



AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4

Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

http://www.cfcopies.com/V2/leg/leg_droi.php

<http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm>

UNIVERSITÉ DE METZ
U.F.R. DE LETTRES ET SCIENCES HUMAINES
CENTRE D'ANALYSE SYNTAXIQUE

MAILLOT Catherine

**LE TRAITEMENT DES PREPOSITIONS SPATIALES
DANS UN SYSTEME DE TRADUCTION AUTOMATIQUE
FRANÇAIS-ALLEMAND, ALLEMAND-FRANÇAIS,
BASE SUR L'UNIFICATION**

Thèse soutenue en vue de l'obtention du Doctorat de l'Université de Metz
le 11 Décembre 1997 à Metz

BIBLIOTHEQUE UNIVERSITAIRE DE METZ



031 151536 7

BIBLIOTHEQUE UNIVERSITAIRE LETTRES - METZ -	
N° Inv.	1997-106L
Cote	LIM3 97/18
Loc.	Magasin

Directeur de Thèse : Monsieur Pierre DIMON, Maître de Conférence
Université de Metz

Jury :
Monsieur Pierre DIMON, Maître de Conférence
Université de Metz
Monsieur Hans HALLER, Professeur
Université de Sarrebruck
Monsieur Jean-Claude LEJOSNE, Professeur
Université de Metz

à mes parents

REMERCIEMENTS

Pour commencer, je remercie mon directeur de thèse, Mr P. Dimon de m'avoir encadrée au cours de ses trois années passées à préparer ma thèse et d'avoir mis à ma disposition la station de travail du bureau 47.

Je remercie également le Professeur J.-C. Lejosne qui m'a autorisée à utiliser le matériel et les ouvrages de son bureau et qui a accepté de faire partie du jury.

Je leur associe, dans ma reconnaissance, le Professeur H. Haller, directeur de l'*Institut für Angewandte Informationsforschung* (I.A.I.) d'avoir accepté que je fasse mon stage dans son centre de recherche lors de mon D.E.S.S. et que je continue à travailler en collaboration avec les chercheurs de son centre, dans le cadre de ma thèse. Je le remercie également de sa présence dans le jury.

Parmi les chercheurs de l'I.A.I., grande est ma reconnaissance envers O. Streiter qui m'a aidé à résoudre bon nombre de problèmes auxquels je me suis heurtée lors de l'implémentation du programme en Cat2.

Parmi mes collègues, je tiens à remercier tout particulièrement R. Thiel pour toute l'aide qu'il m'a apporté au niveau informatique et la patience dont il a fait preuve, ainsi que pour tous les ouvrages qu'il a bien voulu me prêter.

Enfin, je tiens à remercier très sincèrement mes parents qui m'ont permis de mener à bien toutes mes études, je leur suis également reconnaissante pour leur patience et leur soutien tout au long de ma vie étudiante et en particulier au cours de ces trois dernières années. Je tiens également à exprimer toute ma gratitude à ma sœur, Sophie, qui m'a beaucoup aidée dans la frappe et dans la relecture de ma thèse.

Je leur associe dans ma reconnaissance toutes les personnes qui m'ont soutenue ou aidée d'une manière ou d'une autre, et notamment, Emmanuel, Anne, Olivier, Justine, Didier, Isabelle ; Delphine (de l'agence Schon et Brullard) ; ma famille et tous mes amis.

« Arriverai-je un jour à poser au moins les
assises d'une juste théorie de la préposition ?
Je n'ose avoir cette ambition, tant la question
soulève de difficultés »¹

¹ G. GUILLAUME : *Leçons de Linguistique*, 1948-1949, Série B, p. 165, publiées par R. Valin, Paris, Klincksieck, Québec, Presses de l'Université, Laval, 1971.

0. INTRODUCTION GENERALE.....	1
1. GENERALITES SUR LES PREPOSITIONS.....	3
1.1. EN FRANÇAIS	3
1.1.1. Le Robert : Dictionnaire Historique de la langue française.....	3
1.1.2. Le Larousse : <i>Grand Dictionnaire Universel du XIX^e siècle</i>	3
1.1.3. Trésor de la Langue Française	4
1.1.4. <i>Nouveau Larousse illustré</i>	4
1.1.5. Le Robert : <i>Dictionnaire alphabétique et analogique de la langue française</i>	5
1.1.6. G. Mounin : <i>Dictionnaire de la linguistique</i>	6
1.1.7. J.-F. Phelizon : <i>Vocabulaire de la linguistique</i>	6
1.1.8. <i>Grand Larousse de la Langue Française en 7 volumes</i>	6
1.1.9. Le Bidois : <i>Syntaxe du français moderne. Ses fondements historiques et psychologiques</i>	7
1.1.10. M. C. De Boer	7
1.2. EN ALLEMAND	7
1.2.1. Duden : <i>Deutsches Universal Wörterbuch A-Z</i>	7
1.2.2. Hennig Brinkman : <i>Die deutsche Sprache, Gestalt und Leistung</i>	7
1.2.3. Hermann Paul : <i>Deutsche Grammatik</i>	8
1.2.4. Duden, Band 4	8
1.3. POSTPOSITIONS ET CIRCOMPOSITIONS	9
1.3.1. Postpositions	9
1.3.2. Circompositions	9
1.4. CONCLUSION.....	9
2. LES PREPOSITIONS SPATIALES.....	10
2.1. LOCALISATION SPATIALE ET ORIENTATION.....	10
2.1.1. Généralités	10
2.1.2. Les espaces mentaux	12
2.1.2.1. Fauconnier	12
2.1.2.2. Les autres auteurs.....	12
2.2. PRÉPOSITIONS SPATIALES	13
2.2.1. Morphologie.....	13
2.2.1.1. En français	13
2.2.1.2. En allemand	15
2.2.2. Syntaxe.....	16
2.2.2.1. Généralités	16
2.2.2.2. La préposition et son régime.....	18
2.2.2.2.1. En français.....	18
2.2.2.2.2. En allemand.....	20
2.2.3. Sémantique.....	21
2.2.4. Pragmatique	26
2.2.5. Les verbes et la localisation	27
2.2.6. Les différentes prépositions spatiales en français.....	29
2.2.6.1. Les prépositions <i>à, de et en</i>	30
2.2.6.1.1. La notion de «prépositions vides»	30
2.2.6.1.2. La préposition <i>à</i>	31
2.2.6.1.2.1. Généralités.....	31
2.2.6.1.2.2. Emplois spécifiques de la préposition <i>à</i>	33
2.2.6.1.3. La préposition <i>de</i>	34
2.2.6.1.4. La préposition <i>en</i>	34
2.2.6.1.4.1. Généralités.....	34
2.2.6.1.4.2. Emplois spécifiques de la préposition <i>en</i>	35
2.2.6.1.5. Concurrence entre ces prépositions	36
2.2.6.1.5.1. Brøndal	36
2.2.6.1.5.2. Mantchev	36
2.2.6.1.5.3. Emploi de l'article avec ces prépositions.....	36
2.2.6.1.5.4. Concurrence entre <i>à et en</i>	37
2.2.6.1.5.5. <i>à et en</i> les noms d'îles	38
2.2.6.2. Les « paires prépositionnelles »	40
2.2.6.2.1. <i>avant / après</i>	41

2.2.6.2.2. sur / sous.....	42
2.2.6.2.3. dans / hors (de).....	43
2.2.6.2.4. devant / derrière	53
2.2.6.2.5. au-dessus de / au-dessous de	53
2.2.6.2.6. en haut de / en bas de	53
2.2.6.2.7. au-dedans de, en dedans de / au-dehors de, en dehors de.....	54
2.2.6.2.8. Les prépositions adverbialisées	54
2.2.6.3. Les autres prépositions.....	54
2.2.6.3.1. chez.....	54
2.2.6.3.2. contre.....	54
2.2.6.3.3. entre.....	54
2.2.6.3.4. jusqu'(e).....	55
2.2.6.3.5. par	55
2.2.6.3.6. parmi	55
2.2.6.3.7. pour	55
2.2.6.3.8. vers.....	55
2.2.6.3.9. Combinaison de prépositions	55
2.2.6.4. Concurrence entre les prépositions dites « vides » et les autres	56
2.2.6.4.1. à, sur et dans.....	56
2.2.6.4.2. en et dans.....	57
2.2.6.4.3. « Train » et ses prépositions	57
2.2.6.4.3.1. Train et à.....	58
2.2.6.4.3.2. Train et avec	58
2.2.6.4.3.3. Train et dans	58
2.2.6.4.3.4. Train et de / pour	58
2.2.6.4.3.5. Train et en.....	58
2.2.6.4.3.6. Train et par	59
2.2.6.4.3.7. Conclusion sur train et ses prépositions.....	59
2.2.7. Les prépositions spatiales en allemand.....	59
2.2.7.1. Généralités	59
2.2.7.1.1. Le système proposé par Marcq.....	59
2.2.7.1.1.1. Le système 1	59
2.2.7.1.1.1.1. Opposition intérieur ~ extérieur	60
2.2.7.1.1.1.2. Opposition contact ~ non contact.....	61
2.2.7.1.1.1.3. Les autres oppositions.....	61
2.2.7.1.1.2. Le système 2	62
2.2.7.1.1.3. Comparaison entre le système 1 et le système 2	65
2.2.7.1.1.4. Le système 3	65
2.2.7.1.1.5. Cas particuliers	67
2.2.7.1.2. Le système proposé par Schanen et Confais.....	68
2.2.7.1.2.1. Sans autre paramètre.....	68
2.2.7.1.2.2. Paramètre « intérieur~extérieur ».....	69
2.2.7.1.2.3. Paramètre « axes » (locatif = datif, directif = accusatif)	69
2.2.7.1.2.4. Autres paramètres	69
2.2.7.1.3. Le système proposé par Moilanen.....	71
2.2.7.2. Durch	71
2.3. TRAITEMENT INFORMATIQUE DE LA LOCALISATION SPATIALE.....	72
2.3.1. Le système TOUR.....	72
2.3.2. VITRA	74
2.3.2.1. CITY-TOUR.....	74
2.3.2.2. SOCCER.....	75
2.3.3. Système robotique KAMRO	75
2.3.3.1. Évaluation des relations spatiales	76
2.3.3.2. Informatisation des relations spatiales élémentaires	76
2.3.3.3. L'utilisation compositionnelle des prépositions spatiales	76
2.3.3.4. L'implémentation des relations spatiales compositionnelles.....	76
2.3.3.5. L'interprétation des t-uplets de relation	77
2.3.3.6. Vers un dialogue en langue naturelle	78
2.3.3.7. Interprétation des expressions de localisation.....	78

2.3.3.8. Synthèse sur les expressions de référence spatiales	79
2.3.3.9. Résumé	80
2.3.4. Les autres systèmes	80
2.4. CONCLUSION SUR LES PRÉPOSITIONS SPATIALES.....	80
3. GENERALITES SUR LES MODELES.....	82
3.1. LES CAS	82
3.1.1. Généralités	82
3.1.2. La grammaire des cas	84
3.1.2.1. Brøndal	84
3.1.2.2. Guillaume.....	84
3.1.2.3. Fillmore.....	85
3.1.2.4. Simmons	88
3.1.2.5. Schank.....	88
3.1.2.6. Pottier	90
3.1.2.7. Boguraev / Spark-Jones	93
3.1.2.8. Grimes.....	93
3.1.2.9. Feuillet.....	94
3.1.2.10. De Boer.....	94
3.1.2.11. Tableau récapitulatif.....	94
3.1.2.12. Conclusion	96
3.1.3. Cas & prépositions	96
3.1.3.1. Généralités	96
3.1.3.2. En allemand	98
3.1.4. Conclusion	99
3.2. L'UNIFICATION	101
3.2.1. Qu'est-ce que l'unification ?	101
3.2.1.1. Définition.....	101
3.2.1.2. Propriétés	101
3.2.2. Les formalismes basés sur l'unification.....	102
3.2.2.1. Historique	102
3.2.2.2. Généralités	103
3.2.2.3. Les principales notions utilisées.....	105
3.2.2.3.1. Trait.....	105
3.2.2.3.2. Structure de traits	105
3.2.2.3.3. Extension entre structures de traits.....	107
3.2.2.3.4. Subsomption.....	107
3.2.2.3.5. Généralisation	107
3.2.2.3.6. Union prioritaire.....	108
3.2.2.4. Présentation des formalismes.....	108
3.2.2.4.1. Grammaire à clauses définies (DCG, <i>Definite Clause Grammar</i>).....	108
3.2.2.4.2. Grammaire d'unification fonctionnelle (FUG, <i>Functional Unification Grammar</i>).....	109
3.2.2.4.3. Grammaire d'unification catégorielle (UCG, <i>Unification Categorical Grammar</i>) ...	110
3.2.2.4.3.1. Naissance, motivations de départ, objectifs.....	110
3.2.2.4.3.2. Notations utilisées dans UCG	112
3.2.2.4.3.3. Le formalisme UCG.....	113
3.2.2.4.3.3.1. Description du formalisme.....	113
3.2.2.4.3.3.1.1. Le signe	114
3.2.2.4.3.3.1.2. Les règles.....	115
3.2.2.4.3.3.1.3. Les catégories.....	115
3.2.2.4.3.3.1.4. L'ordre.....	115
3.2.2.4.3.3.1.5. La sémantique.....	116
3.2.2.4.3.3.2. FDP (French Dialog Parser) - un fragment UCG pour le français	116
3.2.2.4.3.3.2.1. Le signe	116
3.2.2.4.3.3.2.2. Les règles.....	117
3.2.2.4.3.4. Le traitement des prépositions dans UCG.....	118
3.2.2.4.3.5. Le formalisme PATR-II.....	119
3.2.2.4.4. Grammaire lexicale fonctionnelle (LFG, <i>Lexical Functional Grammar</i>).....	120
3.2.2.4.4.1. Naissance, motivations de départ et objectifs	120
3.2.2.4.4.2. Organisation du modèle.....	121

3.2.2.4.4.3. Notations utilisées dans LFG	121
3.2.2.4.4.4. Le formalisme LFG	121
3.2.2.4.4.4.1. Description du modèle	121
3.2.2.4.4.4.2. Principes et théorèmes	125
3.2.2.4.4.4.3. Tableau comparatif notation traditionnelle / notation LFG	125
3.2.2.4.4.5. Le traitement des prépositions dans LFG	126
3.2.2.4.4.6. Conclusion	127
3.2.2.4.5. Grammaire syntagmatique généralisée (GPSG, <i>Generalized Phrase Structure Grammar</i>)	128
3.2.2.4.5.1. Naissance, motivations de départ et objectifs	128
3.2.2.4.5.2. Organisation du modèle	128
3.2.2.4.5.3. Notations utilisées dans GPSG	129
3.2.2.4.5.4. Le formalisme GPSG	129
3.2.2.4.5.4.1. Description du modèle	129
3.2.2.4.5.4.1.1. Les catégories	129
3.2.2.4.5.4.1.2. Les règles	130
3.2.2.4.5.4.1.2.1. Les règles ID	130
3.2.2.4.5.4.1.2.2. Les règles LP	131
3.2.2.4.5.4.1.2.3. Le format ID /LP	132
3.2.2.4.5.4.1.2.4. Les métarègles	133
3.2.2.4.5.4.2. Principes et théorèmes	133
3.2.2.4.5.4.3. Tableau comparatif notation traditionnelle / notation GPSG	137
3.2.2.4.5.5. Le traitement des prépositions dans GPSG	138
3.2.2.4.5.6. Conclusion	139
3.2.2.4.6. Grammaire syntagmatique guidée par la tête (HPSG, <i>Head Phrase Structure Grammar</i>)	139
3.2.2.4.6.1. Naissance, motivations de départ, objectifs	139
3.2.2.4.6.2. Organisation du modèle	140
3.2.2.4.6.3. Notations utilisées dans HPSG	140
3.2.2.4.6.4. Le formalisme HPSG	140
3.2.2.4.6.4.1. Description du modèle	140
3.2.2.4.6.4.2. Principes et théorèmes	144
3.2.2.4.6.4.3. Tableau comparatif notation traditionnelle / notation HPSG	146
3.2.2.4.6.5. Le traitement des prépositions dans HPSG	147
3.2.2.4.6.6. Conclusion	147
3.2.2.4.7. Grammaire d'arbres adjoints (TAG, <i>Tree Adjoining Grammar</i>)	149
3.2.2.4.7.1. Naissance, motivations de départ, objectifs	149
3.2.2.4.7.2. Notations utilisées dans TAG	149
3.2.2.4.7.3. Le formalisme TAG	149
3.2.2.4.7.3.1. Description du modèle	149
3.2.2.4.7.3.2. Principes et théorèmes	157
3.2.2.4.7.3.3. Tableau comparatif notation traditionnelle / notation TAG	157
3.2.2.4.7.4. Le traitement des prépositions dans TAG	158
3.2.2.5. Conclusion générale sur les grammaires d'unification	160
4. TRAITEMENT AUTOMATIQUE	164
4.1. GÉNÉRALITÉS	164
4.1.1. Lexique	165
4.1.2. Syntaxe	167
4.1.3. Sémantique	168
4.2. LE TRAITEMENT ET LA TRADUCTION DES PRÉPOSITIONS	168
4.3. LES PROJETS D'IMPLÉMENTATION UTILISANT LES GRAMMAIRES D'UNIFICATION	174
4.3.1. Les projets d'implémentation utilisant UCG	174
SYSLID : SYntaktisch-Semantische Linguistikkomponente für sprachverstehende Dialogsysteme..	174
4.3.2. Les projets d'implémentations utilisant LFG	176
4.3.2.1. Le projet LFG PARGRAM	176
4.3.2.2. Le Xerox LFG Grammar Writer's Workbench	178
4.3.2.3. Le Konstanz LFG Workbench	178
4.3.2.4. Le système LFG de Avery Andrew	179
4.3.2.5. Un parser LFG pour le turc	179

4.3.3. Les projets d'implémentation utilisant GPSG	179
4.3.3.1. GPSG-Tools.....	179
4.3.3.2. GPSG parser	179
4.3.3.3. Kit-Fast	180
4.3.4. Les projets d'implémentation utilisant HPSG	182
4.3.4.1. ALE : Attribute Logic Engine.....	182
4.3.4.2. ALEP : Advanced Linguistic Engineering Platform	183
4.3.4.3. BABEL	184
4.3.4.4. JPSG : Japanese Phrase Structure Grammar	184
4.3.4.5. HDRUG	184
4.3.4.6. HPSG-PL.....	185
4.3.4.7. PLEUK	185
4.3.4.8. SOAS HPSG Ellipsis	186
4.3.4.9. TDL : Type Description Language	186
4.3.4.10. TFS : Typed Feature Structures System 6.1.....	186
4.3.4.11. UBS : UnifikationsBasierte Sprache.....	187
4.3.4.12. Projet de Grammaire Syntagmatique du français.....	188
4.3.4.13. Autres travaux.....	188
4.3.5. Les projets d'implémentation utilisant TAG	188
4.3.5.1. Le système XTAG	188
4.3.5.1.1. Présentation générale du système	188
4.3.5.1.2. Analyseur morphologique	189
4.3.5.1.3. Système de balisage des parties du discours.....	189
4.3.5.1.4. Parser.....	189
4.3.5.1.5. Base de données syntaxique	189
4.3.5.1.6. Assignment de cas.....	190
4.3.5.1.7. Les prépositions	192
4.3.6. Les autres projets d'implémentation utilisant les grammaires d'unification.....	192
4.3.6.1. Formalismes grammaticaux et analyse du français	192
4.3.6.2. Projet de recherche : environnement de test grammatical - GTU (Grammatik-test- umgebung).....	193
4.3.6.2.1. Évolutions de GTU - Extension pour les grammaires catégorielles, en particulier UCG	194
4.3.6.2.2. Développement d'un parser GTU-HPSG pour KoGiTum	194
4.3.6.2.3. Développement d'une interprétation sémantique GTU dans les grammaires d'unification	194
4.3.6.2.4. Perfectionnement de GTU.....	195
4.3.6.3. LKP : Linguistic Kernel Processor	195
4.3.6.4. GULP : Graph Unification Logic Programming	195
4.3.6.5. CUF	196
4.3.6.6. Autres travaux.....	196
4.4. EUROTRA	196
5. IMPLEMENTATION.....	198
5.1. PRÉSENTATION DE CAT2.....	198
5.1.1. Niveaux de représentation.....	199
5.1.2. Règles vs objets.....	201
5.2. LE FORMALISME CAT2.....	202
5.2.1. Constantes et variables.....	202
5.2.2. Un système de traits	202
5.2.2.1. Les traits.....	202
5.2.2.2. L'unification	203
5.2.2.3. Les contraintes	203
5.2.2.4. Les attributs fixes.....	204
5.2.2.5. Les valeurs fixes	204
5.2.3. Les niveaux	204
5.2.3.1. Les générateurs	205
5.2.3.2. Les translateurs	205
5.2.3.3. Commun.....	206
5.2.4. Les types de règles	206

5.2.4.1. Les règles b	207
5.2.4.2. Les règles f	207
5.2.4.3. Les règles t	208
5.2.4.4. Les règles tf	208
5.2.5. Processus utilisés dans l'interprétation des règles	208
5.2.5.1. Input	209
5.2.5.1.1. Les unités de traduction	209
5.2.5.1.2. Les tokens	209
5.2.5.2. L'analyse morphologique	210
5.2.5.2.1. Les inputs numériques	210
5.2.5.2.2. Les mots inconnus	210
5.2.5.3. L'analyse syntaxique	211
5.2.5.4. Les transformations	211
5.2.5.5. La synthèse syntaxique	212
5.2.5.6. La synthèse morphologique	212
5.2.5.7. Output	212
5.2.5.8. Le chemin de traduction	212
5.2.5.9. Traduction robuste	212
5.2.5.9.1. Niveau de traducteur robuste	213
5.2.5.9.2. Traduction robuste seulement	213
5.3. PRÉSENTATION DU PROJET	213
5.3.1. Récapitulatif sur les prépositions spatiales	213
5.3.1.1. Les lieux géographiques	214
5.3.1.1.1. Les continents	214
5.3.1.1.2. Les pays	214
5.3.1.1.3. Les régions	215
5.3.1.1.4. Les départements	215
5.3.1.1.5. Les villes	215
5.3.1.1.6. Les îles	216
5.3.1.1.7. Les autres lieux géographiques	216
5.3.1.1.7.1. Les déserts	216
5.3.1.1.7.2. Les montagnes	216
5.3.1.1.7.3. Les mers, les océans	216
5.3.1.1.7.4. Les lacs	217
5.3.1.1.7.5. Les rivières, les fleuves	217
5.3.1.2. Les bâtiments	217
5.3.1.2.1. Les bâtiments	217
5.3.1.2.2. Cas particulier	217
5.3.1.3. La maison	218
5.3.1.3.1. La maison	218
5.3.1.3.2. Porte, fenêtre, mur	218
5.3.1.3.3. Le toit	218
5.3.1.3.4. Cas particuliers	219
5.3.1.4. Les meubles	219
5.3.1.4.1. La table	219
5.3.1.4.2. L'armoire	219
5.3.1.4.3. Le tiroir	219
5.3.1.4.4. Le lit	220
5.3.1.4.5. Les sièges	220
5.3.1.4.5.1. La chaise	220
5.3.1.4.5.2. Le fauteuil	220
5.3.1.4.5.3. Le canapé	220
5.3.1.4.5.4. Le divan	220
5.3.1.5. Les contenants	220
5.3.1.5.1. Cas général	220
5.3.1.5.2. Les verbes de remplissage	221
5.3.1.5.3. Les verbes d'extraction	221
5.3.1.6. Les personnes	221
5.3.1.6.1. après	221

5.3.1.6.2. chez	221
5.3.1.6.3. devant / derrière	222
5.3.1.6.4. Corps humain	222
5.3.1.6.5. Cas particuliers.....	222
5.3.1.7. Divers.....	222
5.3.1.7.1. La terre	222
5.3.1.7.1.1. La terre comme sol	222
5.3.1.7.1.2. La terre vue de la mer	223
5.3.1.7.1.3. Cas particuliers	223
5.3.1.7.2. La mer	223
5.3.1.7.2.1. La mer comme lieu	223
5.3.1.7.2.2. La mer par opposition au bateau.....	223
5.3.1.7.2.3. Cas particuliers	223
5.3.1.7.3. Le port.....	224
5.3.1.7.4. L'eau	224
5.3.1.7.5. Le sol.....	224
5.3.1.7.6. La boue.....	224
5.3.1.7.7. La campagne, la ville.....	224
5.3.1.7.7.1. La campagne.....	224
5.3.1.7.7.2. La ville.....	224
5.3.1.7.8. La route	225
5.3.1.7.9. La rue	225
5.3.1.7.10. Le marché.....	225
5.3.1.7.11. La cour	225
5.3.1.7.12. Le ciel.....	225
5.3.1.7.13. Le sable	226
5.3.1.7.14. Phénomènes météorologiques	226
5.3.1.7.15. Premier plan, arrière plan	226
5.3.1.7.16. La liste.....	226
5.3.1.7.17. La tombe	226
5.3.1.7.18. La place.....	227
5.3.1.7.19. La poche.....	227
5.3.1.7.20. La chaîne, les fers.....	227
5.3.1.7.21. L'air.....	227
5.3.1.7.22. Les points cardinaux.....	227
5.3.1.7.23. Cas particuliers.....	228
5.3.1.7.24 Divers.....	228
5.3.1.8. Conclusion	231
5.3.1.8.1. Tableau français	231
5.3.1.8.2. Tableau allemand	234
5.3.2. Le programme	237
5.3.2.1. La structure morphologique	238
5.3.2.1.1. Le dictionnaire morphologique	238
5.3.2.1.1.1. Les règles b.....	238
5.3.2.1.1.1.1. La ponctuation	239
5.3.2.1.1.1.2. Les déterminants	239
5.3.2.1.1.1.3. Les prépositions	240
5.3.2.1.1.1.4. Les noms	242
5.3.2.1.1.1.5. Les verbes	242
5.3.2.1.1.2. Les règles f	243
5.3.2.1.2. L'analyse morphologique française.....	243
5.3.2.2. Les dictionnaires communs.....	245
5.3.2.2.1. Les déterminants et les prépositions	245
5.3.2.2.2. Les noms	245
5.3.2.2.3. Les verbes	251
5.3.2.3. La structure syntaxique française.....	252
5.3.2.3.1. Les règles du niveau cs.....	252
5.3.2.3.1.1. Les règles b.....	252
5.3.2.3.1.2. Les règles f	253

5.3.2.3.1.3. Les règles kill.....	253
5.3.2.3.2. L'analyse syntaxique.....	253
5.3.2.4. Transformation.....	257
5.3.2.4.1. Les règles t.....	258
5.3.2.4.2. Les règles f.....	258
5.3.2.5. La structure relationnelle française.....	259
5.3.2.5.1. Les règles du niveau IS.....	259
5.3.2.5.1.1. Les règles b.....	259
5.3.2.5.1.2. Les règles f.....	259
5.3.2.5.1.3. Les règles kill.....	260
5.3.2.5.2. L'analyse relationnelle.....	260
5.3.2.6. Le transfert.....	265
5.3.2.6.1. Les règles t.....	265
5.3.2.6.2. Les règles tf.....	266
5.3.2.6.3. Les règles kill.....	266
5.3.2.7. La structure relationnelle allemande.....	267
5.3.2.7.1. Les règles du niveau isde.....	267
5.3.2.7.1.1. Les règles b.....	267
5.3.2.7.1.2. Les règles f.....	267
5.3.2.7.2. La synthèse relationnelle allemande.....	267
5.3.2.8. Transformation.....	272
5.3.2.9. La structure syntaxique allemande.....	273
5.3.2.9.1. Les règles.....	273
5.3.2.9.2. La synthèse syntaxique allemande.....	273
5.3.2.10. La structure morphologique allemande.....	277
5.3.2.10.1. Le dictionnaire morphologique allemand.....	277
5.3.2.10.2. La synthèse morphologique allemande.....	277
5.4. CONCLUSION SUR L'IMPLEMENTATION.....	279
6. CONCLUSION GENERALE.....	281

Glossaire
 Index des noms propres
 Index des projets
 Bibliographie

0. INTRODUCTION GENERALE

Le présent travail s'intéresse au « Traitement des prépositions spatiales dans un système de traduction français-allemand, allemand-français basé sur l'unification » ; il ne représente qu'une étude fragmentaire, le point de départ d'une recherche plus vaste qui devra inclure l'ensemble des prépositions. Nous avons limité le « stock » des prépositions aux prépositions spatiales pour diverses raisons :

- * elles sont les plus concrètes et les plus représentatives et semblent être celles qui se prêtent le mieux à l'implémentation,
- * il est préférable d'approfondir un type de prépositions, et d'essayer d'appliquer le programme obtenu (avec les modifications nécessaires) sur les autres,
- * les équivalences entre les deux langues sont complexes à définir, de plus il faut souvent tenir compte du contexte, par conséquent il vaut mieux bien délimiter le corpus à analyser.

