialité lorsqu'il respecte la structure morphologique et syntaxique du français comme c'est le cas pour le présent corpus. Il est de loin préférable à l'emprunt intégral parce qu'il dénote un certain effort de francisation.

Par ailleurs, le calque en langue de spécialité n'a pas la même valeur qu'en langue courante; il sert à dénommer une réalité inexistante en français mais présente en anglais. Par contre, le calque en langue courante est souvent une pratique abusive car la réalité est déjà dénommée en français. C'est le cas pour l'expression "faire face à la musique" qui est traduite littéralement de l'expression anglaise "to face the music" alors qu'elle comporte dans son registre l'expression équivalente plus qu'acceptable "prendre le taureau par les cornes".

Un calque comme *réseau de transition augmenté* évoque une réalité nouvelle qui émane de l'anglais et, par conséquent, n'a pas d'équivalent en français ni dans toute autre langue. L'anglais est voué à contribuer dorénavant à l'enrichissement de la langue de la même façon que le latin et le grec l'ont fait pour les vocabulaires scientifiques des langues occidentales.

Vu de cette façon, le phénomène du calque dépasse largement les cadres de la linguistique et constitue une particularité des sociétés occidentales et par le fait même des langues qu'on y parle.

CONCLUSION

Les 50 notions étudiées dans le cadre de ce mémoire ne représentent qu'une partie de la terminologie du traitement automatique des langues naturelles. En général, on peut affirmer qu'elle répond aux mêmes modèles de formation du vocabulaire que l'informatique et l'intelligence artificielle.

Composé de deux à six constituants, le syntagme en traitement automatique des langues naturelles ne semble pas comporter de restriction quant à la longueur acceptable. Les syntagmes formés de trois constituants et plus subissent généralement une réduction par la voie du sigle emprunté à l'anglais ou traduit littéralement en français ou par l'omission d'une ou de plusieurs unités lexicales. C'est donc dire que la réduction que subissent ces syntagmes démontre que la lexie simple demeure le modèle que les usagers gardent comme point de référence et vers lequel ils tendent plus ou moins consciemment par le biais de la réduction. Certains syntagmes passent ainsi du statut de lexie complexe à celui de lexie simple ce qui augmente la valeur polysémique de ces termes lorsqu'ils sont pris à l'extérieur de leur champ sémantique propre.

La réduction par acronymie est une autre pratique courante en informatique, en intelligence artificielle et en traitement automatique des langues naturelles même si le présent corpus n'en comporte pas. L'absence d'acronyme serait ici attribuable au manque d'exhaustivité du corpus.

Tout comme l'informatique et l'intelligence artificielle le traitement automatique des langues naturelles a une nette propension à la synonymie causée par la grande variété

des modes de formation du vocabulaire. Il en résulte un flottement dans l'usage et un manque d'uniformité du vocabulaire.

Malgré ces points communs, il est tout de même permis de constater certaines particularités qui font que le vocabulaire du traitement automatique des langues naturelles se démarque de ceux de l'informatique et de l'intelligence artificielle. Aucun cas répondant aux modèles **nom + joncteur + infinitif** et **nom + nom**, opposant deux termes d'égale valeur, n'a été relevé dans le corpus alors qu'ils semblent assez courants dans les autres disciplines. Il faut préciser que le modèle **nom + adjectif** est le plus fréquent en traitement automatique des langues naturelles et c'est un modèle qui assure une très grande cohésion avec pour résultat que les cas de réduction de ce modèle sont inexistants.

Ce qui ressort de cette étude c'est que l'anglais en traitement automatique des langues naturelles tout comme en informatique et en intelligence artificielle influence grandement les modes de formation du vocabulaire lorsqu'il ne l'envahit pas. Mais la maîtrise du calque et de ses mécanismes apparaît comme un mode de formation rapide et efficace, le recours massif au syntagme en est un exemple signifiant, et un moyen de limiter la pénétration de l'anglais dans les langues de spécialité qui sont en constante évolution.