Cette étude ne prendra pas en compte les prépositions qui se rattachent aux verbes, les prépositions qui se traduisent pas une préposition zéro dans l'autre langue (*aller à l'école, die Schule besuchen*), la traduction des idiomes ou proverbes. En effet, pour ces derniers, il semble assez difficile d'établir des correspondances entre les deux langues dans la mesure où les termes ou les expressions n'ont pas du tout la même signification, comme le montre notamment l'exemple suivant : *tourner autour du pot* se traduit en allemand par *wie die Katze um den heißen Brei gehen* (la traduction littérale de l'allemand serait : tourner comme le chat autour de la purée chaude). On constate, en effet, qu'il n'y a aucune équivalence entre les termes des deux langues.

Cette étude est réalisée dans le cadre de travaux de traduction automatique menés à l'*Institut für Angewandte Informationsforschung* (IAI) de Sarrebruck. C'est pour cette raison que nous avons décidé de travailler dans le couple de langues français-allemand et dans l'optique des grammaires d'unification, utilisées notamment dans le programme CAT2 développé à l'IAI et employé dans notre travail.

Dans une première partie, nous donnerons quelques définitions du mot « préposition » proposées dans divers ouvrages et dictionnaires.

Nous nous intéresserons, ensuite, dans une seconde partie, plus particulièrement aux prépositions spatiales. Nous les analyserons du point de vue morphologique, syntaxique, sémantique et pragmatique, puis nous verrons de quelle manière le verbe peut intervenir dans le choix de la préposition. Nous présenterons ensuite les différentes prépositions spatiales françaises et allemandes et nous terminerons par le traitement informatique de la localisation spatiale.

Nous poursuivrons notre étude par une présentation des modèles dans le cadre desquels nous travaillons : tout d'abord les cas puisque 1) l'allemand est une langue casuelle, et que 2) nous devons définir les rôles sémantiques pour le choix des prépositions ; puis les grammaires d'unification puisque le programme que nous utilisons est basé sur l'unification, nous présenterons, pour chacune d'elles, le formalisme, les principes et théorèmes... ainsi que les implémentations informatiques dans lesquelles elles sont employées.

Nous consacrerons la quatrième partie au traitement automatique.

Enfin nous terminerons notre travail en proposant, dans une cinquième partie, l'implémentation de notre système. Nous présenterons tout d'abord le système CAT2 avec tous les éléments qui le composent, puis le projet. Nous établirons, tout d'abord un récapitulatif sur les prépositions spatiales, puis nous expliquerons le programme de traduction en lui-même : les dictionnaires, les règles utilisées aux différents niveaux... et pour terminer nous montrerons, à partir d'un exemple, comment fonctionnent l'analyse et la synthèse dans la traduction d'une phrase.

Nous inclurons différents index (noms propres, notions utilisées, projets), un glossaire (les mots présents dans le glossaire sont repérés dans le texte par le symbole [◊]) et en annexe, l'ensemble du programme.

1. GENERALITES SUR LES PREPOSITIONS

1.1. EN FRANÇAIS

Cette première partie se compose essentiellement de définitions du terme « préposition » tirées de divers dictionnaires et ouvrages de grammaire. Celles-ci sont classées par ordre de généralité : en effet, chaque définition supplémentaire offre des aspects et des notions non encore exposés par les précédentes.

1.1.1. Le Robert : Dictionnaire Historique de la langue française

Le Robert nous propose une définition étymologique et nous indique que le mot préposition est emprunté (XIV^e s.) au latin *praepositio*, *-onis*, proprement « action de mettre devant », spécialisé en grammaire pour désigner un mot grammatical servant à introduire un complément (Cicéron) et employé au sens qualitatif d'« état préférable ». Le mot est dérivé de *praepositum* supin de *praeponere*, « placer devant » et au figuré « mettre en tête, préférer » (→ préposer). Le français n'a gardé que la spécialisation grammaticale du mot latin. Le mot s'est aussi employé pour « fait de placer en avant » (1480), et « action de faire preuve (de courage) » (1531), valeurs propres au moyen français et qui ont disparu au XVII^e siècle.

1.1.2. Le Larousse : *Grand Dictionnaire Universel du XIX^e siècle*

Le Larousse offre une définition plus précise incluant des données syntaxiques, morphologiques et sémantiques :

La préposition est un mot invariable qui se place entre deux termes, pour exprimer un rapport logique qui lie le second avec le premier. Les prépositions inséparables sont celles qui, appartenant à des langues étrangères, leur ont été empruntées avec certains mots composés, et ne sont jamais employées seules, comme *ad*, *per*, *pro*, *ré*, *dis*, dans les mots *admettre*, *protéger*... On a pareillement appelé prépositions inséparables des prépositions qui entrent dans la formation des mots composés comme *avant*, *arrière*, *contre* dans les mots *avant-bras*, *contresens*... Les prépositions servent à exprimer les mille et une nuances de la pensée, et, comme elles sont peu nombreuses, la même préposition a souvent des acceptions très diverses.

Les principaux rapports que les prépositions expriment sont ceux de lieu, d'ordre, de temps, d'union, de but, de cause, de séparation, d'opposition, d'indication :

1. Écrivez les injures sur le sable et les bienfaits sur l'airain
2. Je crains Dieu, et après Dieu, je crains principalement celui qui ne le craint pas
3. La cigale chante pendant l'été
4. Il faut tâcher de bien vivre avec tout le monde
5. Il faut manger pour vivre et non vivre pour manger
6. Il fut exempté, attendu ses infirmités
7. Il travaille toute la semaine excepté le dimanche
8. Un enfant bien élevé ne doit rien faire malgré ses parents
9. Du pain et du fromage, voilà son déjeuner

1.1.3. Trésor de la Langue Française

Dans le *Trésor de la Langue Française*, l'entrée « préposition » propose la définition suivante :

« Partie du discours invariable qui, placé devant un élément à valeur nominale (substantif pour *Pierre*, pronom pour *lui*, adverbe pour *aujourd'hui*, syntagme infinitif pour *le retrouver*, proposition conjonctive pour *qu'elle retrouve*), le lie dans un rapport sémantique donné (approche, éloignement, intériorité, privation...) en le subordonnant à un autre élément de la phrase (substantif : *le livre de Pierre*, adjectif : *capable de réussir*, adverbe : *loin de vous*, verbe : *rêver de quelque chose*) ou à la phrase entière (*à vrai dire, il...*) ; mot ou locution qui appartient à cette partie du discours ».

A l'intérieur du même article, on peut lire une comparaison avec la langue anglaise :

« L'anglais dans le type syntagmatique *goosbery wine* « vin de groseilles », *gold watch* « montre en or »..., exprime par l'ordre pur et simple des termes des rapports que le français moderne marque par des prépositions ».

Les prépositions, comme les conjonctions, n'assument pas de fonctions. Elles ne jouent dans la phrase qu'un rôle de struments pour reprendre le mot de Damourette et Pieton (lat. *struo* « je construis, j'édifie »), c'est-à-dire qu'elles explicitent le rapport syntaxique de deux termes qui, eux, assument une fonction. Les prépositions se distinguent des conjonctions de coordination. Ces dernières, en effet, unissent des termes ou des groupes parallèles.

On trouve également dans cette définition, une citation de Wagner Pinchon (1976) § 511 :

« ... chaque préposition d'un idiome donné dessine, dans ses emplois divers, une certaine figure où se coordonnent son sens et ses fonctions et qu'il importe de restituer si l'on veut donner de l'ensemble de ses particularités sémantiques et grammaticales une définition cohérente. Cette figure est commandée par le même système sublogique qui gouverne les fonctions casuelles. Il va de soi qu'une description guidée par ce principe doit embrasser, pour prendre sa force démonstrative, la totalité des prépositions et la totalité des relations casuelles d'un état de langue » (E. Benveniste, *Problème de linguistique générale*, 1966, p. 13 ?).

1.1.4. Nouveau Larousse illustré

Dans le *Nouveau Larousse illustré*, on peut lire sous l'entrée « préposition » la définition suivante :

« Particule invariable qui sert à unir deux mots en marquant le rapport qu'ils ont entre eux ».

Elle est toutefois complétée par une analyse étymologique et morphologique : les prépositions sont des mots invariables. Cependant plusieurs proviennent de mots variables. Ainsi en français, les prépositions se sont formées de substantifs comme *malgré*, d'adjectifs comme *sauf*, d'impératifs comme *voici* (*vois ici*), de participes présents comme *durant* ou passés comme *excepté*. De même *chez* vient du substantif latin *casa*, « maison » et *près* du participe latin *pressum*, « pressé contre ». En grec certaines prépositions ont la forme casuelle : *kharin* (accusatif singulier), *peri* (locatif singulier)... D'autres ne peuvent être rattachées à aucune catégorie morphologique, telles les prépositions grecques *apo*, *pro*, *en* (latin *ab*, *pro*, *in*).

Du point de vue de leur fonction, les prépositions sont assez rapprochées des adverbes. C'est ce qui apparaît surtout dans les langues comme le sanscrit, le grec et

aussi l'allemand où les prépositions servent même de préfixes aux verbes. En grec, l'augment *e* et en allemand le préfixe *ge* peuvent s'intercaler entre la préposition-préfixe et le radical verbal, grec *peripherô*, imparfait *periepheron*, allemand *umgehen*, participe passé *umgegangen*. Dans le grec homérique et dans les verbes allemands à particules séparables, la préposition peut se détacher du verbe et se placer en un autre endroit de la phrase. Les linguistes enseignent que, dans les langues où les prépositions sont construites avec des mots déclinables, il ne faut pas considérer les prépositions comme régissant tel ou tel cas, mais comme des adverbes qui ne font que renforcer l'idée exprimée par le cas lui-même. Cependant, la liaison habituelle de telle préposition avec tel cas déterminé a fini par donner à la préposition elle-même la valeur du cas dont elle était accompagnée. Dans les langues romanes, et en particulier en français, les prépositions ont peu à peu rendu inutiles les désinences casuelles. Les rapports exprimés par les diverses prépositions sont la tendance (*à, contre, jusque...*), l'origine (*de, par...*), la manière (*avec, par, sans...*), le temps (*avant, après...*), et le lieu (*dans, en, chez...*).

1.1.5. Le Robert : *Dictionnaire alphabétique et analogique de la langue française*

Le Robert reprend l'idée de mot invariable mais intègre une notion nouvelle : celle des cas, et propose la définition suivante :

« Mot grammatical, invariable, servant à introduire un complément en marquant le rapport qui unit ce complément au mot complété. La préposition joue en français le rôle rempli par les cas dans les langues flexionnelles ».

Ce dictionnaire introduit, en plus, le concept de locutions prépositives qu'il présente comme un emploi explétif de certaines prépositions, introduisant non un complément mais une apposition (*la ville de Rome*) ou un attribut (*prendre pour témoin*) ; préposition entrant dans la formation de mots composés (*à-propos, en-cas*). On peut distinguer (de Boer) des prépositions « pleines » exprimant des rapports bien définis (*avant, après, sans...*), des prépositions pouvant exprimer soit un simple cas, soit un rapport précis (*avec, en, par, pour...*) et des prépositions « vides » ou « casuelles » (*à, de*).

Une analyse syntaxique permet à l'auteur de présenter la notion de régime :

« la préposition peut s'employer sans régime avec une valeur adverbiale (*c'est selon*) ou par ellipse (*je voterai contre*). [Elle] se place immédiatement avec son régime, sauf dans des cas où un terme court est intercalé (surtout avant un infinitif) dans des tours interrogatifs avec savoir (*je ne sais qui, je ne sais quoi*) et des réciproques formés de deux indéfinis (*l'un, l'autre*) ».

Dans ce même article, P. Robert nous propose une citation de Dauzat extraite de la *Grammaire rais. de la langue française*, p. 343 et 344 :

« En latin classique déjà, de nombreuses prépositions ... étaient employées pour préciser les cas qui étaient en nombre insuffisant pour exprimer les divers types de rapports. La préposition est arrivée à régir le cas, puis, en développant son emploi, à le rendre utile. En français, la syntaxe des prépositions se confond avec la syntaxe des compléments. Toutefois certains compléments se passent de préposition. Le complément direct est indiqué essentiellement par sa place, après le verbe... Des compléments de temps, de prix, de manière, se passent de préposition... Il reste, en outre, quelques vestiges de l'ancienne syntaxe, tandis que des ellipses modernes ont réintroduit des tournures analogues ».

1.1.6. G. Mounin : *Dictionnaire de la linguistique*

Dans son *Dictionnaire de la linguistique*, G. Mounin nous indique que le terme préposition désigne une classe de mots, ou de locutions invariables, ou particules, qui ont une fonction grammaticale et qui, comme c'est le cas en latin et en grec, se trouvent en général (dans les langues classiques d'Europe) juste avant le nom ou le syntagme nominal auxquels ils confèrent l'autonomie fonctionnelle.

1.1.7. J.-F. Phelizon : *Vocabulaire de la linguistique*

Dans son *Vocabulaire de la linguistique*, J.-F. Phelizon, comme les autres auteurs, considère la préposition comme un mot invariable, appartenant à l'espèce subjonctive et exprimant un rapport de subordination entre plusieurs choses. Il fait cependant une distinction entre deux sortes de prépositions :

1. La préposition forte, indispensable sur le plan sémantique : *J'ai lu chez un conteur de fables* (La Fontaine, III, 18)
2. La préposition faible, qui précise un lien que la seule juxtaposition des unités grammaticales établissait déjà : *ils fonderaient à eux deux une maison de banque* (Balzac, *César Birotteau*).

Cependant certaines prépositions ne sont ni fortes, ni faibles ; elles font partie intégrante, ou d'une forme adverbiale occasionnelle ou d'une forme verbale : *Personne ne répond de mes lettres que moi, et je ne réponds de rien que de mes lettres* (Pascal, *Provinciales*, XVIII).

1.1.8. *Grand Larousse de la Langue Française en 7 volumes*

Le *Grand Larousse* propose une définition fonctionnelle, et présente les caractéristiques communes des prépositions :

- * elles sont invariables,
- * elles marquent la fonction d'un mot qu'elles précèdent dans la chaîne orale, l'élément qu'elles régissent s'appelle le régime, ce dernier peut être un nom, un pronom, un verbe à l'infinitif, un adjectif, un adverbe, éventuellement une proposition entière mais nominalisée par une conjonction : *des souvenirs de quand on était au collège*,
- * elles subordonnent généralement leur régime à un mot de la phrase : on vérifie ce fait en comparant la validité des syntagmes qu'on peut composer avec tel ou tel mot de la phrase et le groupe préposition + régime. Le support du groupe préposition + régime peut être un verbe, un nom, un pronom, un adjectif, un adverbe.

1.1.9. Le Bidois : *Syntaxe du français moderne. Ses fondements historiques et psychologiques*

G. & R. Le Bidois citent plusieurs définitions sur les prépositions.

Le *Dictionnaire général* définit la préposition comme étant :

« une partie du discours invariable qui, placée devant un nom, une proposition infinitive, etc., les lie par un rapport déterminé à un terme précédent ».

Cette définition leur paraît insuffisante car elle pourrait s'appliquer aussi bien à certaines conjonctions (par exemple *et* ou *mais*), et le «etc.» ne nous permet pas de connaître les autres éléments de la phrase que peut unir la préposition.

Dans le *Précis de grammaire historique* (p. 608), la préposition se définit comme :

« un élément du langage dont le rôle le plus ordinaire est d'introduire un complément circonstanciel, c'est-à-dire un complément qui marque un rapport de lieu, de temps, de manière, etc. ».

Cette idée leur paraît exacte pour les prépositions «pleines» telles que *avant*, *après*, *au milieu de...*, mais inapplicable dans les cas où la préposition est «vide», notamment pour *à* et *de*.

1.1.10. M. C. De Boer

Dans son *Essai sur la Syntaxe moderne de la préposition en français et en italien*, M. C. de Boer propose une définition de la préposition par rapport à deux types de syntaxe :

« Dans la syntaxe mobile est préposition : la particule qui relie et subordonne à une autre partie de la phrase : un nomen (nom, adjectif, nom de nombre, participe passé-adjectif ou construction nominale [telles que la phrase substantive conjonctionnelle et la phrase relative]), un pronomen, un infinitif, un adverbe, une construction ou phrase adverbiale ».

« Dans la syntaxe figée est préposition : toute ancienne préposition qui a gardé, lorsque la construction s'est figée, une certaine indépendance par rapport aux autres éléments de la construction ».

1.2. EN ALLEMAND

1.2.1. Duden : *Deutsches Universal Wörterbuch A-Z*

Wort, das Wörter zueinander in Beziehung setzt und ein bestimmtes (räumliches, zeitliches, o.ä) Verhältnis angibt (z. B. *der Ball liegt auf / in / unter dem Schrank*).

« Mot qui relie d'autres mots et qui indique une relation précise (spatiale, temporelle, ...) : *la balle est sur / dans / sous l'armoire* ».

1.2.2. Hennig Brinkman : *Die deutsche Sprache, Gestalt und Leistung*

Beziehungswörter haben keine veränderliche Gestalt wie Substantiv, Adjektiv, Pronomen und Verbum, aber als Präpositionen treten sie in Verbindung mit bestimmten Formen des Substantivs.

« Les mots de liaison n'ont pas de forme variable comme le substantif, l'adjectif, le pronom et le verbe, mais en tant que préposition, ils entrent en relation avec des formes précises du substantif ».

Die Präpositionen, die aus Substantiven hervorgegangen sind, verbinden sich mit dem Genetiv wie das Substantiv selbst.

« Les prépositions, qui sont dérivées de substantifs se combinent avec le génitif comme le substantif lui-même ».

Im allgemeinen unterscheidet unsere Sprache zwischen dem Beziehungswort als Präposition (d. h. in fester Verbindung mit dem Substantiv) und als Modifizierung des Verbuns (*von, ab*) : viele Beziehungswörter können aber in beiden Bereichen antreten (z. B. *durch*).

« En général notre langue différencie le mot de liaison utilisé comme préposition (c'est-à-dire en relation avec un substantif) et utilisé comme modifiant du verbe (*von, ab*) ; cependant, de nombreux mots de liaisons peuvent apparaître dans les deux catégories ».

Für ein Beziehungswort als Präposition können vier Merkmalen formuliert werden :

- 1) Unveränderlichkeit der Gestalt
- 2) Verbindung mit einem Substantiv, das so seinen Bezug enthält
- 3) Kennzeichnung des Substantivs (Genetiv, Dativ, Akkusativ)
- 4) Voranstellung des Beziehungswort

« Pour un mot de liaison, utilisé en tant que préposition, on peut faire quatre remarques :

- 1) Invariabilité de la forme
- 2) Rattachement à un substantif, qui renferme ainsi son rapport
- 3) Marquage du substantif (génitif, datif, accusatif)
- 4) Pré-position du mot de liaison ».

1.2.3. Hermann Paul : *Deutsche Grammatik*

Die älteste Schicht der Präpositionen geht zurück auf Ortsadverbia. Diese traten, ursprünglich nur zu Verben. Stand noch ein Kasus daneben, so war dieser eigentlich von Verbum abhängig, das Adverb diente nur zu genaueren Bestimmung des im allgemeinen schon durch den Kasus bezeichneten Raumverhältnisses. Von hier aus war eine doppelte Entwicklung möglich. Entweder schloß sich das Adverb näher an das Verbum an, oder das Adverb ging eine engere Verbindung mit dem Kasus ein und wurde so zur Präposition.

« Les prépositions tirent leur origine des adverbes de lieu. A l'origine, ceux-ci apparaissaient seulement avec des verbes. S'il y avait un cas, celui-ci dépendait du verbe, l'adverbe servait seulement à donner des indications précises du rapport spatial, qui, en général, était déjà décrit par le cas. A partir de là une double évolution a été possible. Soit l'adverbe s'est rapproché du verbe, soit l'adverbe s'est rapproché du cas et s'est transformé en préposition ».

1.2.4. Duden, volume 4

Die Präpositionen oder Verhältniswörter sind in ihrer Form nach unveränderlich. Sie sind weder Satzglied noch Attribut, sondern werden immer in Verbindung mit einem anderen Wort gebraucht, dessen Kasus in der Regel von ihnen bestimmt wird. Die Präposition steht im allgemeinen am Anfang des Präpositionalgefüges : Präposition + Substantiv : *aus* dem Zimmer (gehen), Präposition + Pronomen : *bei* ihnen (wohnen), Präposition + Adjektiv : *bei* weitem, Präposition + Adverb : *nach* unten (gehen). Durch

die Präposition wird das Wort, das im Präpositionalgefüge steht, in der Regel an ein anderes Wort angeknüpft, und zwar ein Verb, ein Substantiv, ein Adjektiv.

« Les prépositions sont, du point de vue de leur forme, invariables. Elles ne sont ni membres de phrase ni attributs, mais sont toujours utilisées en combinaison avec un autre mot, dont elles régissent ordinairement le cas. La préposition se situe en général au début du groupe prépositionnel : préposition + substantif : (sortir) de la chambre, préposition + pronom : (habiter) chez eux, préposition + adjectif : de beaucoup, préposition + adverbe : (aller) vers le bas. Grâce à la préposition, le mot qui se trouve dans le groupe prépositionnel est en relation avec un autre mot, à savoir un verbe, un substantif, un adjectif ».

Die meisten Präpositionen sind aus Lokaladverbien entstanden, mit denen früher das im allgemeinen bereits durch den Kasus des Substantivs bestimmte Raumverhältnis genauer gekennzeichnet wurde.

« La plupart des prépositions tirent leur origine d'adverbes de localisation, grâce auxquels la relation spatiale déjà définie, d'une façon générale, par le cas du substantif, était plus précisément décrite ».

1.3. POSTPOSITIONS ET CIRCOMPOSITIONS

1.3.1. Postpositions

Dans son *Dictionnaire de la linguistique*, G. Mounin précise dans beaucoup de langues des particules ayant des fonctions grammaticales semblables à celles des prépositions latines, grecques ou françaises se trouvent après le nom ; c'est pourquoi on les appelle postpositions. Il en donne la définition suivante :

« postposition se dit d'un élément (phonème, monème, mot, catégorie grammaticale) placé à la suite d'un autre élément avec lequel il forme un groupe ».

1.3.2. Circompositions

Une circomposition est une préposition composée, dont les deux éléments qu'elle comporte sont situés de part et d'autre de son régime (*von meinem Fenster aus*). Les circompositions n'existent qu'en allemand.

1.4. CONCLUSION

À la fin de cette première partie, on constate qu'il existe des similitudes dans la définition des prépositions dans les deux langues de l'étude :

- * elles sont invariables,
- * elles expriment différents rapports : lieu, temps, but, cause...
- * en règle générale, elles sont rattachées à un substantif.

Il faut également signaler les différences qu'elles présentent :

- * il existe des prépositions, des postpositions et des circompositions en allemand, alors qu'en français on ne connaît que les prépositions,¹
- * les prépositions allemandes sont suivies d'un cas mais pas les prépositions françaises.

¹ Il est intéressant de signaler que l'anglais a un terme générique, le terme adposition, qui regroupe les notions de prépositions, postpositions et circompositions, alors que l'allemand et le français utilisent les termes spécifiques.

2. LES PREPOSITIONS SPATIALES

Les relations spatiales sont celles qui manifestent de la façon la plus concrète l'inscription dans la langue du corps du locuteur et de son expérience physique du monde.¹ Nous les analyserons dans un premier temps avant de passer à l'étude des prépositions spatiales proprement dites. Il faut signaler que les travaux sur les prépositions spatiales et sur les emplois spatiaux des prépositions se sont beaucoup multipliés.

2.1. LOCALISATION SPATIALE ET ORIENTATION

2.1.1. Généralités

Diverses études ont été menées sur la description de l'espace : on peut noter notamment les travaux de Vandeloise (1986) et de Herskovits (1986) qui constituent les références les plus importantes pour le français et pour l'anglais. Les grammaires cognitives et les travaux qui s'y rattachent accordent également une grande importance à l'espace (cf. Langacker, 1987, 1991).

Lang distingue l'orientation induite assignée par rapport à l'axe vertical de l'espace environnant, et la perspectivation induite assignée par rapport à la « ligne de vue » réelle ou potentielle de l'observateur. Il fait une distinction entre :

- * les objets orientés de façon inhérente : ils ont des côtés spécifiques et on peut les interpréter comme étant dans une position normale, sans tenir compte de la façon dont ils sont localisés dans l'espace environnant (ex. : *livre*),
- * les objets orientés de façon canonique : ils doivent être en position normale afin de remplir leur fonction (ex. : *armoire, coffre*),
- * les objets neutres par rapport à l'orientation spatiale et qui n'ont pas de côtés spécifiés (ex. : *balle, cube*).

Les entités mobiles peuvent subir deux types de relations :

- * des relations de position, statiques ou topologiques : elles expriment la position d'une entité par rapport à une autre
- * des relations directionnelles, dynamiques ou projectives : elles expriment un mouvement ou l'orientation d'une entité vers une autre.

Dans de nombreux cas il existe un sens topologique en plus du sens projectif, comme l'équivalent allemand de l'anglais *into* : le *in* allemand est utilisé dans le sens topologique comme dans le sens projectif, alors que l'anglais est lexicalement plus spécifique pour exprimer ces relations spatiales : *in* est utilisé dans des relations topologiques et *into* dans le sens projectif.²

¹ A.-M. BERTHONNEAU, P. CADIOT : *Les prépositions, méthodes d'analyse*, Presses Universitaires de Lille, Lille, 1993, p. 7.

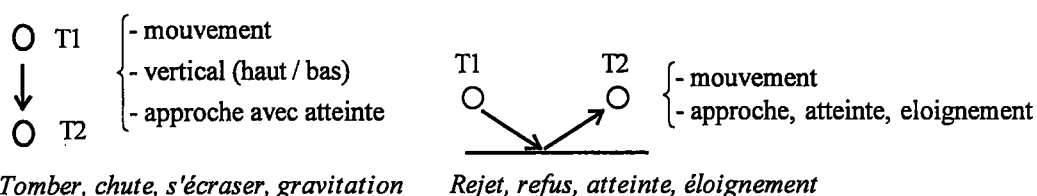
² C. ZELINSKY-WIBBELT : *The semantics of prepositions. From Mental Processing to Natural Language Processing*, Mouton de Gruyter, Berlin, New-York, 1993, p. 7.

Dans chaque relation une partie différente de la scène est considérée comme le premier plan : la grammaire cognitive y fait référence en terme de trajector. La relation implique la façon dont le trajector est localisé par rapport à un arrière plan présumé, auquel on fait référence en terme de landmark. Langacker utilise les termes de trajector et de landmark pour les profils mentaux qui sont spécifiquement construits pour des buts expressifs, le trajector correspondant de préférence aux réalisations linguistiques comme le sujet ou le verbe et le landmark à l'objet. D'autres linguistes tels que Schirra et Hottenroth, partisans de l'approche à deux niveaux (comme Lang, Kaufmann et Pribbenow) préfèrent respectivement les termes d'objet localisé et d'objet de référence. L'objet mobile ou susceptible de l'être dans une scène est aussi celui que l'on doit considérer comme le premier plan (*un chat dans un fauteuil* : le chat est considéré comme le premier plan).

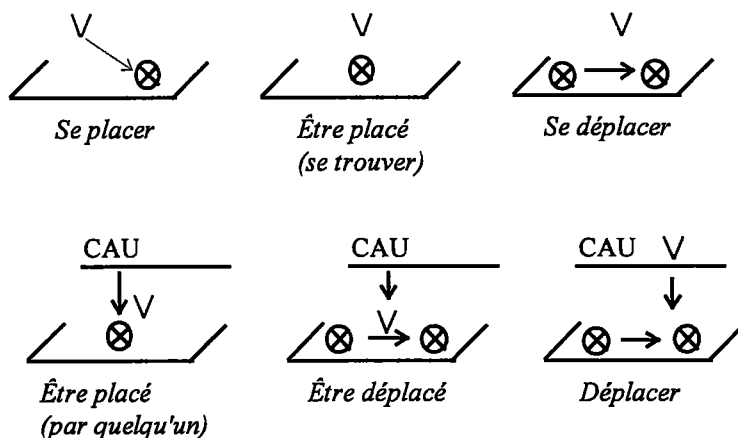
Les noèmes spatiaux proposés par Pottier sont les suivants :

- * mouvement / position
- * approche / éloignement
- * nature du repère (orienté ou non, intériorité ou non)
- * orientation (verticale, horizontale)
- * distance du repère

Exemples ¹ :



Dans les langues naturelles, le point de visée de l'énonciateur conduit à des solutions linguistiques différentes :



Les points importants dans un déplacement sont son origine, son terme et la trajectoire qui les lie.

¹ Les graphiques sont extraits de B. POTTIER : *Théorie et analyse en linguistique*, Hachette Supérieur, Paris, 1992, p. 94.

2.1.2. Les espaces mentaux

2.1.2.1. FAUCONNIER

Les fonctions pragmatiques lient les domaines les uns aux autres (par ex. un livre est relié à son auteur...). Les métaphores et les analogies reflètent la capacité humaine (universelle) à relier des domaines en se basant sur des connexions vécues de types différents. Les capacités cognitives semblent inclure la capacité à établir des cadres, qui permettent de faire une utilisation maximale des données que l'on possède. Ces cadres comprennent également des rôles pour les participants, dont certains n'ont une valeur qu'à un moment précis.

Il existe des principes généraux qui régissent le relation entre une entité et l'expression linguistique utilisée pour s'y référer. Le plus important est le Principe d'Accès (également appelé principe ID) qui établit qu'une expression qui nomme ou décrit une entité peut être utilisée faire référence à une entité qui appartient à un autre domaine seulement si ce dernier est accessible de façon cognitive à partir du premier, et s'il y a une connexion entre les deux entités. La théorie des espaces mentaux fournit un modèle de connexion entre la sémantique et la cognition, ceci permet à Fauconnier de présenter des résultats importants et de proposer des concepts théoriques qui sont supposés rendre compte des régularités observées dans la relation cognition langage. Les travaux de Fauconnier sur les espace mentaux ont donc fourni un modèle général qui permet d'étudier les nombreuses interactions entre les connexions cognitives et le langage naturel.

Dans une conversation, on ne parle pas uniquement de ce que l'on voit, on s'entretient également sur ce qui pourrait être, ce que l'on espère, ce qui pourrait arriver... et l'on se réfère à différents types de choses : le temps, les mots, les états, les attitudes propositionnelles, les modalités épistémiques, les images... Les éléments représentés dans les divers espaces ne sont pas eux-mêmes les référents (au sens standard) et les descriptions ne sont pas les descriptions des éléments mentaux. Si un espace mental correspond partiellement au monde, alors un élément et sa description correspondront d'une certaine façon à la réalité. L'idée de base est que lorsque l'on parle ou que l'on pense, des espaces mentaux se créent, se structurent et sont rattachés grâce à la grammaire, à la culture et au contexte.¹

2.1.2.2. LES AUTRES AUTEURS

Kay a établi le Principe de Parcimonie. Il affirme que l'auditeur construit les interprétations d'un texte de manière à utiliser au maximum les informations héritées ou déjà établies, plutôt que de créer, en permanence, de nouvelles entités.

Selon Bransford, Barclay et Franks (1972), la compréhension d'un texte n'est pas un processus purement interprétatif mais plutôt un processus constructif. C'est en analysant

¹ G. FAUCONNIER, E. SWEETSER : *Spaces, worlds and grammar*, University of Chicago Press, London, 1996, p. 11.

les données linguistiques que le lecteur peut construire une description sémantique de la situation décrite dans le texte.

Van Dijk et Kintsch (1983) sont d'avis que le lecteur construit une représentation de la situation sur la base d'une représentation du discours propositionnel sémantique.

D'après Johnson-Laird (1980,83) et Garham (1981), le concept de modèles mentaux développe ces hypothèses de sorte que le lecteur ne construit pas uniquement une description (propositionnelle) de la situation mais une représentation (non propositionnelle) de la situation elle-même.

Selon Glasgow et Papadias (1992), les modèles mentaux sont représentés dans la mémoire de travail, ils font une distinction, à l'intérieur de celle-ci, entre la représentation spatiale et la représentation visuelle.

« La représentation spatiale d'une image montre symboliquement les composants de l'image, et conserve les propriétés spatiales pertinentes »...

« La représentation visuelle décrit l'espace occupé par une image comme une zone d'occupation. On peut l'utiliser pour retrouver les informations telles que la forme, la distance et la taille relatives ». ¹

Franklin et Tversky (1990) travaillent, quant à eux, sur l'accès des objets dans les modèles mentaux selon leur position par rapport au lecteur.

2.2. PREPOSITIONS SPATIALES

2.2.1. Morphologie

2.2.1.1. EN FRANÇAIS

Par leur origine, on peut classer les prépositions spatiales dans différentes classes. Nous reprenons le classement proposé dans le *Larousse* en 7 volumes (pp. 4588-4589).

1) Prépositions ou adverbes latins conservés :

- à (lat. *ad* et *ab*)
- contre (lat. *contra*)
- de (lat. *de*)
- en (lat. *in*)
- entre (lat. *inter*)
- anc. frçs *fors* (lat. *foris*, «hors de») préposition et adverbe, et frçs anc. et moderne *hors* (tiré de *dehors*, lat. *deforis*)
- anc. frçs *joste* (lat. *juxta*), «près de»
- par (lat. *per*)
- pour (lat. *pro*)
- anc. frçs *rier* (lat. *retro*, adv. et prép.), «derrière»
- sous (lat. *subtus*, adv.)
- sur anc frçs *sor(e)*, *sour(e)*, *seur(e)* (lat. *super* ou *supra*)

¹ J. GLASGOW, D. PAPADIAS : « Computational imagery » in *Cognitive Science* 16, p. 356 cité par R. HÖRNIG ET AL. : *Object Access in Mental Models under Different Perspectives Induced by Linguistic Expressions*, Kit-Report 126, Berlin, 1996, p. 2.

- anc. frçs *tres, tries* (lat. *trans*, «au-delà»), «derrière»
- *vers* (lat. *versus*, adv. combiné avec *in* ou *ad* : *ad Oceanum versus*, «du côté de l’Océan»).

2) Préposition + préposition ou adverbe :

- *dans*, anc. frçs *denz* (lat. *de intus*, «de dedans»)
- *deçà* (lat. *de ecce hac*)
- anc. frçs. *dejoste* (lat. *de juxta*), «à côté de»
- *delà* (lat. *de illac*)
- *de par* (lat. *de per* ?)
- *derrière* (lat. *de retro*)
- *dessous* (lat. *de subtus*)
- anc. frçs *desor, desur* (lat. *de super*)
- *dessus* (lat. *de *susum*)
- *devant* (lat. *de ab ante*)
- *devers* (lat. *de versus*)
- *dusque* anc. frçs (lat. *de usque*, cf. *jusque*)
- *encontre* (lat. *in contra*)
- anc. frçs. *enjosque, enjusque* (lat. *inde usque*), «jusqu’à»
- *envers* (lat. *in versus*)
- *jusque* (lat. *de usque*, ou *enjusque* par aphérèse ¹).

Dans le premier groupe, on retrouve les prépositions qui sont venues directement du latin, dans le second celles que le français a composées avec des mots latins. Mais on peut également ajouter celles que le français a fabriquées :

- * par combinaison de mots français : *vis-à-vis*,
- * par composition : *face à* (préposition et nom), *au milieu de* (deux prépositions et un nom), *loin de* (préposition et mot-outil).

Les locutions composées avec *par-* sont françaises, mais anciennes : *par-dessous*, *par-dessus* (XII^e s.), *par-derrière*, *par-devant*, *par-devers* (XIII^e s.). Puis sont nées, par coalescence ² de syntagmes libres, toutes les prépositions du type *à côté de* (XVI^e s.), *autour de* (XV^e s.)...

Du point de vue de l’origine historique des prépositions, on distingue :

- * celles qui sont venues directement du latin : *à* (*ad*), *de* (*de*), *entre* (*inter*)...
- * celles que le français a composées :
 - d’après des mots latins : *avant* (*ab ante*), *derrière* (*de retro*)
 - par combinaison de mots français :
 - > particules : *depuis* (*de post*)
 - > préposition + nom : *parmi* (*par mi*)
 - par mutation de valeur :
 - > adjectif « transvalué » en préposition : *plein la bouche*
 - par composition :
 - > mot-outil combiné avec des prépositions : *loin de*, *près de*

¹ PHON. Suppression d’un ou plusieurs phonèmes au début d’un mot (ex. : bus pour autobus).

² PHON. Contraction de deux unités phoniques contiguës en une seule.

Les prépositions étaient, à l'origine, des adverbes suivis d'un cas, quand le cas a perdu sa signification, l'adverbe s'est transformé en préposition, c'est pourquoi il y a des ressemblances tant au niveau de leur sens que de leur emploi. Tous ces mots mi-adverbes mi-prépositions indiquent un rapport de lieu et contiennent tous déjà une préposition : *auprès*, *parmi*... et la plupart d'entre eux peuvent servir à former des locutions prépositives : *auprès de*, *au dessus de*...

	à	au	de	dès	en	jusque	par	pour
dedans - dehors - dessus- dessous						?		
devant - derrière								
ici			?					
là				dès là que				
avant							aupar-avant	
arrière								
haut								
bas								
ailleurs								
loin					de - en -			
çà								

TABLEAU 1 : COMBINAISONS POSSIBLES ENTRE ADVERBES ET PREPOSITIONS ¹

2.2.1.2. EN ALLEMAND

On peut classer les prépositions allemandes selon différents critères :

- * celles qui assurent aussi le rôle de « particules verbales » : *an, auf, aus, bei, durch, hinter, mit, nach, über, um, unter, von, vor, wider, zu*.
- * celles qui sont dérivées à partir de :
 - bases nominales : *anfangs, dank*...
 - bases adjectivales : *unweit*...
 - bases participiales : *während, entsprechend*...
 - amalgames syntaxiques, que l'on peut scinder en :
 - GN : *diesseits*...
 - GP : *anstelle / an Stelle*
- * celles qui peuvent être « associées » à une autre préposition ou à un adverbe : *von über dem Fluß her*.

On utilise encore certains adverbes en guise de prépositions : *links, rechts, seitab* + génitif pour *links von, rechts von, seitab von* + datif.

En règle générale, les prépositions se trouvent au début du groupe prépositionnel :

- * préposition + substantif : *an die Wand* (lehnen)
- * préposition + pronom : *bei ihnen* (wohnen)

¹ Ce tableau est extrait de G. & R. LE BIDOIS : *Syntaxe du français moderne. Ses fondements historiques et psychologiques*, 2^e édition revue et complétée, édition Picard, Paris, 1971, p. 590.

On peut avoir des pré-positions ¹, des post-positions et des circom-positions. Dans les deux premiers cas les prépositions sont continues (*aus, mit, nach*) alors qu'en circom-position elles sont discontinues (*von ... herab*).

Il peut y avoir contraction entre une préposition et la marque de l'article :

- * standard (écrit et oral) : *am, beim, im, vom, zum, ans, ins, zur*
- * seulement oral familier (sauf lexies) : *hinterm, überm, unterm, vorm, aufs, durchs, fürs, hinters, übers, ums, unters, vors, hintern, übern, untern*

La contraction est obligatoire dans de nombreuses lexies (*am Leben bleiben...*), devant certains noms géographiques (*im Iran*) et dans certains noms propres qui correspondent à des titres (*Kapitän zur See*). En revanche, elle est impossible lorsque *der, die* ou *das* ont, en plus de leur fonction de marqueur d'identification, une valeur :

- * démonstrative (substitution possible : *dieser* ou *jener*) : *an dem Samstag war ich nicht zu Hause,*
- * anaphorique [◊] : *zu der Zeit war ich nicht zu Hause,*
- * exophorique [◊] : *bei dem [da] möchte ich nicht Lehrling sein,*
- * cataphorique [◊] (substitution possible : *derjenige, diejenige, dasjenige*) : *ich spreche von dem Mann, der neben mir saß.*

DATIF SINGULIER

masculin	neutre	féminin
	am	zur
	beim	
	hinterm	
	im	
	überm	
	vom	
	vorm	
	zum	

ACCUSATIF SINGULIER

masculin	neutre	féminin
	ans	
	aufs	
	durchs	
	fürs	
hintern	hinters	
	ins	
übern	übers	
	ums	
untern	unters	
	vors	

TABLEAU 2 : CONTRACTIONS POSSIBLES ENTRE PREPOSITIONS ET ARTICLES

2.2.2. Syntaxe

2.2.2.1. GENERALITES

On peut classer les prépositions françaises selon leur fonction :

- * les prépositions introduisant les compléments à valeur circonstancielle :
 - celles qui introduisent le complément circonstanciel adjoind. La préposition possède sa pleine valeur de sens ; aucun élément n'impose la

¹ Préposition (en un seul mot) correspond à la catégorie grammaticale, pré-position (avec un trait d'union) signifie que la préposition est placée devant son régime.

présence de telle ou telle préposition, c'est le sens à donner au complément qui est déterminant (il marche dans / vers la ville),

- celles qui introduisent le complément circonstanciel intégré : là encore le choix de la préposition n'est imposé par aucun élément, contrairement au complément d'objet dont la présence est imposée par le verbe (il habite à Paris / en ville / chez ses parents),
- celles qui introduisent le complément du nom ou de l'adjectif à valeur circonstancielle ; la construction est libre, le choix de la préposition se fait en fonction du sens (un week-end à la campagne / chez des amis),
- * les prépositions qui introduisent des compléments imposés : au début, le sens de la préposition a déterminé son emploi après le verbe, mais l'emploi s'est figé et n'est plus susceptible de varier à l'heure actuelle ; ce sont essentiellement des prépositions à faible valeur de sens, elles précèdent :
 - un complément d'objet du verbe (il pense à elle),
 - un complément d'objet de l'adjectif (il est difficile à faire),
 - un complément d'objet du nom (l'arrivée du train).

Au niveau syntaxique, il existe plusieurs types de construction :

- * la forme la plus courante est celle où le nom-cible apparaît comme sujet et le nom-site comme complément prépositionnel introduit par un verbe d'état tel que notamment *être* ou *se trouver*, qui pourraient être considérés comme prototypes ⁰ ; (les prépositions les plus usitées sont *à*, *dans*, *sur*). On obtient des phrases de la forme : $N_{0(\text{cible})} \quad V_{\text{état}} \quad \text{Prep}_{\text{loc}} \quad N_{1(\text{site})}$
- * on peut aussi trouver la forme : $N_{\text{cible}} \quad V_{\text{état}} \quad N_{\text{site}}$ avec des verbes de nature transitive tels que *couvrir*, *occuper*, *remplir*, *inonder*... Rien ne marque syntaxiquement le complément comme complément de lieu : la substitution se fait en *le*, *cela*, *y*, l'interrogation en *quoi* (*les livres occupent toute la table*)
- * on peut trouver une construction $N_{\text{site}} \quad V_{\text{état}} \quad N_{\text{cible}}$ où les rôles syntaxiques sont inversés : le nom-site fonctionne comme sujet et le nom-cible comme complément d'un verbe transitif tel que *contenir*, *emprisonner*, *entourer*, *surplomber*... (*Cette armoire contient tous mes livres*). « L'ordre change très nettement la thématization du discours, et il n'est pas sûr qu'il préserve le rapport que sont censés exprimer les termes de cible et site ». ¹
- * ceci est peut-être encore plus vrai pour un quatrième type de construction $N_{\text{site}} \quad V_{\text{état}} \quad \text{Prep}_{\text{loc}} \quad N_{\text{cible}}$ proche de la précédente mais dans laquelle l'idée de recouvrement ou de saturation s'ajoute au repérage spatial. Les verbes sélectionnent un complément introduit par *de*, *en*, *sous* : *déborder de*, *crouler sous*, *être plein de*... (*l'armoire croule sous les livres*).

¹ A. BORILLO : « A propos de la localisation spatiale », in *Sur les compléments circonstanciels*, Langue Française 86, Larousse, Paris, 1990, p. 76.

Un moyen de faire la différence entre un complément verbal et un complément circonstanciel est de déplacer le complément de lieu en début de phrase :

- * déplacement en tête de phrase sans changement de l'ordre sujet-verbe :
 - Il n'y a pas de problème pour un complément circonstanciel : *dans cette salle, les gens somnolent / dans le lointain, de petites lumières brillent.*
 - Le déplacement sans inversion du sujet et du verbe semble mal accepté par les verbes considérés comme Vloc et dont le complément est obligatoire par nature : **dans la boîte, une vieille gravure se trouve.*

Cette construction inversée révèle parfois de véritables différences de sens dans le verbe : *dans l'armoire, les chaussures étaient rangées / les chaussures étaient rangées dans l'armoire.*

- * déplacement en tête de phrase avec changement de l'ordre sujet-verbe. Il semble que tous les verbes considérés comme Vloc acceptent cette construction : *sur la table était placée une corbeille de fruits / dans la boîte se trouvait une vieille gravure.*

2.2.2.2. LA PREPOSITION ET SON REGIME

2.2.2.2.1. En français

La préposition régit un syntagme nominal, et l'objet désigné par ce syntagme nominal est l'un des deux termes mis en relation par la préposition. Dans le cas d'indications de lieu, Döpke et Schwarze appellent cet objet, l'objet localisant. Le deuxième terme est un objet, dont l'objet localisant spécifie l'endroit en interaction avec la préposition.

Selon Schwarze, l'objet localisé ne peut pas être déterminé de façon uniforme : l'indication de lieu peut porter sur :

- * l'individu désigné par le sujet : *dans son lit, Paul entend les oiseaux*
- * le complément d'objet direct de la phrase : *Paul entend les oiseaux dans le jardin*
- * plusieurs objets à la fois : *Paul répare sa voiture dans la cour.*

Quand une phrase ne contient pas de syntagme dont on puisse dire qu'il désigne un objet localisé, on peut résoudre ce problème en envisageant deux solutions :

- * on fait une analyse de la structure profonde \diamond en y introduisant des individus non mentionnés dans la phrase pour obtenir des objets localisés manquant dans la structure de surface \diamond : *on travaille à la cave.*
- * on considère que la phrase décrit un état de choses (procès, action) dont l'endroit est précisé par l'indication de lieu : *il pleut à Constance.*

Préposition	+ Ø	+ GN	+ Inf	+ Complétive
à	-	+	+	+ (à ce que)
chez	-	+	-	-
contre	+	+	-	-
dans	-	+	-	-
de	-	+	-	+ (de ce que)
derrière	+	+	-	-
devant	+	+	-	-
en	-	+	-	+ (en ce que)
entre	+	+	-	-
jusqu'à	-	+	+	+ (jusqu'à ce que)
par	-	+	(+)	+ (parce que)
parmi	-	+	-	-
pour	+	+	+	+
sous	-	+	-	-
sur	-	+	-	+ (sur ce que)
vers	-	+	-	-

TABLEAU 3 : LES PRINCIPALES CONSTRUCTIONS DES PREPOSITIONS ¹

La préposition peut avoir plusieurs types de régimes :

- * un nom, un pronom : *sur la table - près de lui*,
- * un adverbe de lieu : *il sort d'ici*,
- * un groupe de mots déjà précédé d'une préposition : *il va jusqu'à Paris*,
- * une proposition relative : *il est sorti par où il était entré*.