Une étude plus poussée de la terminologie du traitement automatique des langues naturelles permettrait de comparer les modes de formation utilisés en informatique et en intelligence artificielle afin d'obtenir une vue d'ensemble de la terminologie informatique et d'établir une typologie des modes de formation propre à cette discipline.

PRÉSENTATION DE LA BIBLIOGRAPHIE THÉMATIQUE

La bibliographie thématique comporte deux parties. Premièrement, une bibliographie terminologique qui regroupe tous les ouvrages de référence qui ont servi à la présentation méthodologique et à l'élaboration des dossiers de terminologie particulièrement à la rédaction de certaines rubriques telles que les définitions, les contextes ou les observations.

Et enfin, une bibliographie linguistique qui présente les ouvrages de référence ayant servi à l'analyse des données.

1. BIBLIOGRAPHIE TERMINOLOGIQUE

American Journal of Computational Linguistics (AJCL) (1980) : vol. 6, n° 1, pp. 1-34; (1981) : vol. 7, n° 4, pp. 243-256; (1982) : vol. 8, n° 1, pp. 12-25; (1984) : vol. 10, n° 2, pp. 81-96.

Artificial Intelligence (1980) : vol. 13, n° 3, pp. 231-278; (1982) : vol. 18, n° 3, pp. 327-367.

BONNET, Alain (1984) : **L'intelligence artificielle: promesses et réalités**, Paris, InterÉditions, 271 p.

Byte (1987) : vol. 12, n° 14, décembre, 396 p.

COLMERAUER, Alain (1970) : **Les systèmes-q ou un formalisme pour analyser et synthétiser des phrases sur ordinateur**. Département d'Informatique et de recherche opérationnelle, Université de Montréal, 46 p.

Computational Intelligence (1986) : vol. 2, n° 1, pp. 1-8.

CONDILLAC, Maryse (1986) : **Prolog: fondements et applications**, Paris, Bordas, vi + 314 p.

DANLOS, Laurence (1985) : **Génération automatique de textes en langue naturelle**, Masson, Paris, 240 p.

GAL, Annie, Guy LAPALME, et Patrick SAINT-DIZIER, (1989) : **Prolog pour l'analyse automatique du langage naturel**, Paris, Eyrolles, xii + 169 p.

- GEVARTER, William B (1984) : Artificial Intelligence, Expert Systems, Computer Vision and Natural Language Processing**, Noyes Publications, New Jersey, xiv + 225 p.
- HARRIS, Mary Dee (1985) : Introduction to natural language processing**, Reston Publishing Compagny, Inc., Reston, Virginia, xv + 368 p.
- I.R.I.S.A. (1986) : n° 573, octobre, Rennes, France, ii + 40 p.**
- Langages (1987) : vol. 87, septembre, Librairie Larousse, Paris, 130 p.**
- Linguisticae Investigationes (1980) : vol. 4, n° 1, pp. 39-101; n° 2, pp. 395-414.**
- Linguistics and Philosophy (1982) : vol. 3, n° 4, pp. 471 à 504; (1984) : vol. 7, n° 2, pp. 134-154.**
- Linguistic Structure Processing (1977) : publié par Antonio Zampolli, North-Holland Publishing Company, Netherland, xii + 585 p.**
- Linguistics (1985) : vol. 23, n° 2, pp. 213, 261.**
- Méthodologie-Informatique-Philosophie (M.I.P.) (1985) : n° 2, hiver, pp. 65-101.**
- Natural Language Understanding and Logic Programming (1985) : publié par Veronica Dahl, Patrick Saint-Dizier, Netherland, Elsevier Science Publishers B.V., North Holland, xi + 243 p.**
- PAVEL, Sylvia (1987) : «Intelligence artificielle», Les cahiers de terminologie, Secrétariat d'état du Canada, Direction générale de la terminologie et des services linguistiques, Ottawa, CT-31, juin, 176 p.**
- PELLETIER, Bertrand (1986) : Système d'interrogation de banque de données en langue naturelle, Document de travail #177, Mémoire de maîtrise, Université de Montréal, mars, vi + 269 p.**
- PITRAT, Jacques (1985) : Textes, ordinateurs et compréhension, Paris, Eyrolles, xiv + 202 p.**
- PLANTE, Pierre (1985) : Deredec : logiciel de traitement linguistique d'analyse de contenu des textes, UQAM.**
- POLGUERE, Alain (1984) : Programmation logique des interfaces en langue naturelle, Marseille, 27 p.**
- Pour la Science (1984) : n° 85, novembre, pp. 90-103.**
- R.A.I.R.O Informatique / Computer Science (1980) : vol. 14, n° 2, pp. 137-148.**
- La Recherche (1985) : vol. 16, n° 170, octobre, 1294 p.**
- Revue québécoise de linguistique (1985) : vol. 14, n° 2, 187 p.**