Il existe également des prépositions à régime implicite :

- * il ne s'agit pas de personne : il est fréquent de ne pas exprimer le régime de la préposition s'il a déjà été exprimé avant ou si on peut le déduire facilement du contexte ou de la situation
 - après les prépositions *derrière*, *devant*, l'omission du régime appartient à l'usage le plus général : *Paul marche devant*,
 - à la place des prépositions *dans* (et *en*), *hors*, *sur* et *sous*, on emploie sans régime *dedans*, *dehors*, *dessus* et *dessous* : *je le croyais hors de la maison, il était dedans*,
 - avec les locutions prépositives dont le dernier élément est *de*, on omet régulièrement le régime en même temps que le *de* : *le palais était fermé, autour veillait une garde nombreuse*, (Ac. 1932).
 - avec les prépositions *à* et *de*, l'omission du régime est impossible parce que les pronoms conjoints *y* et *en* remplacent ce type de syntagmes : *parti*

¹ Ce tableau est extrait de P. LE GOFFIC : *Grammaire de la phrase française*, Hachette Supérieur, Paris, 1993, p. 423.

pour le Québec en janvier, il y est resté plus de deux mois et il en est revenu à Pâques,

- * il s'agit d'un nom de personne : le tour normal est de remplacer le nom par un pronom personnel, soit conjoint (préposition omise), soit disjoint selon les prépositions.
 - pour *derrière, dessus...*, cela est assez courant : *il court derrière pour la rattraper* (Dict. contemp.)

L'emploi des prépositions sans régime est limité, il est courant avec certaines d'entre elles, mais en général avec un caractère familier. On peut discuter pour savoir s'il s'agit d'adverbes ou de prépositions à complément zéro (avec ellipse de discours), mais en règle générale, il y a toujours un complément latent.

Certaines prépositions se prêtent à une « réadverbialisation » (*pour* et *contre*, par exemple), elles sont moins abstraites et donc moins polyvalentes, que *à* et *de*, dont l'emploi adverbial n'est guère possible. « On peut supposer que c'est l'une des raisons pour lesquelles leur charge notionnelle est sentie suffisante pour qu'on puisse lui faire jouer le rôle d'apport - c'est ce qui se passe dans leur emploi adverbial. Pour les prépositions *dans, sur, sous*, non employées comme adverbes, un autre facteur intervient : l'existence d'une série parallèle *dedans, dessus, dessous*, mots dans lesquels, selon G. Guillaume (*Leçon* du 12 avril 1956, p. 8), l'élément *de-*, exerçant son rôle « d'inverseur », « d'agent de mutation », a renversé la préposition du champ de la transprédicativité au champ de la prédicativité ; en fonction d'adverbe, ce sont ces termes que l'on emploie. »

2.2.2.2. En allemand

Dans la plupart des grammaires on fait la différence entre prépositions et adverbes par le fait que les prépositions, contrairement aux adverbes, sont souvent suivies d'un régime (*auf und ab* = adverbe, *er wartet auf mich* = préposition). Comme beaucoup de prépositions sont dérivées d'adverbes, on peut considérer comme tout à fait sensé que certaines prépositions puissent apparaître avec ou sans régime. *Abseits* et *eingangs* peuvent ou non être suivis d'un régime. Selon la terminologie grammaticale allemande, ils seront utilisés soit comme une préposition introduisant une spécification adverbiale (*er steht abseits der Autobahn*), soit comme un adverbe (*er steht abseits*). Attention *ausgangs* ne peut pas être utilisé sans régime. De même, on pourra omettre le régime, lorsqu'il est sous-entendu, mais identifiable à coup sûr : *er setzte den Hut auf (dem Kopf)*.

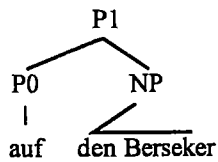
Du point de vue de la syntaxe, on peut faire une différence entre les groupes prépositionnels simples et les groupes prépositionnels complexes : ¹

a) ich warte [auf den Berserker]

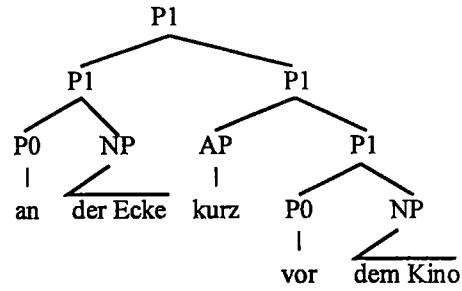
b) ich warte [an der Ecke kurz vor dem Kino]

¹ Ces graphiques ont été repris dans N. FRIES : *Präpositionen und Präpositionelphrasen im Deutschen und im Neugriechischen*, Max Niemeyer Verlag, 1988, p. 91.

Théorie \bar{X}



GP simple



GP complexe

- * la substitution du groupe P1 se fait dans des conditions différentes entre a) et b) :
 - a) darauf / worauf warte ich
 - b) *daran / davor warte ich
- * on peut construire le groupe P1 par un lexème de spécification PostP dans les phrases de type b) mais pas dans celles de type a) :
 - er wartet [den Sommer über]
 - er sucht sich [nach Callgirls] mais * er sucht sich [Callgirls nach]
- * dans les groupes prépositionnels de type a), il ne se présente qu'un nombre très limité de lexèmes P, dans les groupes prépositionnels de type b) se présentent tous les lexèmes P possibles. Il s'agit des lexèmes suivant : *an, auf, aus, bei, durch, für, gegen, in, mit, nach, über, um, unter, von, vor, wider, zu, zwischen* (*außer* seulement dans les constructions à verbe support^o).
- * les groupes prépositionnels de type b) peuvent être construits par des lexèmes P intransitifs (non pronominaux), mais pas les GP de type a) : *Ich warte [oben]*

Les deux types de groupes prépositionnels présentent une distribution syntaxique différente :

- les groupes prépositionnels simples ont une position déterminée soit par un lexème de spécification lexicale, soit par la structure syntaxique,
- les groupes prépositionnels complexes apparaissent soit de façon relativement libre dans n'importe quelle phrase, soit comme complément de différents lexèmes.

2.2.3. Sémantique

Au niveau sémantique, la localisation spatiale est représentée par une relation statique entre deux termes, l'un étant l'objet à localiser et l'autre un repère par rapport auquel se précise la localisation.

Selon les auteurs, l'objet et le repère ont différents noms : Herskovits (82), Langacker (86) les appellent *figure* et *ground*, Jackendoff (83) *theme* et *reference object*, Talmy (83) *primary object* et *secondary object*, Vandeloise (86) *cible* et *site*.

Dans leur sens prototypique, la plupart des prépositions sont des prédicats spatiaux et, en tant que tels, elles peuvent converger sur différents aspects de notre environnement physique.

En conceptualisant le monde physique et en parlant, nous sommes confrontés à un très grand nombre d'entités ayant une signification spatiale, qui ont une extension spécifique dans l'espace tridimensionnel selon leur tridimensionnalité, leurs liens internes et externes, leur forme, leur orientation et leur mobilité. C'est grâce à ces propriétés spatiales que les entités construisent, entre elles, les relations possibles et ainsi les prédictions spatiales possibles exprimées par les prépositions.

Döpke et Schwarze (1981) font une distinction entre :

- * la localisation actantielle : les régularités selon lesquelles une indication de lieu est appliquée aux différents objets désignés par une phrase reposent sur une interaction complexe entre l'information sémantique et l'information extralinguistique : savoir, croyances, probabilité... : *Paul travaille à la cave*, et
- * la localisation propositionnelle : dans certaines phrases, l'indication de lieu joue le rôle d'une espèce de cadre ou de condition par rapport à la description de procès généralisés : *en Angleterre, on roule à gauche*.

Attention dans la localisation actantielle, on n'exclut pas que l'indication de lieu puisse signaler l'endroit où se trouve(nt) le (les) actant(s). On peut, en principe, englober la localisation actantielle dans la localisation propositionnelle en dérivant l'information relative à la localisation des actants de l'information que le syntagme prépositionnel fournit au sujet de la localisation de l'état de choses considéré comme un tout. Cette solution ne convient pas à Döpke et Schwarze car elle ne permet pas, comme l'a remarqué R. Bartsch, d'analyser de façon unitaire les indications de lieu :

Paul téléphone à Mannheim → il téléphone à quelqu'un qui se trouve à Mannheim

Paul regarde dans la boîte aux lettres → il peut ou non y avoir quelque chose dans la boîte aux lettres.

Döpke et Schwarze proposent une première solution (qui sera modifiée par la suite) qui permet un traitement unitaire des différents types d'indication de lieu. Ils partent du fait qu'une préposition exprime premièrement un cas locatif et deuxièmement une relation de localisation spécifique. En utilisant la préposition pour exprimer cette relation spécifique, ils proposent de formuler la relation locative générale ainsi :

« il existe un x qui participe à l'état de choses e en tant qu'actant et pour lequel est valable la formule ' x préposition y '. Quant au rôle que x joue dans la phrase, il dépend des données sémantiques et pragmatiques de chaque cas particulier ». ¹

Ils présentent ensuite une deuxième solution afin d'analyser des cas du type *dans la boîte aux lettres*.

« Au lieu de dire que l'information spécifique fournie par la préposition est une relation entre deux individus, on peut dire qu'il s'agit d'une fonction, qui, dans le cas des prépositions locales, attribue à l'objet localisant un segment d'espace ».

¹ C. SCHWARZE : *Analyse des prépositions*, IIIème Colloque franco-allemand de linguistique théorique du 2 au 4 février 1981 à Constance, Max Niemeyer Verlag, Tübingen, 1981, p. 24.

Celui-ci participe à l'état de choses décrit par la phrase, dans un rôle déterminé, qui n'est pas forcément le locatif. La relation locative est redéfinie comme une relation entre un état de choses e et un segment d'espace s :

« il existe un x qui participe à e et qui se trouve à s ». ¹

Il semble y avoir trois problèmes de base pour la sémantique des prépositions :

- 1) Les prépositions offrent une variabilité considérable d'interprétations dépendantes du contexte : une préposition peut exprimer une large étendue de relations (ceci reste vrai même si on s'intéresse seulement aux utilisations spatiales d'une même préposition) ; ceci pose problème aux théories « abstractionnistes » qui partent de l'hypothèse selon laquelle il faut définir le sens des mots en terme de conditions nécessaires et suffisantes pour les utiliser de façon adéquate.
- 2) Le second problème vient du fait qu'il y a une interaction sémantique étroite entre la préposition et son contexte linguistique (si on pense à une préposition spatiale, on a à l'esprit un lieu ou une chose). On ne peut pas concevoir le sens d'une préposition sans référence à un argument interne ; une information configurationnelle concernant les arguments semble faire partie du sens des prépositions spatiales.

La sémantique des prépositions doit débarrasser et isoler le contenu sémantique spécifique de la préposition individuelle des facteurs contextuels qui influencent ses différentes interprétations et expliciter de quelle manière ces facteurs sont en interaction avec le sens des prépositions.

- 3) Le troisième problème émane du fait que l'on peut voir une même situation selon différentes perspectives, elle peut être conceptualisée de plusieurs façons. On est capable de créer des structures mentales différentes afin d'interpréter une scène extralinguistique, on peut utiliser plusieurs prépositions pour faire référence à une même situation linguistique (*quelque chose sur / dans l'assiette*). ²

On ne peut pas extrapoler les conditions d'utilisation des prépositions à partir des propriétés géométriques objectives des situations extralinguistiques qu'elles peuvent exprimer. Il faut rechercher les conditions et les principes cognitifs qui rendent possible la conceptualisation de certaines situations extralinguistiques d'une manière ou d'une autre et le choix de telle ou telle préposition.

Zelinsky-Wibbelt suit une distinction faite par Herskovits en 1986 et classe les facteurs qui influencent le sens d'une préposition selon deux dimensions :

- * les contraintes directes / inhérentes, indépendantes du contexte spécifique d'un discours ; elles incluent les restrictions entre le prédicat prépositionnel et ses arguments de même que les relations spatiales possibles entre les entités reliées par la préposition,
- * les contraintes dépendantes du contexte ; elles font référence aux conditions spécifiques fournies par la situation de discours sous-jacente à une expression.

¹ Cf. *op.cit.* [Schwarze81], p. 25.

² Cf. *op.cit.* [Zelinsky-Wibbelt93], pp. 179-180.

Lang, partisan de l'approche à deux niveaux, pense que l'utilisation intrinsèque ou extrinsèque des prépositions projectives ne dépend pas de leurs différentes représentations sémantiques. L'interprétation de différents concepts est le résultat d'une interaction entre deux niveaux distincts représentant l'information sémantique et l'information conceptuelle, au cours de laquelle, on assigne des valeurs aux paramètres qui font partie de la représentation sémantique. Au niveau conceptuel, les objets spatiaux sont représentés comme des schémas d'objets comportant des informations sur la dimensionnalité, la forme, la taille et l'orientation de l'objet. La position et la localisation de l'objet dans l'espace sont déterminées par l'espace perceptuel primaire (*Primary Perceptual Space*) qui comprend les trois axes suivants : l'axe vertical, l'axe de l'observateur et l'axe horizontal. On peut représenter la sémantique d'une préposition grâce à deux composantes de la signification : l'inclusion qui est responsable de l'interaction des propriétés de dimensionnalité de l'objet situé et de l'objet de référence, et le chemin qui contient le paramètre de passage $D(x)$ établi par le chemin effectué à partir des informations conceptuelles sur l'objet situé x .

Selon Pribbenow, les expressions de localisation sont définies par les fonctions leur permettant d'affirmer la localisation statique ou dynamique des entités, cette fonction étant typiquement désignée par les prépositions. Hottenroth, quant à elle, affirme que la représentation sémantique de la préposition doit être une abstraction d'une certaine gamme d'utilisations « prototypiques », et non de toutes les interprétations contextuelles possibles. Cuyckens pense que la signification est encyclopédique par nature et explique les différents sens d'une préposition comme une structure de ressemblance de famille et le flou de sens de certaines prépositions comme la conséquence naturelle du fonctionnement de l'homme dans son environnement social.

En grammaire cognitive les prépositions considérées comme expressions spatiales sont représentatives de la nature de la signification linguistique : la conceptualisation est basée sur l'expérience physique issue de notre fonctionnement dans un environnement spatial, c'est ce que Lakoff appelle l'expérience préconceptuelle. Elle est encodée mentalement sous forme de schémas imagés qui servent de modèles cognitifs.

On considère que les prépositions expriment trois types de relation : la place, le but et le passage. On peut éventuellement en rajouter une quatrième : la source. Ces relations constituent ce que Johnson (1987) a appelé « les schémas d'images »¹. Il affirme que c'est l'abstraction et la généralité de ces schémas qui nous permettent de structurer indéfiniment de nombreuses perceptions, images et événements. Les objets n'occupent pas seulement une place, ils bougent également le long d'une trajectoire, à partir d'un point de départ et vers un point d'arrivée (ces trois notions se présupposent mutuellement). Toutefois, dans un sens de passage, une préposition dessine seulement la trajectoire sans tenir compte du point de départ et du point d'arrivée ; dans un sens de but, elle s'occupe du point d'arrivée mais ne tient pas compte de la trajectoire ni du point de départ.

Les prépositions spatiales indiquent la relation *spatiale* entre deux arguments x et y , c'est-à-dire, la façon dont x et y sont en relation l'un par rapport à l'autre dans l'espace.

¹ Exemples cycliques de l'expérience impliquant « le mouvement du corps humain, la manipulation d'objets et les interactions perceptuelles ».

Les prépositions spatiales décrivent la place / la localisation de x (en d'autres mots, elles assignent une place particulière à x), en utilisant l'argument y comme référence ; elles servent à localiser x par rapport à y , ou plutôt, la place de y . Dans de nombreux cas, la préposition spatiale ne sert pas à localiser l'entité entière x par rapport à y , mais un élément dans sa structure sémantique.

Les langues ordinaires s'occupent de l'espace relativiste, espace dans lequel la localisation des objets est déterminée par rapport à d'autres objets. Notre conception tridimensionnelle de l'espace peut être attribuée à notre appréciation visuelle de la profondeur et des relations verticales et horizontales entre les objets. Toutes les prépositions ne parlent pas de l'espace par rapport à une origine et des axes coordonnés. Certaines d'entre elles peuvent être interprétées par rapport au locuteur ou à l'observateur.

En règle générale, les prépositions qui dénotent une position du trajector peuvent aussi dénoter un but, c'est-à-dire une place que le trajector va occuper par rapport au landmark. On peut voir ce but sans tenir compte de la décomposition de la place (un point, une configuration de points uni-, bi- ou tridimensionnelle).

Il y a deux modes d'interprétation de la préposition spatiale :

- * déictique : la relation spatiale désignée par la préposition est liée à l'orientation visuelle d'un observateur, qui sert comme point de référence pour la localisation des objets,
- * inhérente : l'orientation ne joue aucun rôle, (la face de) l'objet localisant est le point de référence (attention la face d'une armoire correspond aux portes, alors que la face d'une voiture correspond au volant ; la face des objets dépend de conditions différentes).

Le référent spatial peut ou non se situer dans le champ visuel de l'observateur.

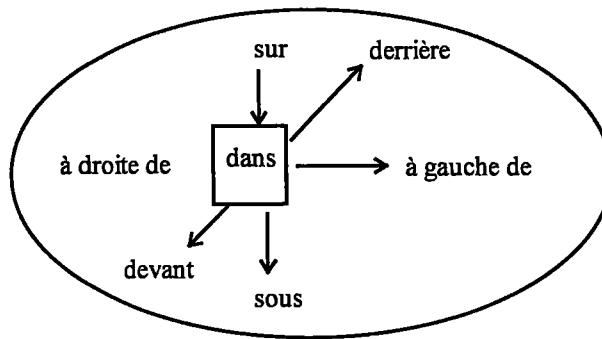
Les prépositions non-déictiques ne mettent en relation que deux éléments : un point de référence, et un élément situé par rapport à celui-ci, alors que les prépositions déictiques, mettent en relation trois éléments : un point de référence, un élément situé par rapport à celui-ci et le locuteur-observateur (c'est par rapport à sa position que sont définis, dans chaque situation, le « devant » et le « derrière » de l'objet de référence).¹

Il faut tenir compte des propriétés spatiales des objets localisés et des objets de référence ainsi que des principes de conceptualisation et de catégorisation des objets pertinents pour l'utilisation d'une préposition individuelle et de son interprétation.

Herweg propose la représentation sémantique suivante : $\lambda y \lambda x$ (LOK (x , PREP* (y))). Cette hypothèse est basée sur l'observation de prépositions telles que *dans*, *sur*, *sous*, *devant*, *derrière*... qui se différencient les unes des autres par le fait que chacune d'entre elles met en relation l'objet localisé dans une région caractéristique différente par rapport à l'objet de référence. Les prépositions doivent contenir les informations qui déterminent le type de région qu'elles peuvent identifier par rapport aux différents

¹ Cf. *op. cit.* [Schwarze81], p. 258.

objets de référence possibles. Les régions sélectionnées par les prépositions pour un objet tridimensionnel sont : ¹



Hottenroth suit une proposition faite par Wunderlich dans le sens où elle suppose que la signification d'une préposition spatiale est une relation de localisation entre l'objet localisé x et une région caractéristique par rapport à un objet de référence y .

On peut dire que les prépositions spatiales ont une composante commune : elles indiquent qu'un objet x est localisé dans une région définie par rapport à un objet y . LOC est la relation de localisation entre un individu et une région. REG (y) est une métavariable utilisée pour définir la région voisine spécifique de l'individu y : $\lambda y \lambda x$ (LOC (x , REG (y))). La forme sémantique des prépositions spatiales est représentée par une structure sémantiquement décomposée en une composante de localisation et une autre composante qui impose des contraintes supplémentaires. ²

2.2.4. Pragmatique

« Lorsque la pragmatique indexicale a des applications possibles en sémantique prépositionnelle, ce ne sont pas des particularités individuelles internes, propres à tel locuteur ou à tel allocutaire, qu'elle vise, mais une situation spatiale ou temporelle où pourrait se trouver n'importe qui (cf. ce qu'implique un énoncé comme *La chaise est derrière la table* concernant la position dans l'espace des interlocuteurs). » ³

Dans certains cas, l'attribution d'un sens précis au rapport prépositionnel nécessite une connaissance du contexte et/ou de la situation, en raison du caractère insolite de la mise en rapport, de la « polysémie » de la préposition et de l'un ou l'autre des termes qu'elle relie.

La production et la compréhension d'un rapport prépositionnel mettent en jeu des connaissances relatives non plus aux coordonnées spatio-temporelles de l'énonciation et à la situation des actants de l'énonciation dans l'espace mais aux êtres, notions, objets, phénomènes et circonstances dont il est parlé.

Une phrase telle que *Il s'est suicidé au café* est ambiguë car la préposition est polysémique : *au café* peut signifier 1) dans un café, 2) au moment du café après le dessert, 3) après avoir bu trop de café. On pourrait conjuguer 1) et 2), il est mort dans un

¹ Le schéma est extrait de *op. cit.* [Zelinsky93], p. 183.

² *Ibid.*, pp. 221-222.

³ J. CERVONI : *La préposition. Étude sémantique et pragmatique*, Duculot, Paris / Louvain-la-Neuve, 1991, p. 239.

café au moment du café, et on pourrait même conjuguer les trois : il est mort au café, au moment du café et avec du café. Dans le cas 3), *il s'est suicidé* devient ambigu : le café peut être le moyen ou la cause de la mort ; dans le cas de la cause, on peut se demander s'il est mort à la suite d'un excès ou de plusieurs excès répétés pendant une période ; pour désambiguïser *il s'est suicidé*, il faut faire appel à des connaissances spécifiques.

En plus des connaissances spécifiques, on fait aussi appel à des connaissances encyclopédiques, connaissances liées aux différentes acceptions du mot café :

- * boisson obtenue à partir des grains de caféier,
- * lieu public,
- * moment, où dans un repas, on sert le café – ces connaissances sont telles que :
 - a) liées à 1), on pense à des notions para-médicales sur les effets nocifs du café ;
 - b) liées à 2) on fait appel à la connaissance d'une sorte de rituel en vertu duquel le dessert vient après le fromage et le café après le dessert,

ces connaissances peuvent influencer sur l'ambiguïté : si un auditeur ne possède pas le savoir a) ou le savoir b) l'énoncé est moins ambigu puisqu'une possibilité d'interprétation disparaît et s'il ne possède ni l'un ni l'autre, l'énoncé cesse d'être ambigu.¹

Searle (1979), quant à lui, part de l'exemple *Le chat est sur le paillason*.² L'applicabilité de *sur* est déterminée, entre autres, par les éléments suivants :

- * le chat et le paillason sont situés dans un champ de gravitation à l'intérieur duquel être sur et être sous ont une signification précise (dans un espace sidéral où la gravitation est nulle, cette signification serait perdue),
- * le chat est un vrai chat et il repose de tout son poids sur son support,
- * le paillason est à plat sur le sol.

Cette même phrase pourrait être utilisée pour les cas bizarres imaginés par Searle (chat et paillason flottant dans l'espace, chat suspendu par des fils, paillason formant un angle avec le sol) sans cesser d'être vraie au sens littéral. Les hypothèses préalables seraient différentes de celles qui entrent en jeu dans la « situation standard » ; selon Cervoni :

« cela prouve, pour l'auteur, que le sens littéral n'est pas indépendant du contexte et, par voie de conséquence, « qu'il n'existe pas de distinction nette entre la compétence linguistique du locuteur et sa connaissance du monde ».

2.2.5. Les verbes et la localisation

Quand on analyse les différents membres d'un GP, on remarque que le contenu sémantique d'une préposition est souvent déterminé par le contenu des éléments nominaux situés à sa droite. L'exploitation des traits morphosyntaxiques dans le GP ne donne pas de clarification satisfaisante du contenu de la préposition, c'est pourquoi on

¹ Cf. *op. cit.* [Cervoni91], pp. 245-246.

² *Ibid.*, pp. 243-244

s'appuie plus sur une information sémantique du GN qui la suit. Toutefois le rôle des classes nominales en combinaison avec l'information morphosyntaxique est limité car le français est une langue non flexionnelle. Par exemple, dans les exemples tels que *Le livre est sur la table*, *Il pose le livre sur la table*, si on analyse le nom situé près la préposition, on a une indication de lieu mais on n'a pas d'indication de mouvements ou de direction, il faut donc élargir le contexte.

Les verbes jouent alors un grand rôle : les classes verbales précisent et déterminent une valeur sémantique exacte :

- * verbes de mouvement dirigé : *tomber, monter...*
 - impliquant un point de départ : *venir, s'enfuir de* + GN
 - impliquant une arrivée : *aller, arriver à / dans* + GN
- * verbes de mouvement non dirigé : *marcher, se promener, flâner...*
- * verbes duratifs : *sommeiller, dormir ...*
- * verbes inchoactifs : *se mettre en route ...*
- * verbes perfectifs : *terminer, finir, trouver...*

Différents types de verbes sont utilisés dans la localisation spatiale :

- * les verbes simples : *être, demeurer, siéger...*
- * les verbes pronominaux dérivés de verbes transitifs : *s'élever, s'étendre, se dresser...*
- * les formes du type : être + participe passé : *être posé, être rangé...*

On peut regrouper les verbes en fonction de leur trait dominant de caractérisation, il semble en effet assez difficile de les regrouper en fonction d'une décomposition en termes de traits (certains aspects se combinent entre eux dans le verbe) :

- * nature : non animé/humain ou animal : *être placé, habiter, reposer, figurer, régner...*
- * nombre : *être éparpillé, s'empiler, grouiller, se presser, se répartir...*
- * dimension horizontale, verticale, volume : *s'étendre, s'élever, se dresser, reposer, pendre, contenir (des)...*
- * persistance dans l'état : *rester, demeurer, stagner, siéger, planer...*
- * posture : *être debout, être à plat, être posé, être appuyé...*
- * consistance (de la cible ou du site) : *baigner (dans), croupir, flotter, nager, stagner...*

Certains verbes de mouvement ou d'action peuvent se transformer en verbes d'état : *des ballons s'élevaient lentement dans les airs / une statue s'élevait au centre du square* (cette transformation est liée à l'emploi de l'imparfait ou du présent qui interdit la mention d'indicateurs aspectuels pour indiquer un déroulement, un début ou une fin d'action). Il en est de même pour certaines phrases passives : *la vaisselle est immédiatement empilée dans l'armoire / la vaisselle est empilée dans l'armoire.*

Vandeloise affirme :

« les verbes de déplacement sont en fait une classe de verbes de mouvement particulière, pour lesquels le système de référence par rapport auquel le mouvement est évalué doit être extrinsèque à l'entité mobile. Il y a déplacement de l'objet entier s'il occupe successivement plusieurs positions de l'espace ». ¹

Il y a différents types de verbes de mouvement :

- * verbes comme *toucher*, *poser* : leur usage ne nécessite pas le déplacement de l'entité entière mais seulement celui d'une de ses parties,
- * verbes comme *marcher*, *courir*, *ramper* : leur usage est lié à un déplacement qui reste en arrière plan, leur fonction essentielle étant de spécifier la manière dont le mouvement s'effectue,
- * verbes comme *aller*, *venir*, *arriver* : ils décrivent un déplacement évalué par rapport à un système de référence extérieur à la cible mobile.

Le verbe *aller* est inutilisable si la nature du déplacement n'est pas précisée ; avec le verbe *venir*, la nature est spécifiée implicitement grâce au caractère déictique du verbe et correspond normalement au lieu d'énonciation : * *Paul va*, *Paul vient / arrive / entre / part*. ² Dans l'immense majorité des cas, le verbe de mouvement appelle la mention de la direction et de l'origine du mouvement, apparemment sans que l'on puisse les supprimer. Avec ces types de verbes, il n'est pas facile d'isoler la contribution d'une préposition au sens de la phrase.

« Le sémantisme verbal intervient dans le système des fonctions grammaticales portées par les cas et par les prépositions comme un élément perturbateur » ³ : dans une phrase telle que *je vais à Rome*, le complément de lieu est obligatoire, alors que dans l'exemple suivant : *il est mort de la fièvre à Rome*, il n'est pas obligatoire.

2.2.6. Les différentes prépositions spatiales en français

Nous scinderons cette étude sur les différentes prépositions spatiales en français en trois grandes parties : la première concernera les prépositions *à*, *de* et *en*, souvent traitées à part par les linguistes, car elles sont généralement considérées comme « prépositions vides », la seconde s'intéressera aux prépositions que l'on peut souvent regrouper en « paires prépositionnelles », et la troisième traitera des autres prépositions.

¹ C. VANDELOISE : « L'expression du mouvement », in *Langue Française* 76, Larousse, Paris, 1988, p. 85.

² *Ibid*, p. 88.

³ F. LETOUBLON : « Sur les genoux des Dieux », in *Sur les compléments circonstanciels*, Langue Française 86, Larousse, Paris, p. 25.

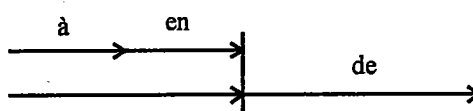
2.2.6.1. Les prépositions *à*, *de* et *en*

2.2.6.1.1. La notion de «prépositions vides»

Pour certains linguistes, il existe deux types de prépositions : les prépositions « vides » (*à*, *de* et parfois *en*) et les autres. C. de Boer, quant à lui, les classe en trois groupes : les prépositions casuelles, les prépositions semi-casuelles, les prépositions non-casuelles, que F. Brunot et Ch. Bruneau appellent respectivement prépositions vides, demi-vides et pleines.¹

Il existe deux raisons essentielles pour lesquelles certains linguistes séparent *à*, *de* et parfois *en* des autres prépositions :

- * leur indice de fréquence est très élevé (elles apparaissent dans un groupe où prédominent des mots typiquement grammaticaux),
- * elles sont monophonémiques² (comme des mots typiquement grammaticaux tels que *le*, *que*, *où*) G. Moignet tire parti de cette caractéristique, il établit le micro-système suivant :



« Les deux éléments vocaliques – le second n'est que la variante nasalisée du premier – sont placés en première tension, l'élément consonantique en seconde tension. Cette structure phonologique élémentaire est en rapport avec la ténuité, la généralité et le degré abstrait des sémantèses en cause ».

Il est ainsi amené à distinguer « dans l'ensemble des prépositions, une zone de sémantèse transcendante, de ténuité maximale, représentée par les prépositions monophonémiques *à*, *en* et *de*, et une zone de sémantèse immanente, plus particulière, de densité sémantique moyenne, où figurent *dans*, *vers*, *par*, *pour*, *sur*, *sous*, *chez*, *entre*, *avec*, *sans*, prépositions di- ou triphonémiques (*avec* étant la seule ayant quatre phonèmes) » (*op. cit.*, p. 230, § 385).

G. Moignet ajoute que c'est seulement avec des « termes plus étoffés qui fonctionnent aussi (...) comme adverbes : *avant*, *après*, (...) ou qui sont des mots prédicatifs : adjectifs, participes, substantifs en emploi prépositionnel : *sauf*, *excepté* (...) *malgré* » et plus encore, avec les locutions prépositionnelles que l'on pénètre dans un domaine de relations plus particulières, plus précises. L'impression de vide sémantique est renforcée si on les compare avec des prépositions telles que *avec*, *sans*, *avant*, *après*..., ou si on s'intéresse aux cas où elles commutent soit avec zéro (*traiter d'un sujet / traiter un sujet*) soit entre elles (*commencer à / commencer de*), où la signification n'en est pas modifiée.

¹ Cf. *op. cit.* [Cervoni91], p. 138.

² Composé d'un seul phonème c'est-à-dire de la plus petite unité sonore ayant une signification fonctionnelle.

C. de Boer considère que les prépositions vides *à* et *de* ne sont utilisées que pour combler un « hiatus syntaxique », Grévisse estime qu'elles n'ont d'autre fonction que celle de « cheville syntaxique ».

Gougenheim a fait une étude sur le thème « Y a-t-il des prépositions vides en français ? » Il adopte une autre position que ses prédécesseurs en ce qui concerne la préposition *à* :

« En somme, dans tous les emplois de la préposition *à* en français moderne, nous retrouvons l'idée de la ponctualité statique ou dynamique. *à* ne saurait donc être considéré comme une prépositions vide ». ¹

En ce qui concerne la préposition *de*, il adopte une position moins nette : il distingue une série d'emplois où *de* est une préposition ayant une valeur intrinsèque :

« La « valeur intrinsèque » d'une préposition, « indépendante de la valeur fonctionnelle », c'est son « sens ». ²

et une série d'emplois où *de* peut exprimer une idée d'origine, d'éloignement, une dérivation de la cause vers le moyen, une partie, une matière..., *de* peut aussi apparaître dépourvu de valeur intrinsèque et même sortir du cadre des prépositions (indice d'infinitif, élément d'articles partitifs...) (*op. cit.*, p. 16) ; il fait une distinction entre les emplois prépositionnels et non prépositionnels de *de*. Selon lui, *de* peut être employé. Il conclut :

« C'est en ce sens que nous pouvons voir dans *de* la seule préposition «vide» du français ».

Spang-Hanssen, quant à lui, ne parle pas de prépositions vides mais de prépositions incolores.

« Parmi les étiquettes prépositions vides, prépositions abstraites, prépositions incolores, nous préférons la dernière pour son imprécision même. Elle ne prétend pas à une rigueur en discordance avec le principe d'analyse et c'est probablement celle qui a le plus de chance d'être acceptée par tous les grammairiens, puisqu'elle n'implique pas une seule manière de voir les problèmes » (Spang-Hanssen 1963, pp. 13-14). ³

Le terme incolore est emprunté à W. von Wartburg et P. Zumthor (*Précis de syntaxe du français contemporain*). Il estime donc que *à*, *en* et *de* ont un caractère « incolore » et « abstrait » et montre que le choix d'une de ces prépositions ou son remplacement par une préposition plus « concrète » obéit souvent à des facteurs non sémantiques tels que la variation (*l'admiration de mon frère*, *l'admiration de Voltaire* mais *l'admiration de mon frère* pour *Voltaire*), l'influence des articles (*le livre de mon père* mais *un livre à mon père*).

2.2.6.1.2. La préposition *à*

2.2.6.1.2.1. Généralités

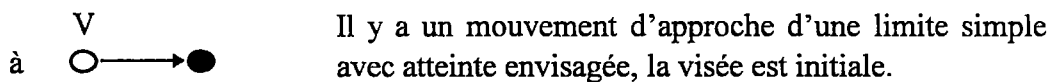
à sert à marquer le terme d'un mouvement dans l'espace, la destination : *aller à Paris*.

¹ Cité dans *op.cit.* [Cervoni91], p. 136.

² *Ibid.*, p. 137.

³ *Ibid.*, p. 132.

Le schème psychique proposé par Pottier, définissant le signifié de *à* est le suivant :



Mais, avec un verbe statique, *à* marque le locatif : *on m'attend à Paris*.

Selon Vandeloise, derrière l'auxiliaire *être* la préposition *à* est en règle générale :

- * acceptable quand le site est désigné par un nom propre (*Léopold est à Liège*)
- * acceptable quand l'article défini introduit un site dont la position est spécifiée (*Léopold est au Luxembourg* / * *Léopold est au rocher* mais *Léopold est au rocher de la Pierre Blanche*). * *Léopold est à l'arbre* est agrammatical alors que *Léopold est devant l'arbre* est tout à fait correct ; pour expliquer l'exactitude de l'énoncé *Léopold est à la fenêtre* / *à la plage*, Vandeloise considère que sa présence près de la fenêtre ou sur une plage évoque un rituel social auquel la cible participe.
- * inacceptable, à quelques exceptions près, quand le site est introduit par l'article indéfini (* *Léopold est à un arbre* / *Léopold est près d'un arbre*).¹

	article défini référentiel	article indéfini
<i>à</i>	souvent	rarement
<i>près, sur, dans</i>	toujours	toujours

Quand elle introduit un site spécifié, *à* a une fonction strictement localisatrice et est indifférente à la relation entre sa cible et son site. En revanche, *près de*, *sur* et *dans* établissent la relation d'accès, de porteur à porté et de contenant à contenu entre leur cible et leur site.

à est une préposition qui a été étudiée par Ruwet dans un article intitulé « A propos des prépositions de lieu en français ». Ruwet fait une étude comparative entre les compléments de lieu introduits par *à* (*le livre est à la bibliothèque*) et ceux introduits par une autre préposition (*le livre est sur la table, dans le tiroir...*). Leur identité en « structure de surface » est fallacieuse. La seule ressemblance que l'on peut trouver entre ces exemples est qu'ils répondent tous à la question « Où est le livre ? ». Si on effectue différents tests, on constate qu'il n'existe en fait que des différences : par exemple le test du comportement à l'égard de la coordination, on note qu'il est souvent possible de coordonner des prépositions de lieu qui ont un complément identique. Mais qu'en revanche, il est absolument impossible de coordonner *à* et toute autre préposition. Selon Ruwet, cette différence indique qu'en « structure sous-jacente », *à* et les autres prépositions de lieu ne doivent pas être placées dans la même catégorie. Il se pourrait que *à* ne soit pas une préposition mais fasse partie des « traits (analogues à des cas) appartenant aux syntagmes nominaux ». Il ne fournit pas d'arguments décisifs en faveur de cette hypothèse. A son avis, *à* est un « simple marqueur de subordination, qui ne contribue en rien à l'interprétation sémantique », alors que les prépositions de lieu y

¹ Cf. *op. cit.* [Vandeloise88], pp. 79-80.

contribuent. *à* est la « plus neutre » des prépositions de lieu. Elle apporte une information minimale par rapport aux prépositions de lieu (ex. : *Pierre est à la Sorbonne => Pierre est devant / derrière / à l'intérieur du bâtiment ?*). C'est la « valeur ponctuelle » de *à* qui permet cette imprécision : la préposition *à* possède une propriété (évidemment sémantique) grâce à laquelle on peut, en l'utilisant, parler d'un lieu, si étendu soit-il comme un point dans l'espace. Selon Vandeloise (1986, p. 14), il est préférable de dire que « la préposition *à* exprime la coïncidence de son sujet avec un point de son objet ». Il ne réduit pas l'objet à un point. Il vise à « substituer aux notions géométriques qui sont à l'arrière-plan, en général, de l'analyse linguistique des prépositions « spatiales » des concepts plus enracinés dans l'expérience d'eux-mêmes et du monde que peuvent avoir les usagers du langage ».

Dans les emplois où la préposition *à* est préférée à d'autres prépositions à valeur locative, elle sert, selon H. Frei (cf. *La grammaire des fautes*, p. 149), de « locatif générique ». Sa généralisation répond au « besoin d'invariabilité », qui est l'un des deux aspects sous lesquels se manifeste le « besoin d'économie » (l'autre aspect étant le « besoin de brièveté »). La définition du « besoin d'économie » précise certaines des conditions du recours à l'implicite :

« Pour le parleur et l'entendeur pressés, le jeu de la parole et de l'interprétation doit se dérouler aussi rapidement que possible. Le parleur abrège ou supprime plus ou moins inconsciemment tout ce qui dans une situation donnée va de soi, c'est-à-dire tout ce qui, étant connu de l'interlocuteur, forme le fond commun de leur conversation. L'entendeur, de son côté, au lieu de soumettre la parole de son interlocuteur à une analyse serrée, cherche à comprendre avec le minimum d'effort et de temps » (cf. *La grammaire des fautes*, p. 107).¹

2.2.6.1.2.2. Emplois spécifiques de la préposition *à*

La préposition *à* s'utilise devant² :

- * les noms de pays masculins, singuliers ou pluriels, et *à* initiale consonantique, elle se contracte avec l'article : *au Luxembourg, aux Pays-Bas*
- * les noms féminins ou masculins, commençant par une voyelle ou une consonne, de villes françaises ou étrangères : *à Paris, à Moscou, à La Rochelle*; *à* se contracte avec l'article si le nom de la ville est masculin et précédé d'un déterminant : *au Caire*. Pour les noms de ville commençant par A, certains auteurs utilisent *en*, au lieu de *à* : *en Avignon, en Arles ; en Alger* (rarement usité),
- * les noms de petites îles lointaines : *à la Réunion*,
- * les noms de petites îles d'Europe : *à Malte*,
- * les noms masculins de grandes îles lointaines : *à Madagascar*,³
- * les noms d'îles au pluriel, avec contraction de la préposition et de l'article : *aux Seychelles, aux Antilles*.

¹ Cf. *op. cit.* [Cervoni91], p. 265, note 125.

² M. GREVISSE : *Le Bon Usage*, Duculot, treizième édition refondue par André Goosse, Paris / Louvain-la-Neuve, 1993, p. 882.

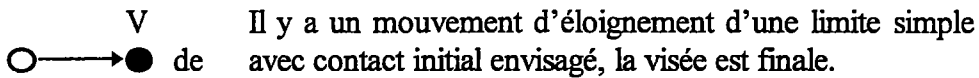
³ Précisons toutefois que *Le Petit Larousse* de 1997 indique que Madagascar est du genre féminin.

2.2.6.1.3. La préposition de

Avec un verbe dynamique, la préposition *de* marque le point de départ :

- * séparation extraction départ : *arracher un clou du mur*
- * provenance : *il vient de Paris.*

Le schème psychique, proposé par Pottier, définissant le signifié de *de*, est le suivant :



Avec un verbe statique, elle marque l'origine : *il est de Metz.*

La préposition *de* peut se combiner avec d'autres prépositions spatiales mais garde toujours son sens premier de point de départ, d'origine : *il vient de chez Paul.*

Tout comme la préposition *à*, *de* se contracte avec l'article masculin singulier : *il arrive du Caire*, et avec l'article pluriel : *il revient des Seychelles.*

2.2.6.1.4. La préposition en

2.2.6.1.4.1. Généralités

Dans son sens spatial, la préposition *en* marque la destination : *il va en Afrique*, la localisation : *il est en ville*, ou le but : *il est arrivé en France.*

J.-J. Franckel et D. Lebaud, dans une étude sur la préposition *en*, affirment :

« un indice essentiel du fonctionnement de *en* est d'emblée fourni par les caractéristiques du terme qu'il régit. [...] Cela tend à suggérer que *en* traite le terme qui le suit sur un mode purement qualitatif, et non comme une occurrence autonome ». ¹

On peut interpréter un exemple tel que *Il est en ville* ² grâce à deux propriétés : d'une part, l'énoncé fait référence à une localisation momentanée et non à une localisation permanente, et d'autre part le mot *ville* fait l'objet d'une forme de « centrage qualitatif », être en ville signifie être au centre, près du centre. Dans cet énoncé, il y a donc une actualisation temporelle de ce qui se présente comme un rapport de localisation spatiale, au niveau de l'interprétation, et une spécification de la localisation. *en* ne pourra pas s'employer avec tous les termes qui, à l'instar de *ville*, ont une valeur sémantique de localisateurs spatiaux. C'est ainsi que des phrases telles que *il est en village*, *il est en désert* ne sont pas acceptables. De même *il est en forêt* est correct alors que *il est en bois* n'a pas de valeur localisatrice, *bois* ayant ici une fonction de matériau. *Il est en campagne* confère à *campagne* la valeur d'une campagne militaire. Si on ajoute un marqueur de centrage tel que *plein* : *il est en plein désert*, *il est en pleine campagne* deviennent des énoncés acceptables.

Dans un exemple comme *être en mer*, on a les mêmes phénomènes d'actualisation (le poisson n'est pas plus *en mer* que la statue n'est *en ville*) et de centrage (perte de contact

¹ Cité dans [Berthonneau91], p. 56.

² *Ibid.*, pp. 58-59.

avec la rive). En revanche, on pourra dire d'un marin ou d'un bateau qu'il est *en mer* (mais pas d'une bouteille ou d'un animal marin) car ces termes sont intrinsèquement pourvus de valeur « maritime ».

Dans le premier exemple, on a utilisé des propriétés d'actualisation et de centrage, dans le second celles de fonction intrinsèque. Enfin, il en existe un troisième type, au niveau spatial, dans lequel on retrouve les propriétés d'actualisation et de fonction intrinsèque mais pas de centrage : *il est en classe* fait référence à l'instituteur ou à l'écopier, *il est en prison* au prisonnier, *il est en conférence* au conférencier...

J.-J. Franckel et D. Lebaud cite quelques auteurs qui ont étudié la préposition *en* :

« Ces deux prépositions [*en* et *dans*] nous mettent en présence d'un fait entièrement nouveau : la déformation de la fonction, au lieu d'être une variation d'une valeur que rien d'extérieur ne révèle, se dénonce, au contraire, par un changement de forme apparente. *en* est, en effet, dans la langue, la valeur déformée (...) de *dans* ». ¹

« La préposition *en*, issue de la préposition latine *in* est à l'origine une préposition de lieu qui a d'abord annulé le sens de *dans* et celui de *sur*, puis s'est peu à peu restreinte au premier avant d'être supplantée par *dans*, dans la plupart de ses emplois ». ²

2.2.6.1.4.2. Emplois spécifiques de la préposition *en*

La préposition *en* s'emploie, sans article, devant :

- * les noms de pays féminins : *en Chine*,
- * les noms de pays commençant par une voyelle, qu'ils soient masculins : *en Iran*, ou féminins : *en Allemagne*,
- * les noms féminins de grandes îles proches ou lointaines : *en Sardaigne*, *en Nouvelle-Guinée*, attention toutefois à : *à Terre-Neuve*, *à Java*,
- * certains noms de départements français, tels que les noms composés singuliers féminins : *en Haute-Vienne*, et avec certains noms simples, sans doute confondu avec les noms des anciennes provinces : *en Dordogne*. En revanche, *en* est impossible avec les noms composés singuliers masculins : **en Lot-et-Garonne*, les noms pluriels : **en Vosges*, **en Bouches du Rhône*, les noms masculins commençant par une consonne : **en Jura*. Grévisse affirme que les noms de départements se construisent plutôt avec la préposition *dans* et l'article (sauf *la Corse*, *la Savoie* et *Paris*),
- * les noms de provinces féminins : *en Normandie*, *en Bavière*,
- * les noms de provinces commençant par une voyelle : *en Anjou* ; dans ces deux cas, *dans* est possible,
- * les noms de provinces masculins et à initiale consonnantique : *en Poitou*, on utilisera volontiers *dans le* mais rarement *au*,
- * les noms de continents : *en Afrique*, *en Océanie*.

¹ GUILLAUME : *Le problème de l'article et sa solution dans la langue française*, pp. 265-266, 1919, cité dans *op. cit.* [Berthonneau91], p. 76.

² GOUGENHEIM : « Valeur fonctionnelle et valeur intrinsèque de la préposition *en* en français moderne », in *Grammaire et Psychologie*, Journal de psychologie, p. 178, Paris, 1950, cité dans *op. cit.* [Berthonneau91], p. 77.

2.2.6.1.5. Concurrence entre ces prépositions

2.2.6.1.5.1. Brøndal

Viggo Brøndal, dans ses *Essais de linguistique générale* (1943), dénombrait dix neuf prépositions et les expliquait par les concepts de logique. Selon lui, à et en étaient des corrélatifs, et désignaient respectivement un point et une ligne, le continu et le discontinu, le potentiel et l'actuel. Alors que de désignait tous les rapports à la fois : la direction et l'origine, le sujet et l'objet, la partie et le tout.

2.2.6.1.5.2. Mantchev

Mantchev vise une intégration de toutes les prépositions du français dans un système à trois niveaux :

I	à	en	de
II	avec	comme	sans
		malgré	
III	pour	par	contre

Le premier trinôme connaît « une différenciation interne » construite sur le même modèle, ce qui donne :

1.	à			en			de		
2.	devant	entre	derrière	sur	dans	sous	vers	chez	hors (de)
	avant	pendant	après						
	depuis	durant	jusque (à)						
3.	envers			selon			sauf		

TABLEAU 4 : TRINOME INSTITUTIONNEL DE MANTCHEV ¹

Mantchev établit un « trinôme institutionnel » à-en-de qui structure tout le système qui, réduit à sa partie stable se compose de huit trinômes. Il distingue un « système primaire stable », de 24 prépositions (19 chez Brøndal), et « un ensemble secondaire instable », constitué de « variantes prépositives et de prépositions complexes ».

En I, les prépositions ont une affinité avec être : les deux termes qu'elles relie, minimalement séparés, sont vus intériorisées dans une unité sémantique, rapportée à l'idée de être : *je suis à la maison = je suis chez moi ; il est en prison = il est prisonnier ; elle est de France = elle est française.*

En II, les prépositions ont une affinité avec avoir.

En III, les prépositions ont une affinité avec faire.

2.2.6.1.5.3. Emploi de l'article avec ces prépositions

Les noms de villes se construisent sans article sauf quand celui-ci fait partie intégrante du nom propre. Les noms de petites îles d'Europe et les noms masculins

¹ Le tableau a été repris dans *op. cit.* [Cervoni91], p. 160.

d'îles lointaines se construisent sans article : *Malte, Majorque, Chypre, Oléron, Jersey, Madagascar, Bornéo, Java, Ceylan, Formose ou Taiwan, Sakhaline, Cuba, Haïti* mais *la Corse, la Sardaigne, la Sicile, l'Irlande, l'Islande, la Nouvelle-Zélande...*¹

	Déterminants					Préposition			
	Ø	l'	le	les	la	à	au	aux	en
a) Lisbonne	+					+			
Malte	+					+			
La Rochelle	(+)					+			
Arles	+					+			+
b) Afghanistan		+							+
Allemagne		+							+
Islande		+							+
c) Groenland			+				+		
Pérou			+				+		
d) Etats-Unis				+				+	
Canaries				+				+	
e) Grèce					+				+
Sicile					+				+
f) Guadeloupe					+	+(la)			+
Jamaïque					+	+(la)			+
Martinique					+	+(la)			+

TABLEAU 5 : UTILISATION DES ARTICLES ET DES PREPOSITIONS EN FRANÇAIS²

On pourrait présenter le tableau précédent de la façon suivante :

Déterminants	Prépositions
la	en
l'	
Ø	
le	à
les	

En présence de la préposition *en*, le déterminant disparaît ; en revanche, avec la préposition *à*, il y a contraction de la préposition et du déterminant masculin ou pluriel.