SAINT-DIZIER, Patrick (1987) : **Initiation à la programmation en Prolog**, Paris, Eyrolles, xii + 178 p.

SÉBILLOT, Pascale et Patrick SAINT-DIZIER (1986) : «Un processus de filtrage par arbres syntaxiques dans les grammaires discontinues», **Proceedings Journées Prolog-CNET**, pp. 467-471.

TAUM 73 (1973) : Groupe de recherche pour la traduction automatique, Rapport d'étape, Université de Montréal, iv + 262 p.

Technique et science informatiques (1982) : vol. 1, n° 2, pp. 32-133.

Terminogramme; Bulletin d'information terminologique et linguistique (1988) : Gouvernement du Québec, Office de la langue française, éditeur officiel du Québec, n° 46, janvier, 32 p.

TREMBLAY, Diane (1988) : «Traitement des langues naturelles et développement de banques de données linguistiques», dans **Terminogramme**, n° 46, janvier, pp. 11-14.

2. BIBLIOGRAPHIE LINGUISTIQUE

AUGER, Pierre (1979b) : «La syntagmatique terminologique, typologie des syntagmes et limite des modèles en structure complexe» , dans **Table ronde sur les problèmes de découpage du terme**, Ve congrès de l'Association internationale de linguistique appliquée, Montréal, Éditeur officiel du Québec, pp. 11-26.

AUGER, Pierre et Jean-Louis ROUSSEAU (1978a) : «**Méthodologie de la recherche terminologique**», coll. Études, recherches et documentation, Québec, Office de la langue française, Éditeur officiel du Québec, 82 p.

BONNARD, Henri (1978) : «Le syntagme», dans **Grand Larousse de la langue française**, Paris, Librairie Larousse, vol. 7, pp. 5899-5902.

BOULANGER, Jean-Claude (1978b) : «Néologie et terminologie», dans **Néologie en marche, série b: langues de spécialités**, n° 4, Office de la langue française, Québec, Éditeur officiel du Québec, p. 127.

BOULANGER, Jean-Claude (1985c) : **Initiation à la terminologie**. Notes de cours et travaux pratiques, Québec, Université Laval, août, 325 p.

BOULANGER, Jean-Claude (1987d) : «Un miroir aux alouettes : le calque en intelligence artificielle» , dans **Meta**, vol. 32, n° 3, septembre, pp. 326-331.

BOULANGER, Jean-Claude (1988e) : «Le syntagme en informatique: un projet de recherche», dans **Terminogramme**, n° 41, janvier, Montréal, pp. 22-23.

BOULANGER, Jean-Claude (1989f) : «La place du syntagme dans les dictionnaires de langue» , dans **Meta**, vol. 34, n° 3, septembre, pp. 360-369.