2.2.6.1.5.4. Concurrence entre à et en

Doit-on dire *à la Guadeloupe* ou *en Guadeloupe* ? *à* possède le trait « ponctuel » (l'île est considérée comme un point), *en* possède le trait « surface » (on considère l'étendue de l'île)

« Concurrence *à/en* : La préposition *en* n'alterne avec *à* que devant certains substantifs féminins et quelques substantifs masculins à initiale vocalique qui ne sont pas incompatibles avec l'idée d'intériorité. L'article *livre* en combinaison avec *à* a une valeur de détermination, celle de généralité étant fournie par *en* qui exclut l'article ; on notera que la préposition *en* figure dans des syntagmes plus ou moins figés : *il travaille à l'usine / il travaille en usine ; à l'appartement / en appartement ; à la bourse / en bourse ; à la campagne / en campagne* » (*Trésor de la langue française*, p. 10).³

¹ Cf. op. cit. [Grévisse93], p. 883.

² Ce tableau est extrait de S. REBOUL : « À la Guadeloupe : en Guadeloupe », in *Le Lexique : construire l'interprétation*, Langue Française 103, pp.68-78, Larousse, Paris, p. 71.

³ Cf. op. cit. [Reboul90], p. 70.

Dans certaines locutions prépositives, on retrouve les prépositions *en* et *à* (ou plutôt *au*), mais l'usage n'a pas défini leur rôle de manière cohérente : *en bas de*, *au bas de*, *à bas de* ; *au-dedans de*, *en dedans de*, il faut toutefois faire attention à la différence d'orthographe. De même on peut trouver *remettre quelque chose en place ou à sa place*, mais on dit uniquement *aller au devant de quelqu'un*. Enfin les grammairiens opposent souvent *au-dessous de* et *en dessous de* : la première locution a davantage le sens de « sur la surface inférieure de », et la seconde de « dans la partie inférieure de ».

2.2.6.1.5.5. *à* et *en* les noms d'îles

Au cours de notre étude, nous avons pu constaté qu'il n'est pas facile d'établir des règles d'utilisation des prépositions *à* et *en* avec les noms d'îles car il existe de nombreuses exceptions. En effet, Grévisse affirme dans *Le Bon usage* (p. 1506, § 1003 a) 2°) que l'on utilise *à* devant les noms de petites îles lointaines (*à la Réunion*), les noms de petites îles d'Europe (*à Malte*), les noms masculins de grandes îles lointaines (*à Madagascar*), les noms d'îles au pluriel, avec contraction de la préposition et de l'article (*aux Seychelles*, *aux Antilles*) et *en* devant les noms féminins de grandes îles proches ou lointaines : (*en Sardaigne*, *en Nouvelle-Guinée* mais *à Terre-Neuve*, *à Java*).

Il semble, en effet, relativement difficile d'utiliser uniquement ces critères puisqu'il est quasiment impossible de définir avec exactitude le genre de certains noms d'îles. Nous n'avons pas trouvé d'ouvrage qui puisse nous donner ces indications avec précision. Certains dictionnaires unilingues donnent le genre de quelques noms propres (selon *Le Petit Larousse* de 1997, Cuba est du genre féminin, mais il n'y a aucun renseignement concernant Bornéo), il existe également des dictionnaires bilingues qui indiquent le genre de quelques noms propres.

Prenons un exemple : dans le dictionnaire *Harrap's Weis Mattutat* et dans *Le Petit Larousse*, il est écrit que Chypre est du genre féminin, dans le second ouvrage on apprend que l'île se situe en Europe et qu'elle mesure 9251 km². D'après une étude plus approfondie que nous avons menée, il semble qu'une île soit lointaine si elle ne se situe pas en Europe, et qu'elle soit considérée comme grande lorsque sa superficie dépasse 4000 km². On peut donc affirmer que Chypre est une grande île, proche et de genre féminin, donc d'après Grévisse, on devrait employer la préposition *en*, ce qui est faux puisque l'on dit *aller* ou *être à Chypre*.

En orientant notre étude vers l'emploi des déterminants avec les noms d'îles, nous en avons conclu que ce critère semblait le plus fiable et le plus généralisable. En effet, si on se réfère au tableau ¹ suivant, il apparaît que l'on utilise *à* quand le nom de l'île :

- * ne peut pas être précédé d'un déterminant : *à Chypre*,
- * est précédé d'un déterminant au pluriel : *aux Seychelles*,

¹ Le tableau a été établi ex nihilo. Nous avons cherché les noms d'îles dans le dictionnaire afin de déterminer leur localisation, leur superficie, et leur genre (quand cela était possible ; le point d'interrogation dans la colonne genre signifie que cette information n'apparaissait pas, le dictionnaire de référence est *Le Petit Larousse 1998*).

2. Les prépositions spatiales

* est précédé d'un déterminant masculin singulier : *au Japon*, bien que dans ce cas, le nom de l'île se confonde avec le nom du pays.

et *en* quand le nom de l'île est féminin et s'emploie avec le déterminant, il faut toutefois noter qu'avec la préposition *en*, le déterminant disparaît : *en Islande, en Nouvelle-Guinée...*

Nom	Genre	Article	Lieu	Taille ¹	Prép.
Antigua-et-Barbuda	f		Antilles	722	à
Bali	?		Pacifique	5.561	à
Belle-Ile	?		Europe	90	à
Bornéo	?		Pacifique	750.000	à
Bougainville	?		Pacifique	10.600	à
Chypre	f		Europe	9.251	à
Cuba	f		Antilles	111.000	à
Haïti	m		Antilles	27.750	à
Java	?		Pacifique	130.000	à
La Barbade	f		Antilles	431	à
La Dominique	f		Antilles	751	à
La Réunion	f		Océan Indien	2.511	à
Madagascar	f		Océan Indien	587.000	à
Madère	?		Europe	740	à
Majorque	?		Europe	3.640	à
Malte	f		Europe	246	à
Minorque	?		Europe	702	à
Panay	?		Pacifique	12.500	à
Porto-Rico	?		Antilles	8.897	à
Rhodes	?		Europe	1.400	à
Saint Barthélemy	?		Antilles	25	à
St Vincent et les Grenadines	m		Antilles	388	à
Sumatra	?		Pacifique	473.600	à
Taïwan	f		Japon	36.000	à
Terre Neuve	?		Atlantique Nord	406.000	à
Trinité et Tobago	f		Antilles	5.128	à
Les Açores	pl		Europe	2.247	aux
Les Antilles	pl		Antilles	240.000	aux
Les Baléares	pl		Europe	5.000	aux
Les Bermudes	pl		Antilles	53,5	aux
Les Canaries	pl		Europe	7.300	aux
Les Comores	pl		Océan Indien	1.900	aux
Les Kerguelen	pl		Océan Indien	7.000	aux
Les Maldives	pl		Océan Indien	300	aux
Les Philippines	pl		Pacifique	300.000	aux
Les Seychelles	pl		Océan Indien	410	aux
Groënland	m	le	Atlantique Nord	2.186.000	au
Japon	m	le	Japon	373.000	au
Royaume Uni	m	le	Europe	243.500	au
Sri Lanka	m	le	Océan Indien	66.000	au

¹ En km².

Guadeloupe ¹	f	la	Antilles	1.709	à la / en
Jamaïque	f	la	Antilles	11.425	à la / en
Martinique	f	la	Antilles	1.100	à la / en
Irlande	f	l'	Europe	70.000	en
Islande	f	l'	Atlantique Nord	103.000	en
Corse	f	la	Europe	8.680	en
Crète	f	la	Europe	8.336	en
Grande Bretagne	f	la	Europe	230.000	en
Nouvelle Bretagne	f	la	Pacifique	35.000	en
Nouvelle Calédonie	f	la	Pacifique	16.750	en
Nouvelle Guinée	f	la	Pacifique	800.000	en
Nouvelle Irlande	f	la	Pacifique	9.600	en
Nouvelle Zélande	f	la	Pacifique	270.000	en
Papouasie Nouvelle Guinée	f	la	Pacifique	463.000	en
République Dominicaine	f	la	Antilles	484.000	en
Sardaigne	f	la	Europe	24.090	en
Tasmanie	f	la	Pacifique	68.000	en

TABLEAU 6 : LES ILES : GENRE, DETERMINANT, LOCALISATION, SUPERFICIE

2.2.6.2. Les « paires prépositionnelles »

L'idée que les prépositions vont par paire correspond à une intuition. En effet, l'opposition entre *à* et *de*, *sur* et *sous*, *devant* et *derrière*... semble être évidente ; toutefois si on peut opposer *de* à *à* dans des emplois « spatiaux », « temporels » ou « notionnels », on peut aussi opposer *de* à *en* (matière => *un anneau d'or*, *une bague en or*), *à sur* (propos => *donner une explication de / sur*), *à pour* et *par* (cause => *louer de / pour ; par / de crainte de*).

En fait, on peut trouver un fondement théorique aux paires prépositionnelles en visant un très haut niveau d'abstraction. Chez G. Guillaume, le binarisme prépositionnel est en relation avec le mécanisme fondamental qui divise la totalité du pensable en un Univers-Espace et un Univers-Temps, puis, par répétitions successives de son propre fonctionnement, modèle le système des parties du discours sur cette partition première ». ²

« On peut et l'on doit (...) parler en l'occurrence de micro-systèmes, de binarités repérables dans le plan sémantique et mettant face à face des notions d'opérations contradictoires ou du moins opposables (...). On notera : *dans/hors* (inclusion / exclusion), *avant/après* (antéposition temporelle / postposition), *devant/derrière* (antéposition spatiale / postposition), *sur/sous* (superposition / sub-position), *avec/sans* (addition / soustraction), *par/pour* (transition / destination) ». Moignet (1981, p. 227)

Cervoni explique la pensée de Moignet ainsi :

« C'est bien en terme de positivation et de négativation que l'on peut décrire les oppositions les plus nettes du système des prépositions. Le contraste du positif et du négatif qui fonde le binarisme des paires prépositionnelles est un trait fondamental de la logique basiale issue de la « mécanique intuitionnelle ».

En fait les autres paires prépositionnelles sont des oppositions secondaires limitées à un domaine sémantique particulier. Ringenson (1926) explique l'alternance entre *par* et

¹ Pour la distinction entre la préposition *à* et *en*, se référer § 2.2.6.1.1.5.4. supra.

² Cf. *op. cit.* [Cervoni91], p. 153.

de dans les compléments d'agents par le sens des verbes dont ils dépendent ; Pottier (1974, p. 132) met l'accent sur l'influence de la classe puissant (*par*) ou non-puissant (*de*) du substantif complément. Les oppositions de ce type ne doivent donc pas être mises sur le même plan que les binarités résultant de la structuration du système prépositionnel par une double tension – positivante ou négativante – dont la source est la mécanique intuitionnelle.

2.2.6.2.1. avant / après

Selon Vandeloise, il y a trois domaines d'emploi pour ces prépositions : l'espace, le temps et les ordres abstraits (nombres, alphabet et « compétitions ») ; comme ils ont la même structure, toute l'information nécessaire pour traiter les différents emplois est fournie par le sens de *avant / après*, ce qui est un point crucial dans son analyse.

Ce sont des prépositions antonymes à la fois spatiales et temporelles. Les emplois nominaux des prépositions sont en principe communs aux deux domaines. Le site, repère à partir duquel on localise la cible, de *avant* peut toujours devenir cible de *après* et réciproquement. Le mouvement, élément indépendant du site et lexicalement inexprimé, qui joue un rôle capital dans la localisation de la cible a un rôle très important dans la relation *avant / après*.¹ En effet, il a un rôle dans l'espace (mais pas dans le temps et les ordres abstraits) à condition que les termes ordonnés soient mobiles.

Situation 1 : ordres dynamiques (fig. 1 p. 159)

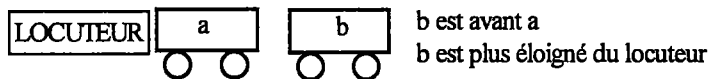


FIGURE 1 : AVANT / APRES AVEC LES ORDRES DYNAMIQUES

Situation 2 : ordres statiques (fig. 1 p. 159).

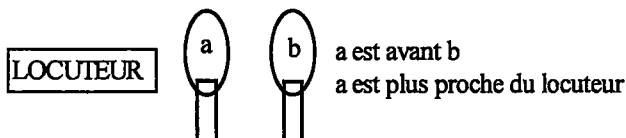


FIGURE 2 : AVANT / APRES AVEC LES ORDRES STATIQUES

La place du locuteur est pertinente pour les objets statiques-uniquement, et la notion de pôle est donc cruciale : c'est sur le déplacement possible du locuteur vers a, b - leur « rencontre potentielle » - que repose l'assimilation des ordres statiques aux ordres dynamiques, et donc la généralisation du mouvement.

L'ordre a un rôle très important : soit les référents sont intrinsèquement ordonnés (*Mars est avant Juin, 1 est avant 3...*), soit ils ne le sont pas et c'est *avant / après* qui en instaurent un.² Vandeloise illustre la « rencontre potentielle » en prenant l'exemple d'une voiture jaune et d'une voiture bleue situées sur des routes perpendiculaires qui se croisent à une intersection : on dira « *la voiture jaune est avant la voiture bleue* » si elle

¹ Cf. *op. cit.* [Berthonneau93], pp. 44-45.

² *Ibid.*, pp. 46-48.

est plus près du carrefour que la voiture bleue. Si l'itinéraire des deux voitures ne doit pas se croiser, on ne peut pas utiliser *avant* / *après*. Toutefois si les véhicules se suivent, on ne voit pas comment ils pourraient se rencontrer.

Introduit grâce aux ordres dynamiques, c'est en réalité pour les ordres statiques auxquels il s'applique en vue de généraliser le mouvement que le terme de « rencontre potentielle » est le plus approprié, le locuteur, s'il se déplace, atteignant d'abord *a*, puis *b*.¹ Les pôles sont définis par les échelles spatiales. Il existe évidemment un pôle terminal (but du mouvement partagé des deux mobiles) pour les ordres dynamiques. Pour qu'il y ait une rencontre potentielle dans les ordres statiques, il faut que le locuteur soit dit aligné avec *a*, *b*, quelle que soit la position de *a*, *b* devant ou derrière lui.

Il n'y a pas d'ordre préconstruit dans l'espace, on a donc besoin d'un troisième terme pour définir l'axe selon lequel *a*, *b* sont alignés ; ce manque est pallié par la rencontre potentielle qui instaure un pôle mais qui impose une contrainte plus forte, en identifiant le locuteur au pôle. Un site inexprimé ne peut référer qu'à une position « virtuelle » du locuteur ; *avant* / *après* n'ont jamais le locuteur pour site inexprimé, ils ont un usage identique pour le chemin et la ligne du regard dans l'espace visuel du locuteur. Une moitié de l'espace est toujours cachée au locuteur. L'usage de *avant* / *après* doit être inverse pour les deux échelles.

Vandeloise choisit le mouvement pour introduire un ordre et une orientation selon un axe dans l'espace. Pour cela, le locuteur doit être la source de tout alignement entre *a* et *b*, pour disposer du rôle sans lequel il n'est pas de rencontre potentielle. Le locuteur joue un rôle et en tant qu'énonciateur il définit un ordre car il est au pôle initial s'il n'indique pas le repère qui est à l'origine de l'orientation. Cette opposition entre la personne du locuteur et l'énonciateur est également éclairante pour les ordres dynamiques. L'échelle des pôles dynamiques est un chemin, une trajectoire du mouvement partagé par deux mobiles dont le but est le pôle terminal. Le locuteur n'a pas de rôle dans l'orientation des ordres dynamiques. Il n'y a pas de changement dans l'usage de la préposition même si le locuteur est mobile. La place du locuteur est indifférente, l'énonciateur est source de l'ordre. « L'ordre est affaire de point de vue plutôt que de distance » (Vandeloise).

Pour l'espace, l'ordre mutuel de deux notions peut être changé, *avant* sera remplacé par *après* et inversement. Or pour le temps, ce n'est pas possible ni par la position de l'énonciateur ni en prenant comme origine un repère objectif : * *vu du début de l'année Pâques est avant l'Ascension, mais vu de la fin de l'année, l'Ascension est avant Pâques.*

2.2.6.2.2. *sur* / *sous*

Ce sont avant tout, des prépositions de lieu (*sur* / *sous la table*) mais aussi des prépositions de causalité (*sous le choc, il pâlit ; sur la recommandation d'un pont, il eut la légion d'honneur*) et des prépositions temporelles (*sur ce bon mot, il partit ; sous le règne de Louis XIV...*) composées d'une sous-classe représentée par les deux séries : *sur* (ordre + écoute + proposition + ...) et *sous* (contact + traitement + prétexte + ...).

¹ Cf. *op. cit.* [Berthonneau93], p. 51.

Vandeloise (1986) relève cinq caractéristiques pour ces deux prépositions :

1) si S2 est *sur/sous* S1, S1 est généralement plus bas/haut que S2

* *sur* : dans le cas d'énoncés statifs, l'ordre vertical est habituel : *le chat est sur la table*,

dans le cas d'énoncés dynamiques, le verbe de mouvement définit l'ordre «vertical» : *Max a appuyé sur la touche* (même si l'interrupteur ne peut se mouvoir qu'horizontalement),

* *sous* : l'analyse est plus complexe pour un exemple tel que *le chat est sous la table* : en effet, doit-on considérer la table comme un plateau, comme un volume ? En ne tenant compte que de l'axe vertical habituel, un objet découpe un certain espace ; *sous* indique l'inclusion dans cet espace, ou l'espace que découpe l'objet de par sa structure ou son usage (*dans* désigne l'espace propre à l'objet lui-même).

2) si S2 est *sur* S1, il y a généralement contact (indirect) entre S1 et S2,

3) si S2 est *sous* S1, S2 est généralement rendu inaccessible par S1,

4) dans les relations S2 est *sur/sous* S1, S2 est généralement plus petit que S1,

5) si S2 est *sur* S1, l'action de S1 s'oppose à l'action de la pesanteur sur S2.¹

2.2.6.2.3. *dans / hors (de)*

En règle générale, *hors* se construit avec la préposition *de*. Toutefois avec des verbes du type *sortir, tirer*, on utilise seulement *de* ; en effet, *sortir hors de* est un pléonisme, et un germanisme. Cette préposition s'emploie avec des verbes statifs et marque la position ou le résultat d'un changement de position : *il est hors de l'eau*, et avec des verbes dynamiques et indique un changement de lieu dont le point d'origine se trouve à l'intérieur de quelque chose : *il court hors de la maison*.

Il existe de nombreuses analyses de cette préposition *dans* : Lindkvist (1950), Bennett (1975), Cooper (1968), Leech (1969), Miller & Johnson-Laird (1976), Moilanen (1979), Herskovits (1982, 86, 88), Wunderlich (1982, 85), Vandeloise (1984, 85, 86), Lutzeier (1985), Hawkins (1988), Garrod & Sanford (1988), Bierwisch (1988), Herweg (1989), Pribbenow (1989), Borillo & Vieu (1992) et Hottenroth (1991).

Selon Vandeloise, « *a est dans... b* si le site [i.e., y] contient (partiellement)... la cible [i.e., x] » (1986:224) ; il spécifie cette relation de contenant comme suit : « Le contenu est inclus, au moins partiellement, dans le contenant ou dans la fermeture convexe de sa partie contenante » (1986:225).²

Si on compare les deux couples de phrases suivants :

- | | | |
|------------------------------------|---|----------------------------------|
| a) <i>les pêches dans le bol</i> | / | b) <i>la fissure dans le bol</i> |
| c) <i>les objets dans la boîte</i> | / | d) <i>le trou dans la boîte</i> |

on remarque qu'il y a deux régions possibles par rapport à l'objet de référence : la région incluse dans l'objet (a et c), et la matière de l'objet, c'est-à-dire l'objet de

¹ Cf. *op. cit.* [Berthonneau93], pp. 138-142.

² Cf. *op. cit.* [Zelinsky-Wibbelt93], p. 41.

référence (b et d). Les propriétés de l'objet localisé déterminent la sélection de la région à laquelle on pense parmi toutes les « régions-dans » identifiées. Il en est de même avec *l'oiseau dans l'arbre* et *la hache dans l'arbre*.

Pour des exemples tels que ceux qui suivent, la différence d'interprétation est due à l'inclusion totale ou partielle, mais la région identifiée par *dans* est la même dans chacune des paires :

<i>la pipe dans la bouche</i>	/	<i>le bonbon dans la bouche</i>
<i>le paquebot dans l'eau</i>	/	<i>le poisson dans l'eau</i>
<i>les fleurs dans le vase</i>	/	<i>l'eau dans le vase</i>

Ce sont les propriétés géométriques caractéristiques de ces régions qui déterminent l'utilisation de *dans* et non celle de toute autre préposition, alors que la différence d'interprétation entre l'inclusion partielle ou totale provient de la connaissance générale des relations spatiales habituelles entre l'objet localisé et l'objet de référence. Ceci est également vrai pour les différences d'interprétation dans les exemples suivants où l'objet localisé est compact (objet séparé dans l'objet de référence) ou il se mêle à l'objet de référence (par mélange, dissolution, réaction chimique...) : *la cuillère dans le café*, *le sucre dans le café*, *le lait dans le café*, *l'oxygène dans l'eau*.

L'interaction entre *dans* et l'objet de référence semble d'abord venir du processus d'interprétation, et c'est seulement à ce moment que les processus secondaires opérant sur le résultat de cette interaction sélectionnent une des « régions-dans » identifiées, ou déterminent les différences d'interprétation telles que l'inclusion totale ou partielle...

En règle générale, on admet que les informations codées dans la représentation sémantique doivent être une généralisation recouvrant toutes les instantiations possibles sur le niveau conceptuel. Les instantiations de *dans*, c'est-à-dire les régions que *dans* peut identifier, peuvent avoir des propriétés géométriques diverses.

<i>la lettre dans la boîte</i>	<i>le vin dans la bouteille</i>
<i>le vin dans le verre</i>	<i>la viande dans l'assiette</i> (aussi sur)
<i>l'homme dans la foule</i>	<i>l'oiseau dans l'arbre</i>
<i>le clou dans la planche</i>	<i>la fissure dans le bol</i>
<i>un trou dans le papier</i>	<i>le ballon dans l'air</i>
<i>la vie dans l'espace</i>	<i>le monastère dans l'enceinte du fossé</i>
<i>la ville dans ses murailles</i>	<i>l'homme dans le cercle</i>
<i>les maisons dans cette région</i>	<i>un vieil homme dans son fauteuil</i>
<i>le bébé dans les bras de son père</i>	<i>l'homme dans l'encadrement de la porte</i>
<i>la grosse tabatière dans ses doigts</i> (aussi entre)	

On peut constater l'hétérogénéité des « régions-dans ». Les critères de classification seront les suivants : quelle est la dimensionnalité de la région, a-t-elle ou non des limites, est-elle matériellement occupée ? On appellera les différentes classes d'interprétations les types d'utilisation.¹

¹ Pour plus de détails sur l'étude de la préposition *dans*, cf. [Zelinsky-Wibbelt93].

Type d'utilisation 1 :

La « région-dans » est tridimensionnelle, creuse (c'est-à-dire qu'elle n'est pas occupée par l'objet de référence) et matériellement bornée de tous les côtés, l'objet de référence est un contenant clos, il peut y avoir une relation contenant/contenu. Ça semble être l'utilisation la plus fréquente et la plus prototypique de *dans* (*les objets dans la boîte, les robes dans l'armoire, le vin dans la bouteille*). A ces contenants, on peut ajouter les maisons, les voitures et autres véhicules, les organes « entourants » (*les enfants dans la maison, le sang dans le cœur, le liquide dans l'estomac, le bonbon dans la bouche*). Dans ces exemples, la région entourante est une partie constitutive de l'objet de référence, mais ce n'est pas toujours le cas (*les fleurs ou le cadeau dans le papier jaune, les jetons dans la main*). Il faut comme prérequis pour ce mécanisme que les objets de référence soient malléables, c'est-à-dire que l'on puisse lui donner une forme afin qu'il puisse en entourer un autre.

Type d'utilisation 2 :

La « région-dans » est tridimensionnelle, creuse mais partiellement entourée par l'objet de référence (*les pêches dans le bol, le vin dans le verre*). L'objet localisé est hors des limites de l'objet de référence (plus l'objet entourant est ouvert et plus c'est vrai). On utilise *dans* car on présuppose que la région est partiellement entourée et qu'elle fait partie de l'objet de référence. Les objets ne sont pas entourés de tous les côtés. Toutes les catégories d'objets du type d'utilisation 1 peuvent apparaître ici puisque les contenants peuvent être plus ou moins ouverts (*les robes dans la valise, le vin dans la bouteille*). Les objets de référence n'entourent pas obligatoirement la « région-dans » de tous les côtés (*la femme dans la petite barque, un homme dans son fauteuil*).

Les « régions-dans » peuvent présenter des limites variées : ainsi on peut trouver des contenants de plus en plus ouverts (*la viande dans l'assiette*), des objets de référence qui entourent partiellement des régions ouvertes sur plus d'un côté tels que les tuyaux, les fossés, les tranchées, les cages... (*le gaz dans le tuyau, l'eau dans la tranchée, les voitures dans la rue*).