BOULANGER, Jean-Claude et Michelle RIVARD (1976a) : «Définition de la néologie» et «Présentation méthodologique» , dans **Néologie en marche, série b: langues**

de spécialités, cahier préparé par Jean-Claude Boulanger et Henriette Dupuis sous la direction de Pierre Auger. Régie de la langue française, Québec, Éditeur officiel du Québec, pp. xii-xxix.

CAYER, Micheline et Lise LEBEL-HAROU (1983) : **Guide de présentation des cahiers de Néologie en marche**, Gouvernement du Québec, Éditeur officiel du Québec, Québec, 80 p.

DUBUC, Robert (1985) : **Manuel pratique de terminologie**, Montréal, Linguatex, CILF, 102 p.

Essai de définition de la terminologie (1975) : Actes du colloque international de terminologie, tenu à Québec au Manoir du Lac Delage, du 5 au 8 octobre 1975 par la Régie de la langue française, 209 p.

GOFFIN, Roger (1979) : «Le découpage du terme à des fins lexicographiques : critères formels, sémantiques, quantitatifs et taxinomiques», dans **Table ronde sur les problèmes de découpage du terme**, Ve congrès de l'Association internationale de linguistique appliquée, Montréal, Éditeur officiel du Québec, pp. 159-168.

GUILBERT, Louis (1965a) : «Les cadres morphologiques du néologisme dans la formation du vocabulaire de l'aviation», dans **La formation du vocabulaire de l'aviation**, chapitre 5, Paris, Librairie Larousse, pp. 255-329.

GUILBERT, Louis (1970b) : «La dérivation syntagmatique dans les vocabulaires scientifiques et techniques», dans **Les langues de spécialité. (Analyse linguistique et recherche pédagogique)**. Actes du Stage de Saint-Cloud, stage tenu du 23 au 30 novembre 1967, Publication du Conseil de l'Europe, Strasbourg, Aidela, pp. 116-125.

GUILBERT, Louis (1973c) : «Théorie du néologisme», dans **Cahiers de l'Association internationale des études françaises**, n° 5, mai, pp. 9-29.

GUILBERT, Louis (1975d) : «Terminologie et linguistique», dans **Essai de définition de la terminologie**, Régie de la langue française, Québec, Éditeur officiel du Québec, pp. 13 à 26.

JASTRAB de SAINT-ROBERT, Marie-Josée (1987) : «Les syntagmes nominaux complexes en anglais et en français: éléments de réflexion», dans **Meta**, vol. 32, n° 3, septembre, pp. 260-266.

MARTINET, André (1979) : «Syntagme et syntème», **Table ronde sur les problèmes de découpage du terme**, Ve congrès de l'association internationale de linguistique appliquée, Montréal, Éditeur officiel du Québec, pp. 185-189.

MEUNIER-CRESPO, Mariette (1987) : «Les anglicismes dans la presse économique espagnole», dans **Meta**, vol. 32, n° 3, septembre, pp. 272-277.

NAKOS, Dorothy (1989) : «Étude comparée des modes de formation des lexies complexes», dans **Meta**, vol. 34, n° 3, septembre, pp. 352-357.

La normalisation linguistique, 1974) : Actes du colloque international de terminologie, tenu au Lac Delage (Québec) par l'Office de la langue française les 16, 17, 18 et 19 octobre 1973, iv + 257 p.

SANTOYO, Julio Cesar (1987) : «Traduction, fertilisation et internationalisation», dans **Meta**, vol. 32, n° 3, septembre, pp. 240-249.

Travaux de terminologie et de linguistique 1 (1978) : Office de la langue française, Québec, Éditeur officiel du Québec, 108 p.

VINAY, Jean-Paul (1979) : «Problèmes de découpage du terme», dans **Table ronde sur les problèmes de découpage du terme**, Ve congrès de l'Association internationale de linguistique appliquée, Montréal, Éditeur officiel du Québec, pp. 83-101.