Comme dans le cas de constructions ayant un contenant ouvert comme objets de référence, le prérequis pour l'utilisation de *dans* est que l'on considère les régions partiellement entourées comme mentalement fermées et qu'on les traite comme partie de l'objet de référence. La limite de la région peut être réduite à un cercle ou à quelques points (*la serviette dans l'anneau, la vis dans l'écrou, l'homme dans l'encadrement de la porte*).

Plus la limite de la région que *dans* identifie est réduite (dans les cas ci-dessus), plus la fermeture mentale de la région, et ainsi la possibilité d'utiliser *dans*, semble dépendre de la fonction de l'objet de référence comme contenant. La fermeture semble être plutôt motivée visuellement et non fonctionnellement dans les cas suivants : *l'oiseau dans l'arbre* (dans le sens de « haut de l'arbre entre les branches et les feuilles »), *l'homme dans la foule*.

Type d'utilisation 3 :

La « région-dans » est tridimensionnelle, (normalement) occupée par le matériel de l'objet de référence, et bornée de tous les côtés (*le clou dans la planche, le couteau dans le pain*).

On trouve également les objets de référence qui sont une quantité de substance bien que, très souvent, on ne se concentre pas sur les limites de la « région-dans » dans les interprétations respectives (*le clou dans le bois, le sucre dans le café, le pied dans la neige*). Les objets de référence de *dans* peuvent aussi être des substances en général, au lieu d'être des quantités limitées de substance.

Type d'utilisation 4 :

La « région-dans » est tridimensionnelle, vaguement ou non bornée (*les galaxies dans l'espace, la vie dans l'univers*). Mentalement, on crée quand même certaines limites, mais le mécanisme n'est pas identique à celui des contenants ayant des limites partielles car les limites que l'on imagine n'autorisent pas l'utilisation de mots tels que hors de, entrer, sortir...

Type d'utilisation 5 :

La « région-dans » est bidimensionnelle, et bornée (*le point dans le cercle, la boule dans le carré*).

Type d'utilisation 6 :

La « région-dans » est tridimensionnelle, bornée mais seulement deux dimensions sont nécessaires pour l'utilisation de *dans* ; l'objet de référence est souvent une région géographique, un territoire (*les caravanes dans le désert, les lapins dans le champ, les chalets dans les Alpes*). En fait on exprime la localisation uniquement par rapport au plan horizontal. Ce sont les limites territoriales qui motivent l'utilisation de *dans*. On a également ce type de régions avec des constructions où l'objet de référence est une délimitation telle qu'un fossé, un mur... (*la ville dans ses murailles, les moutons dans la clôture*).

Type d'utilisation 7 :

La « région-dans » est bidimensionnelle, ou deux limites sont suffisantes pour l'utilisation de *dans*, partiellement, vaguement ou non bornée ; on peut retrouver des régions géographiques (*le point dans l'angle, les maisons dans le voisinage, cette petite tente dans l'immense étendue*). Soit on complète les bornes (cas 1), soit on les crée mentalement (cas 2), soit on n'en tient pas compte (cas 3).

Type d'utilisation 8 :

La « région-dans » est unidimensionnelle (*l'intervalle dans la ligne*) ; c'est un type d'utilisation très rare. La classe d'objets localisés semble être très limitée.

En résumé les « régions-dans » possibles sont tri-, bi- ou unidimensionnelles. Elles peuvent être :

- * totalement, partiellement, vaguement ou non bornées
- * creuses (et peuvent être une partie constitutive de l'objet de référence ou être entourées par celui-ci sans en être une partie) ou matériellement occupées par l'objet de référence.

Dans d'autres cas, les analyses sont classées selon :

A) Les outils conceptuels qu'elles utilisent. L'expression *a est dans b* peut être décrite par :

- la tridimensionnalité de l'objet prépositionnel *b*,
- l'inclusion de *a* dans *b*,
- la relation contenant / contenu.

B) Le nombre de niveaux selon lesquels elles organisent les usages de la préposition *dans*.

- plus de deux niveaux (souvent reliés par des principes associatifs),
- deux niveaux considérés comme des modules indépendants,
- pas de séparation de niveaux.

C) La justification des niveaux.

- la représentation ou le nombre d'exemples motivés par un même principe,
- la simplicité générale du lexique qui se manifeste par la monosémie de la préposition et la compositionnalité de l'expression *a est dans b* (entièrement prédictible à partir du sens des parties),
- quant aux analyses à un seul niveau, si elles sont descriptivement aussi adéquates que les théories à plusieurs niveaux, elles se justifient directement, selon Vandeloise, en faisant l'économie d'une distinction inutile.

En 1988, Hawkins fait une analyse de la catégorie MEDIUM constituée par les objets prépositionnels des prépositions. Page 21, elle définit le Principe d'Association Central (PAC) comme « le principe qui permet de catégoriser un nombre maximum d'éléments de la catégorie et qui est aussi distinct que possible des principes d'association centraux qui caractérisent les autres catégories », et les Principes Associatifs Dépendants (PAD) comme « les principes qui relient les usages périphériques de la catégorie aux éléments centraux ».

Constatant que 60% des membres de son corpus sont tridimensionnels, elle choisit ce principe comme caractéristique connue de la catégorie MEDIUM. 2 DIM représente le PAC de la catégorie surface, et 1 DIM celui de la catégorie CHANNEL.

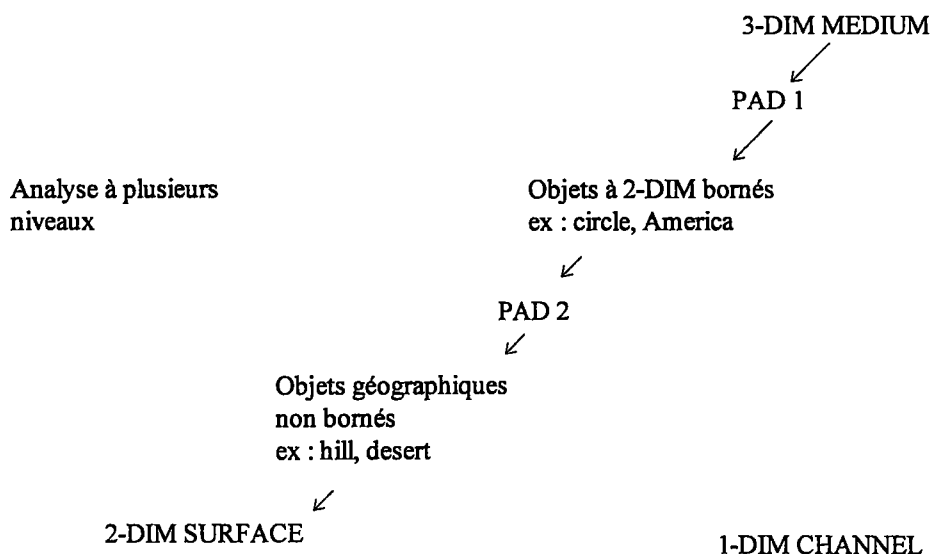


FIGURE 3 : LES CATEGORIES « MEDIUM » DE HAWKINS ¹

Hottenroth (1991) considère qu'une partie seulement des objets tridimensionnels sont les objets prépositionnels les plus caractéristiques de la préposition *dans*, à savoir ceux qui sont creux et matériellement bornés de tous les côtés (auxquels on peut ajouter la fermeture mentale des contenants ouverts). Elle considère que les régions géographiques sont tridimensionnelles. Elle introduit la notion de dimensions pertinentes pour les régions géographiques (deux pour celles-ci) et pour l'eau (une seule pour cette dernière).

Miller & Johnson-Laird (1976) utilisent, quant à eux, la notion d'inclusion et estiment que x est au moins partiellement inclus dans y (définition monosémique).

« In (x, y) : referent x is in a relatum y if : part (x, z) et Incl (z, y) [p. 385] ».

Vandeloise pense que le deuxième sens, l'inclusion partielle, émerge du premier sens, l'inclusion totale (définition bisémique).

Herweg (1989) travaille dans le cadre des analyses modulaires. Il se base sur les descriptions de Wunderlich (1982, 85). Il y a deux modules indépendants : un module sémantique et un module conceptuel. Il propose la définition suivante, et paraphrase Wunderlich :

« (lok(x), place (y)). Place (y) umfaßt nicht nur den Raum, den y materielle einnimmt, sondern gegenbenenfalls auch den inneren «leeren», d. h. von den materiellen Teilen von x (eventuell nur partiell) umschlossenen Raum ».

(lok(x), place (y)). Place (y) ne comprend pas seulement l'espace que y occupe matériellement, mais au contraire, également le « vide » intérieur, c'est-à-dire, l'espace circonscrit par les éléments de x (éventuellement seulement en partie).

x est *dans* y s'il existe un objet x et s'il existe un objet y tels que la place de x (comprenant la partie matérielle de x et certains de ses vides) soit une partie spatiale de la place de y (comprenant la partie matérielle de y et certains de ses vides).

¹ La figure est extraite de *op. cit.* [Berthonneau93], pp. 21.

Herweg écrit :

« das beschriebene funktionale Konzept eines Behältnisses ist kein Bestandteil der Wortbedeutung der Präposition *in* in der hier vertretenen Konzeption der lexikalischen Semantik. Das Behältniskonzept muß keinesfalls in allen Verwendungen von *in* erfüllt sein ». (Herweg (1989, p.121))

Le concept fonctionnel décrit d'un récipient n'est pas un élément de la signification de la préposition *in* dans la conception soutenue ici de la sémantique lexicale. Le concept de récipient ne doit, en aucun cas, être rempli dans tous les emplois de *in*.

Il donne comme exemple : *Peter ist in Amerika*, (*Pierre est en Amérique*) où l'Amérique n'exerce aucune force sur Pierre.

On peut faire une comparaison entre les analyses d'Hawkins et d'Herweg :

	Hawkins	Herweg
Interprétation directe	3-DIM (complètement borné ou non)	complètement borné (à 3-DIM ou 2-DIM)
Interprétation indirecte	2-DIM	non borné

Pour la phrase : *the flowers in the vase* (*les fleurs dans le vase*), Hawkins utilise le Principe d'Association Central (tridimensionnalité), Herweg le Module conceptuel car il y a une inclusion partielle.

Selon Vandeloise, il est impossible de décrire complètement la préposition *dans* en se limitant à la forme des objets qu'elle met en relation.



La lampe est dans la douille

* La bouteille est dans le capuchon

le contraste est justifié par des facteurs d'ordre dynamique : la douille exerce une force sur la lampe et détermine sa position, l'inverse se produit entre la bouteille et le capuchon.

Le terme qui décrit l'objet déterminant la position du second objet impliqué dans la relation doit se trouver à droite de la préposition *dans* et non à gauche. *b* joue le rôle d'un site qui détermine la position de la cible *a* (connue pour les autres prépositions spatiales) (Vandeloise (1986)).

Il n'y a plus de doute sur l'influence de la force sur l'usage de la préposition *dans*. L'interaction dynamique est un trait dominant de la relation contenant / contenu. Dans la phrase *les fleurs sont dans le vase*, c'est le vase qui détermine la position des fleurs et non l'inverse (qu'elles soient partiellement ou totalement incluses dans celui-ci).

Vandeloise décrit les relations contenant / contenu grâce aux traits suivants :

- 1) le contenant contrôle la position du contenu
- 2) le contenu se déplace vers le contenant, exemple :

* le doigt dans la bague
mets ton doigt dans la bague.

- 3) le contenu est inclus (au moins partiellement) dans le contenant ou dans la fermeture convexe de sa partie contenante.

Contrairement aux solutions géométrique et topologique, l'analyse fonctionnelle permet de traiter sur le même niveau tous les usages spatiaux de la préposition *dans*.

La formulation des traits de la relation contenant /contenu est modifiée. Le contrôle exercé par le contenant sur le contenu dans le premier trait peut se manifester de trois manières différentes :

- 1) un mouvement du contenant entraîne un mouvement similaire du contenu
- 2) le contenant limite les mouvements du contenu
- 3) le contenant empêche l'accès du contenu aux éléments extérieurs.

Cuyckens affirme que la préposition néerlandaise¹ *in* exprime une relation de COINCIDENCE entre les arguments *x* et *y*, qu'elle peut aussi impliquer une relation partie-tout entre *x* et *y*, et qu'elle lexicalise des relations spatiales statiques de même que des relations dénotant un passage. Les deux principaux éléments de la sémantique de *in* sont la relation COINCIDENCE et le «medium» de configuration spatiale associé à l'entité *y*.

Il faut donner une préférence au traitement de *in* en terme de relation de COINCIDENCE plutôt abstraite entre *x* et une configuration spatiale spécifique associée à *y* plutôt qu'à une description en terme d'information relationnelle plus spécifique.

On appellera la configuration spatiale «medium». Ce terme est introduit par Hawkins (1985, p. 95), qui définit le «medium» comme « une condition, une atmosphère dans laquelle quelque chose peut fonctionner, se développer » (Webster).

En somme, on admet que *in* lexicalise une relation de COINCIDENCE entre *x* et la configuration spatiale « medium » associée à l'entité *y*.

NON PARTIE - TOUT

medium 3DIM, borné, perméable/poreux

Le premier type de «medium» associé à une entité particulière correspond à l'intérieur défini par ses parois externes solides. On trouve dans cette catégorie toutes les entités qui peuvent être caractérisées comme « enveloppe ou construction ayant des limites externes solides qui circonscrivent un intérieur tridimensionnel avec une consistance quelque peu différente de celle des limites » : *the books in the cupboard, the jewels in the box...* [Hawkins (1988:253)]. Il est relativement rare de trouver des entités avec une fermeture complète, « si bien que certaines limites de l'intérieur seront imaginaires » : *nous sommes dans le tunnel...* [Hawkins (1986, p. 151)].

On trouve dans le deuxième type de «medium» des espaces limités par deux plans ou deux objets cylindriques qui se rejoignent en un angle : *the man was sleeping in the gutter...*, les limites ne sont pas incluses dans la zone active.

¹ L'étude de Cuyckens est donnée à titre indicatif car elle peut apporter des éléments nouveaux par rapport aux travaux précédents. Elle est présentée dans *op. cit.* [Zelinsky-Wibbelt93].

On trouve dans le troisième type de « medium » des entités entières. La zone active de *y* est un espace tridimensionnel, perméable et borné par un autre objet : *he was standing in the door....*

medium 3DIM, borné, non perméable/non poreux

La zone active de *y* qui entre en relation de COINCIDENCE avec *x* est le morceau entier de matériau indiqué par *y*. Les entités qui font partie de ce type particulier de «medium» sont des morceaux solides de matériaux tels que des planches, des murs... Contrairement au premier type de « medium », il n'y a pas de contraste entre la consistance interne et les limites de ce « medium ». Ce morceau solide peut exister par lui-même ou être une enveloppe qui circonscrit un intérieur. Parmi les entités qui figurent dans ce type de « medium », on trouve aussi des substances moins solides, qui doivent être placées dans un contenant afin qu'elles ne se répandent pas de façon incontrôlable : *I didn't put any sugar in my coffee yet, the crumbs in the jam...*

medium 3DIM, non borné, perméable/poreux

La zone active de *y*, et en même temps son « medium » dénote l'entité *y* elle-même. Il faut ajouter que si *y* désigne une grande partie d'espace, il peut se réduire à un morceau spécifique avec des limites vagues : *the plane in the fog, the birds in the air...*

medium 3DIM, non borné, non perméable/non poreux

Dans ce type de « medium », on trouve des entités dont les limites sont vagues ou inexistantes : *they were sitting with their lags in the sand, fish swim in water...*

medium 2DIM, borné

Dans ce type de « medium », on trouve :

- * des entités géographiques (cercle, carré...), ces surfaces bidimensionnelles sont limitées par une ligne unidimensionnelle qui, elle-même, ne fait pas partie du « medium » ; en fait, cette ligne de bordure constitue à elle seule une zone active : *the children were standing in the circle*
- * un ensemble de lignes qui se rencontrent à un angle et circonscrivent une surface bidimensionnelle ; ces lignes ne font pas partie du « medium » mais ne forment pas non plus une zone active à elles-seules : *draw a circle in the corner formed by the two lines*
- * une partie de page, délimitée par une ligne (imaginaire) qui marque la division de la page, cette ligne ne fait pas partie du « medium » : *your marks are in the margin*
- * des entités géographiques et géopolitiques, les limites de ces entités peuvent constituer à elles seules une zone active : *he lives in Germany*

medium 2DIM, illimité

Dans ce type de « medium », on trouve des entités géographiques et géopolitiques qui sont marquées par rapport au reste de la surface de la terre, mais dont les limites ne sont pas claires : *we live in the north/the desert...*

PARTIE - TOUT

Contrairement à *in* qui exprime une relation non partie-tout, *in* dans les relations partie-tout ne lexicalise pas une relation de COINCIDENCE avec le «medium» associé à *y*. C'est l'emboîtement de *x* dans *y* qui motive l'utilisation de *in* dans ce type de relations.

Le concept de «medium» peut être associé à une structure de ressemblance de famille. Il existe des «mediums» tridimensionnels poreux et non poreux, bornés et non bornés mais aussi poreux et bornés, poreux et non bornés... Il en est de même avec les «mediums» bidimensionnels. Les «mediums» bidimensionnels bornés doivent marquer une certaine division. «L'objet de référence doit être une des aires provenant de la division d'une surface» (Herskovits, 1986, p. 153). Il peut y avoir des divisions sur une page, des figures géométriques, des entités géographiques ou géopolitiques et sur le corps humain. Attention : toutes les entités qui illustrent un «medium» borné ou non borné qui marque une division sur la surface de la terre ne demandant pas *in* ; les noms d'îles, par exemple, demandent *op*. Il semble que plus une île peut être vue comme une unité politique (tel un pays sur un continent, avec des frontières délimitées), plus les limites de l'île entrent en compte et plus on utilise *in*.

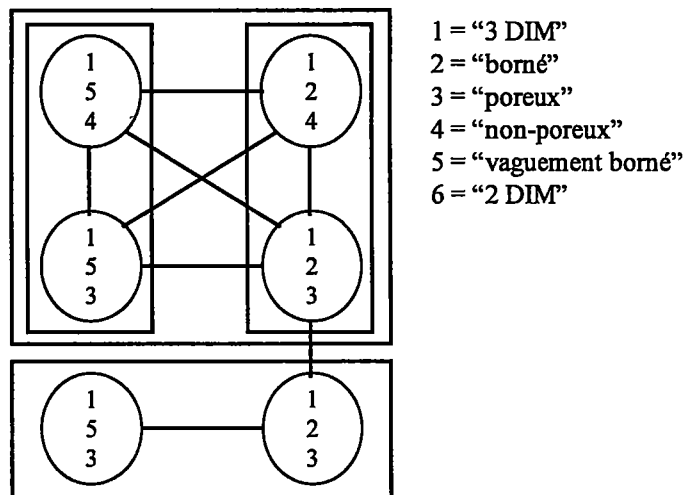


FIGURE 4 : STRUCTURE DE RESSEMBLANCE DES FAMILLES DE « MEDIUM »¹

PASSAGE

in peut assigner *x* à une série de localisations ou de passages par rapport à *y* : *he threw his tolls in the box*. *in* dénote un passage commençant à un point non spécifié dans l'espace et dont la localisation terminale est le «medium» associé à *y*.

Il se pose un problème avec les «mediums» tridimensionnels ouverts sur le dessus. Les contenants tels que les verres demandent, sans aucun doute, *in*. Mais plus les côtés verticaux diminuent en taille, plus on sera tenté d'utiliser *op* à la place de *in*. Il semble à Cuyckens que, tant que quelqu'un voit *y* dans sa fonction de contenant, *in* est autorisé. On peut difficilement assigner à des entités ayant des côtés verticaux presque inexistantes

¹ La figure est tirée de *op. cit.* [Zelinsky-Wibbelt93], p. 58.

une fonction de contenant, si ce n'est pour des assiettes : *there was a potato on / in the plate.*

2.2.6.2.4. devant / derrière

Ce sont des prépositions spatiales antonymes. *derrière* signifie soit en arrière de, au dos de : *se cacher derrière un arbre*, soit à la suite de : *marcher derrière quelqu'un*. *devant* a le sens de en avant de, en face de : *marcher devant quelqu'un*, *regarder devant soi*.

Pour expliquer l'opposition entre ces deux prépositions, on pourrait reprendre l'analyse que Vandeloise a faite de *avant / après*, car ces deux paires prépositionnelles se ressemblent. Le site, repère à partir duquel on localise la cible, de *devant* peut toujours devenir cible de *derrière* et réciproquement. On peut également dire que le mouvement joue un rôle capital dans la localisation de la cible et a un rôle très important dans la relation *avant / après*. En effet, il a un rôle dans l'espace à condition que les termes ordonnés soient mobiles. On peut réutiliser les figures proposées par Vandeloise et les adapter aux préposition *devant / derrière*.

Situation 1 : ordres dynamiques

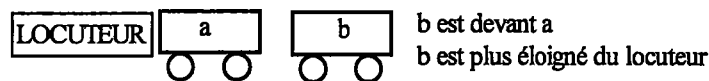


FIGURE 5 : DEVANT / DERRIERE AVEC LES ORDRES DYNAMIQUES

Situation 2 : ordres statiques

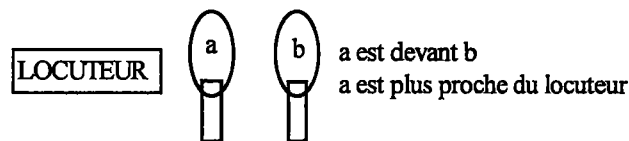


FIGURE 6 : DEVANT / DERRIERE AVEC LES ORDRES STATIQUES

Là encore, la place du locuteur est pertinente pour les objets statiques uniquement.

2.2.6.2.5. au-dessus de / au-dessous de

au-dessus de et *au-dessous de* expriment les mêmes rapports que *sur* et *sous* mais sans qu'il y ait contact. *au-dessus de* signifie dans la partie supérieure de et *au-dessous de* dans la partie inférieure.

Comme le dit Grévisse dans *Le bon usage* p. 1527, « certains grammairiens opposent *au-dessous de* à *en dessous de*, auquel ils ne laissent que le sens « sur la face inférieure de, dans la partie inférieure de ». Cependant la deuxième locution concurrence la première.

2.2.6.2.6. en haut de / en bas de

en haut de et *en bas de* expriment les mêmes rapports que *au-dessus de* et *en-dessous de* mais avec un contact.

2.2.6.2.7. *au-dedans de, en dedans de / au-dehors de, en dehors de*

Il n'y a pas de grandes différences entre *au-dedans de* et *en dedans de*. *au-dehors de* est rare et ne s'emploie qu'au sens local, en revanche *en dehors de* est plus fréquent et peut s'employer dans des expressions figurées.

2.2.6.2.8. *Les prépositions adverbialisées*

Certains mots peuvent fonctionner comme des prépositions ou des adverbes.

1. Voici la rue, arrêtez-vous devant le magasin.
2. Voici le magasin, arrêtez-vous devant

Dans l'exemple 2, la construction « absolue » de *devant* évite l'emploi de *lui*.

Ce genre de constructions est possible pour *avant, après, devant, derrière, depuis*, et en français familier pour : *contre, entre*. On peut aussi ajouter les locutions prépositionnelles composées de *par* : *par-delà...*

Il existe une correspondance formelle entre les prépositions *dans, hors, sous, sur* et les adverbes *dedans, dehors, dessous, dessus* : *Mets la dans la boîte, mets la dedans*.

Si on enlève *de* à de nombreuses locutions prépositionnelles, on obtient des locutions adverbiales : *à côté, au-delà, en dehors...*

On peut également combiner des prépositions et des adverbes : *d'ici, de là, au loin, de près...*

2.2.6.3. Les autres prépositions

2.2.6.3.1. *chez*

chez signifie « dans la maison de » et s'emploie pour des personnes : *il habite chez ses parents*, ou des indications de lieu en rapport avec des personnes, « dans le pays de » et même « dans la personne de, dans l'œuvre de » : *chez Flaubert*.

Selon le Littré, *chez* ne peut avoir pour régime que des noms de personnes ou d'êtres personnifiés. Mais depuis 1988, l'Académie accepte que l'on emploie cette préposition avec des noms d'animaux : *chez les rapaces, le bec est généralement corné*.

2.2.6.3.2. *contre*

contre indique qu'un objet est en contact avec un autre dans un axe vertical : *l'échelle est contre le mur*.

2.2.6.3.3. *entre*

Dans son sens spatial, *entre* indique un intervalle (*entre Paris et Marseille*). *entre* peut également renforcer *de* après un verbe exprimant une idée de séparation, et quand le lieu dont on s'éloigne implique la préposition *entre* (*on l'a retiré d'entre ses mains*).

2.2.6.3.4. jusqu'(e)

jusque prend la forme *jusqu'à* quand le second terme du syntagme prépositionnel est introduit par un article ou contient une forme verbale. Cette préposition introduit la fin d'un mouvement (*Paul va jusque Liège*).

2.2.6.3.5. par

Dans son sens spatial, *par* marque la transition : *il passe par Paris*.

La préposition *par* peut être précédée par la préposition *de* et signifie alors quelque part dans, à travers : *il voyage de par le monde*.

2.2.6.3.6. parmi

Au sens propre, *parmi* indique le lieu et signifie « au milieu de » (*parmi vingt personnes*).

2.2.6.3.7. pour

Dans son sens spatial, la préposition *pour* indique la destination : *il part pour Paris*.

2.2.6.3.8. vers

vers s'emploie pour indiquer le déplacement d'une cible en direction d'un site. Le site peut être une direction : *Paul va vers le Nord* ou un lieu : *Paul va vers le château*.

2.2.6.3.9. Combinaison de prépositions

dessus = de + sur

dessous = de + sous

dedans = de + dans

dehors = latin tardif de *deforis*

devant = de + avant

derrière = latin tardif de *retro*

Il est possible de combiner ces prépositions avec une autre préposition :

de : de dessus (N) / de dessous (N)

de dedans (N) / de dehors (N)

de devant (N) / de derrière (N)

à : au-dessus (de N) / au-dessous (de N)

au-dedans (de N) / au-dehors (de N)

au-devant (de N) / inusité

en : en dessus (de N) / en dessous (de N)

en dedans (de N) / en dehors (de N)

par : par-dessus (N) / par-dessous (N)

par-devant (N) / par derrière (N)

On peut également cumuler les prépositions entre elles :

d'en-dessous (de N)

de par dessous (N)

par en-dessous (de N)

Exemples d'emplois spatiaux :

sauter par-dessus un obstacle

retirer quelque chose de derrière un meuble.

Le sens de ces combinaisons dépend du sens de chacune des prépositions ou locutions prépositives : *par-dessus* indique bien un mouvement de transition et une élévation, *de derrière* précise la position initiale de l'objet et le sens du mouvement.

2.2.6.4. Concurrence entre les prépositions dites « vides » et les autres

Slatka, dans un de ces articles, affirme : « *dans* et *à* cèdent de plus en plus certains de leurs emplois à *sur* : *je travaille à Paris, je travaille sur Paris ; je l'ai lu dans le journal, je l'ai lu sur le journal.* Ou encore : *dans l'île, sur l'île.* Les gens de ma génération emploient *à* et *dans* ; nos enfants emploient *sur*. Il faut bien admettre que la représentation de l'espace a changé, et que certains objets ou lieux sont perçus d'abord comme des surfaces ; d'où l'emploi de *sur*. » ¹

2.2.6.4.1. à, sur et dans

1) En règle générale, on utilise *à* quand on envisage le lieu comme un point, *sur* comme une surface et *dans* comme un volume : *s'asseoir à son bureau, déposer un objet sur son bureau, enfermer une lettre dans son bureau.* Mais il existe des applications particulières : on emploie *dans la rue* (les maisons qui la bordent forment un volume), *sur la rue* (quand il s'agit d'une maison, d'une fenêtre ou d'une porte donnant sur la rue) et *à la rue* (quand quelqu'un est sans logis). De même, on dit *à la côte* quand on envisage la réalité du côté de la mer, mais *sur la côte* quand le littoral est vu du côté de la terre. ²

2) Il peut y avoir concurrence entre *sur* et *dans* sans que le sens ne diffère :

On s'assoit *dans un fauteuil* (si on envisage le volume) ou *sur un fauteuil* (si on considère la surface sur laquelle on s'assied). L'usage sera le même pour une bergère. En revanche, on utilisera la préposition *sur* avec les mots *canapé, divan, sofa* et évidemment *chaise*.

On lit, généralement, un article *dans le journal*, mais on peut le lire *sur le journal*, si ce dernier est étendu devant soi.

On préfère *sur* à *dans* avec les noms de livres qui ont la forme de répertoires (registre, annuaire, agenda, carnet...), et avec le mot *liste*. En revanche, on peut aussi bien dire *sur un cahier* que *dans un cahier* : Les deux prépositions alternent d'ailleurs dans les définitions du *Trésor* :

Cahier de brouillon [...] = Cahier SUR lequel l'élève rédige ses devoirs [...]. Cahier de correspondance [...] = Cahier DANS lequel sont inscrits [sic] les notes de l'élève [...]. Cahier de textes [...] = Cahier SUR lequel l'élève inscrit la liste des devoirs à faire [...]. ³

¹ SLATKA : « Le « bon » génie de la langue et le rouge à lèvres », in *Europe* 738, p. 92-103, Europe et Messidor, Paris, 1990.

² Cf. *op. cit.* [Grévisse93], p. 1498.

³ Définitions citées dans *op. cit.* [Grévisse93], p. 1500.

On dit *aller sur le balcon* mais *dans une loggia* car on considère que c'est un balcon fermé. Grévisse affirme que l'on peut utiliser *dans, sur ou sous une véranda*.

On peut se promener *dans* une prairie, un pâturage, un champ ou un pré. En revanche, on utilise la préposition *sur* avec le mot *pré* quand il s'agit du lieu convenu pour un duel, avec le mot *champ* dans des acceptions du type *sur le champ de foire, sur le champ de tir*.

3) En ce qui concerne une chose qui ne pénètre pas entièrement :

Grévisse affirme qu'il est logique de dire que la clé est *à la serrure* ou *sur la serrure* mais l'usage courant veut qu'on dise que la clé est *dans la serrure*.

De même il est plus fréquent d'avoir une cigarette *à la bouche* que *en bouche* ou *dans la bouche*.

2.2.6.4.2. en et dans

On utilise généralement la préposition *dans* lorsqu'il s'agit d'un volume, d'un espace à trois dimensions (*dans son bureau*). *en* s'emploie moins fréquemment, et plutôt dans des expressions plus ou moins figées, son régime est souvent sans déterminant (*en enfer mais attention au paradis*).

2.2.6.4.3. « Train » et ses prépositions

Les prépositions pouvant introduire le mot « train » sont *en, dans, avec, par, et à*.¹

« Train » peut introduire un complément en *de* ou en *pour*.

« Train », en fonction de sujet, peut induire un complément pour la phrase, où il figure, introduit par *à, pour, vers, sur*.

Mode de transport : *en* : *Guy se rend à l'école en train*.

Type de transport : *avec* : *Guy se rend à l'école avec le train*.

Objet physique individualisé : *dans* : *Guy monte dans le train*.

Lieu-instrument (pour transport collectif) : *par* : *Guy se rend à l'école par le train*.

Lieu-fonction : *à* : *J'irai chercher Guy au train*.

1. *Ce train va à / * pour / ? vers Genève* : cette phrase est possible pour un train à l'arrêt ou en mouvement, attention la destination doit être spécifique (*?* Paul va à l'arbre*)

2. *Ce train part pour / vers / ? à Genève* : *pour Genève* ne comporte en aucun cas l'anticipation que la cible est atteinte

2. *Ce train retourne sur / vers / * pour Genève* : « Au contraire de retourner, le verbe repartir ne construit pas un espace de représentation continu, mais au contraire, une discontinuité : d'où l'acceptabilité de repartir pour Genève (remarque de P. Encrevé).

¹ Cette étude a été menée par P. Cadiot dans un article intitulé « 'Train' et ses prépositions. Modes de donation du référent et principes cognitifs » in *Cahiers de Lexicologie* 58, Didier Edition, Paris, 1991, pp. 63-79.

2.2.6.4.3.1. Train et à

J'irai chercher Guy au train : le régime de la préposition est un espace extérieur à l'objet physique « train ». Le train est le lieu de repère et non l'objet contenant, c'est pourquoi Pierre Cadiot parle de lieu-fonction.

2.2.6.4.3.2. Train et avec

Guy se rend à l'école avec le train / * *Guy se rend à l'école avec un train* : type de transport autonomisé, concrétisé dans ses représentations, présence d'un déterminant défini obligatoire.

2.2.6.4.3.3. Train et dans

Le train est vu comme un espace concret, la préposition alloue une référence matérielle à son régime.

Guy se rend à l'école dans un train bondé - * *dans un train de nuit* : *dans* ne peut pas être utilisé pour un sous-type avec un déterminant indéfini.

Guy n'était pas dans le train de nuit (sous-type et non exemplaire individualisé).

Guy est dans le train - * *en train*

Guy vient en train - * *dans le train*

s'il y a mouvement on utilise *en*, s'il n'y a pas mouvement on utilise *dans*.

2.2.6.4.3.4. Train et de / pour

1. *Le train de Genève* ≠ 2. *le train pour Genève*

En 1. la préposition est désémantisée, le train peut venir de Genève ou partir pour Genève alors qu'en 2. elle ne l'est pas : *Je te laisse sinon je vais rater le train (pour/de) Genève*.

En 1. il s'agit d'un type (rubrique sur les panneaux des gares ou dans les écrits spécialisés) alors qu'en 2. il s'agit de l'occurrence d'un type : *Le train de Genève arrive au quai 12* / * *Le train pour Genève arrive au quai 12* // * *Tu devrais attraper un train de Genève et revenir dans la soirée* / *tu devrais attraper un train pour Genève et revenir dans la soirée*.

En 1. il s'agit d'un contenant, d'un objet matériel alors qu'en 2. il s'agit d'un temps, d'une durée : *Le train de Genève a les portes qui grincent* / ? * *Le train pour Genève a les portes qui grincent* // * *Pendant le train de Genève, j'ai tout le temps de penser à toi* / *Pendant le train pour Genève, j'ai tout le temps de penser à toi*.

En 1. il s'agit d'un bloc référentiel opaque alors qu'en 2. il s'agit d'une suite semi-syntagmatique : *Pour Dijon, il faut prendre le train de Genève*.

2.2.6.4.3.5. Train et en

1 *en* ∅ *train* / 2 *en train de nuit* (sous-type) / 3 * *en train bondé* : il y a une distinction entre le train vu comme objet physique individualisé (3) et train vu comme moyen de transport (1 et 2).

2.2.6.4.3.6. Train et par

Dans *Guy se rend à l'école par le train de 7h47* : le référent est conçu comme un mode extensif, alors que dans *Le voyage se fera par train de nuit* : le référent est conçu comme un mode intensif (sous-type).

2.2.6.4.3.7. Conclusion sur train et ses prépositions

Pour étendre cette description au-delà des seuls trains, il est nécessaire de prendre au minimum en compte une distinction de type transports individuels vs. transports en commun (remarque de D. Leeman). Les moyens de transports individuels ne s'accoutent pas notamment de la préposition *par* : *J'irai là-bas par train / bateau / avion / *vélo / *voiture*.

Une généralisation de cette description à tous les moyens de transport se heurte à d'autres difficultés : l'inacceptabilité de **Guy se rend au marché en cheval* doit-elle amener à conclure que « cheval » ne peut pas être considéré comme un « mode de transport » ?

« Il faudrait d'autre part examiner dans quelle mesure ces « types sémantiques idéalisés » se déduisent naturellement (dans le domaine des moyens de transport collectif) des caractéristiques prototypiques, générales, des prépositions en cause. Ainsi la notion de « mode de transport » peut être considérée comme une application (au domaine en cause) de la propriété prototypique de la préposition *en* de ne pas circonscrire d'espace concret, et donc de s'en tenir à une référence de type « notionnel » ou « intentionnel » (ce qui explique notamment sa mauvaise compatibilité avec le déterminant défini). Ainsi les propriétés de *dans* dans le domaine de la réalité étudié sont les mêmes que ses propriétés générales (circonscrire un espace concret), etc. Je ne suis pas sûr qu'un pareil optimisme soit fondé dans les cas de *avec* ou *par*. Cette question reste ouverte » (sur « avec » cf. Cadiot, à paraître).

2.2.7. Les prépositions spatiales en allemand

2.2.7.1. GENERALITES

2.2.7.1.1. Le système proposé par Marcq

Philippe Marcq a travaillé sur les prépositions spatiales en allemand. Nous nous sommes, entre autre, référés à son ouvrage¹ pour notre étude. Selon lui, on peut représenter le système des prépositions spatiales selon des caractéristiques géométriques : droite verticale (*über~unter*), droite de bout (*vor~hinter*), droite de front (*neben*). On peut le scinder en trois : système 1, système 2, système 3.

2.2.7.1.1.1. Le système 1

C'est un système situé dans un espace pratique défini. Le repère est un volume, c'est-à-dire une portion d'espace limitée par une surface.

¹ P. MARCQ : *Prépositions spatiales et particules « mixtes » en allemand*, Vuibert, 1972.

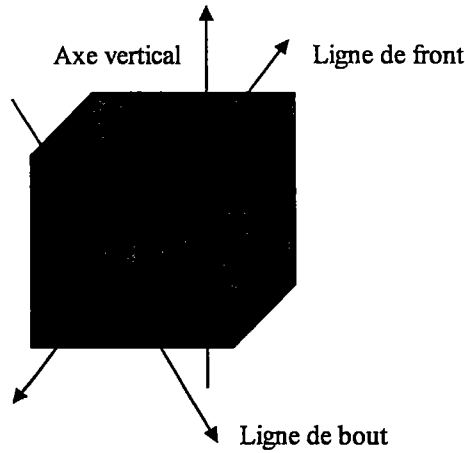


FIGURE 7 : LE SYSTEME DES PREPOSITIONS SPATIALES

Il se compose de quatre sous systèmes :

- * sous-système Ubi ? → relation locative [Wo ? Où ?]
- * sous-système Quo ? → relation directive (avec changement de lieu) [Wohin ? Où ?]
- * sous-système Qua ? → relation de passage [/ Par OÙ ?]
- * sous-système Unde ? → relation d'origine [Woher ? D'où ?]

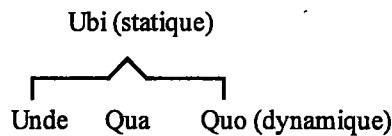


FIGURE 8 : LE SYSTEME 1 DE P. MARCQ

Ubi et *Quo* comprennent les mêmes prépositions disposées selon les mêmes oppositions, la marque du cas est la différence fondamentale.

On peut classer les prépositions selon des oppositions du système auxquelles les relations prennent part. On obtient trois axes :

- intérieur ~ extérieur
 - dans extérieur : contact ~ non contact
 - dans non contact : haut ~ bas
 - avant ~ arrière
 - à côté (droite) ~ à côté (gauche)

2.2.7.1.1.1. *Opposition intérieure ~ extérieure*

- >> Ubi → in + datif
- >> Quo → in + accusatif
- >> Qua → durch + accusatif le cas ne signifie rien, il est présent pour des raisons historiques
- >> Unde → aus + datif

2.2.7.1.1.2. Opposition contact ~ non contact

- >> Ubi → an / auf + datif *an* détermine le contact en toute position et
- >> Quo → an / auf + accusatif *auf* détermine le contact à l'appui de la pesanteur
- >> Unde → von + datif [origine à partir du contact]
- >> Qua → (pas d'opposition)

über + accusatif [passage par dessus avec ou sans contact], (*quer*)*über*... *hinaus*, (*quer*)*über*... *hinweg*, *querüber*

um + accusatif [passage sur le côté avec ou sans contact], *um*... *herum*

unter + datif ... *durch* [passage par dessous, circomposition], mode de passage rare dans l'expérience humaine, c'est peut-être pourquoi il n'y a pas de préposition spécialisée. On ne l'a rencontrée qu'une fois sous la forme « *under* + datif ... *hin* » dans un Sermon de Berthold von Regensburg.

2.2.7.1.1.3. Les autres oppositions

- >> Ubi : opposition haut ~ bas → *über* + datif ~ *unter* + datif
opposition avant ~ arrière → *vor* + datif ~ *hinter* + datif
- >> Quo : opposition haut ~ bas → *über* + accusatif ~ *unter* + accusatif
opposition avant ~ arrière → *vor* + accusatif ~ *hinter* + accusatif
opposition à côté ~ à côté → *neben* + accusatif ~ *neben* + accusatif
- >> Qua : déjà exposé dans le « contact »
- >> Qui : *unter* + datif ... *hervor* = de dessous (axe vertical)
hinter + datif ... *hervor* = de derrière (ligne de bout)

Les circompositions contiennent la détermination d'une portion d'espace par rapport à un repère et l'expression de la sortie de ce qui était dans cette portion d'espace.

	INTERIEUR	EXTERIEUR			
		CONTACT	NON-CONTACT		
			haut/bas	avant/arrière	à côté
UBI	in + datif	an/auf + datif	<i>über</i> + datif <i>unter</i> + datif	<i>vor</i> + datif <i>hinter</i> + datif	<i>neben</i> + datif
QUO	in + accusatif	an/auf + accusatif	<i>über</i> + acc. <i>unter</i> + acc.	<i>vor</i> + acc. <i>hinter</i> + acc.	<i>neben</i> + acc.
QUA	<i>durch</i> + acc.	L'opposition contact/non-contact ne fonctionne pas <i>über</i> + accusatif <i>unter</i> + acc. <i>durch</i>			<i>um</i> + acc.
UNDE	<i>aus</i> + datif	<i>von</i> + datif	<i>unter</i> + acc. <i>hervor</i>	<i>hinter</i> + acc. <i>hervor</i>	

TABLEAU 7 : TABLEAU RECAPITULATIF DU SYSTEME 1

Il n'est pas impossible de donner les correspondants français des prépositions du système allemand mais il faut avoir recours trop souvent aux verbes et on n'aboutit pas à un système prépositionnel français.

Formes renforcées de von :

- *von ... her* détermine le point de départ situé à une certaine distance de l'observateur, exprime le rapprochement de l'agent du procès à partir de ce point de départ par rapport à l'observateur,
- *von ... aus* détermine le point de départ où est situé l'observateur, exprime l'éloignement de l'agent du procès à partir de ce point de départ,
- *von ... an* est plutôt temporel mais on trouve *VON dieser Stelle AN bis dorthin*.

2.2.7.1.1.2. Le système 2

Le système 2 n'a pas les trois axes du système 1. Il n'y a pas d'opposition intérieur ~ extérieur et contact ~ non contact. Il y a toujours deux participants en cause. Ce système contient l'ensemble des indications de position ou de déplacement des deux participants l'un par rapport à l'autre. Il se compose de quatre sous-systèmes, indications de référence à l'intérieur desquelles les relations peuvent jouer :

A) détermination de l'espace compris entre deux participants ou plus, il y a cinq relations :

- a) locative → *zwischen* + datif, *unter* + datif (survivance du temps où *zwischen* n'existait pas [*unter ihnen* = au milieu d'eux], exprimant l'appartenance à un ensemble),
- b) directive → *zwischen* + accusatif, *unter* + accusatif (même acception que dans a)),
- c) de passage → *zwischen* + datif ... *hindurch*,
- d) d'origine → *zwischen* + datif ... *hervor*,
- e) de déplacement dans le même sens que les 2 participants se déplaçant eux-mêmes côte à côte → *zwischen* + datif ... *her*.

B) déplacement de deux participants orientés dans le même sens, il n'y a pas de relation possible mais seulement des modalités de réalisation :

a) précédant et suivant se déplacent à même allure

- * précédant → *vor* + datif ... *her*, datif + *voran*,
vor ... her et *voran* ne sont pas interchangeables. *voran* est plus restrictif et suppose une bonne « entente » entre les participants qui doivent être animés.

- * suivant → *hinter* + datif ... *her*, datif + *nach*

b) précédant fixe et suivant mobile → datif + *voraus*

c) précédant mobile et suivant fixe → *nach* + datif

d) deux participants se déplaçant côte à côte → *neben* + datif ... *her*

C) il existe six phases de rapprochement de deux participants orientés l'un vers l'autre :

1. Rapport statique : (six phases)

- 1) les deux participants se font face ou évoluent à une certaine distance l'un de l'autre → datif + *gegenüber*
- 2) les deux participants sont en contact → *an* + datif

- 3) un des participants est recouvrant, l'autre recouvert → *über* + datif
~ *unter* + datif
- 4) un des participants est entourant, l'autre entouré → *um* + accusatif
~ *in* + datif
- 5) comme phase 3 (mais de l'autre côté)
- 6) comme phase 2 (mais de l'autre côté)

2. Rapport dynamique : (six phases)

- 1) les deux participants sont à une certaine distance et se font face :
 - a) les deux sont mobiles et se rapprochent → datif + *entgegen*
 - b) l'un est mobile et se rapproche de l'autre → *auf* + accusatif
... *zu*
- 2) les deux participants entrent en contact → *an* + accusatif (comme dans le système 1)
- 3) il y a opposition recouvrant ~ recouvert → *über* + accusatif ~ *unter* + accusatif
- 4) il y a opposition entourant ~ entouré → *um* + accusatif ~ *in* + accusatif
- 5) l'entouré sort → *aus* + datif
l'entourant s'éloigne → *von* + datif
- 6) il y a ponctuellement perte de contact entre les participants → *von* + datif
dans un ancien système, on avait *ab* + datif, qui est encore utilisé aujourd'hui dans le domaine commercial (*Ware ab unserem Lager*).

D) Évitement, il y a quatre réalisations :

- a) collision
- b) traversée
- c) contournement
- d) passage

- * ponctuel : les participants circulent dans des plans différents et passent par un point de rapprochement maximal sans se rencontrer
- * linéaire : il y a une ligne de rapprochement maximal, le participant mobile est, ou n'est pas, en contact avec le participant fixe qu'il longe.

- 1) les participants ne circulent pas dans des plans parallèles. Il n'y a que l'opposition ligne droite, ligne courbe : collision / traversée ~ contournement → *gegen* + accusatif / *durch* + accusatif ~ *um* + accusatif.

Il ne s'agit plus des positions ou déplacements relatifs d'un participant par rapport à l'autre mais de l'opposition entre les déplacements possibles d'un seul participant, l'autre pouvant être considéré comme fixe

- 2) les participants circulent dans des plans parallèles :

- rapprochement maximal ponctuel ¹ →
an + datif ... *vorüber* (on ne s'arrête pas)
an + datif ... *vorbei* (on s'arrête au point de repère)

¹ *an* est le premier terme de circomposition le plus courant, mais d'autres ne sont pas exclus

- rapprochement maximal linéaire →
an + datif ... *entlang* (rapprochement par contact extérieur)
 accusatif ... *entlang* (rapprochement par intégration)
dans l'état actuel de la langue, plusieurs prépositions sont en concurrence avec « an ... entlang » : längs + génitif, entlang + datif (datif + entlang n'est pas exclu), accusatif + entlang.

A. Détermination de l'espace par rapport à deux participants ou plus	
zwischen / unter + D	verbe d'état ou exprimant un procès non orienté entre / parmi
zwischen / unter + A	verbe exprimant un procès orienté entre / parmi
zwischen + D (hin)durch	passer entre
zwischen + D hervor	sortir de (d'entre)
zwischen + D her	déplacement entre
B. Déplacement de deux participants orientés dans le même sens	
vor + D her	procès + devant, précéder
D + voran	précéder, façon de se déplacer + devant
hinter + D her / D + nach	suivre, procès + après / derrière
nach + D	procès + vers
neben + D her	déplacement + à côté de
C. Rapprochement de deux participants	
<i>1. Rapport statique (verbe exprimant un procès non orienté +)</i>	
D + gegenüber	en face de
an + D	cf. système 1
über + datif	sur / par dessus, recouvrir
unter + D	sous / par dessous / en dessous
um + A	autour de, entourer
in + D	cf. système 1
<i>2. Rapport dynamique (verbe exprimant un procès orienté +)</i>	
D + entgegen	à la rencontre de / au devant de
auf + A zu	vers, sur à, dans la direction de
gegen + A	vers
über + A / unter + A	sur, par dessus / sous, par dessous
um + A	autour de, entouré
in + A	cf. système 1
aus + D / von + D	cf. système 1
D. Evitement	
gegen + A	contre
durch + A	cf. système 1
um + A	tourner autour de
an ... vorbei	passer à côté de
an ... vorüber	passer devant, à côté de
an ... entlang A + entlang längs + G	} longer, le long de

TABLEAU 8 : TABLEAU RECAPITULATIF DU SYSTEME 2

2.2.7.1.1.3. Comparaison entre le système 1 et le système 2

Le système 1 est une émanation abstraite du système 2 et correspond à une toute petite partie de ce dernier.

Le système fondamental est le même dans les deux systèmes : il y a quatre oppositions mais les réalisations sont différentes selon la présence ou l'absence d'axes :

- 1° l'opposition « *um~in* » du système 2 est une opposition : entourant~entouré ou extérieur~intérieur. L'opposition « tout sauf *in~in* » du système 1 est une opposition extérieur~intérieur. S'il n'y a pas d'axe, on a que *um*, s'il y a des axes *um* se différencie en : *über~unter, vor~hinter, rechts neben~links neben*.
- 2° *an* détermine la position « en contact ». Il n'y a pas *auf* dans le système 2, s'il y a des axes, on a *auf* dans le système 1 (appui de la pesanteur).
- 3° *aus* et *von* appartiennent aux deux systèmes : au système 2 car il fonctionne même sans intervention des axes, au système 1 car il participe aux oppositions intérieur~extérieur et contact~non contact.
- 4° l'opposition *gegen/durch~um* de l'évitement dans le système 2 est une opposition de modalité de l'évitement « droite~courbe » sans plus. Quand il y a des axes, la droite (collision + traversée) n'est pas affectée, la courbe se différencie en *über/um/unter...durch* (sous-système « QUA »).

2.2.7.1.1.4. Le système 3

C'est le système de la cooccurrence¹. Le plan en question est le plan défini par la ligne de bout et la ligne de front du repère. Les prépositions sont *bei* + datif, *zu* + datif, *nach* + datif (ce *nach* est différent du *nach* du sous-système de la poursuite de deux participants).

Un double problème se pose :

- celui de la détermination des portions de plan (axe vertical exclu) :
 - bei* → que dans l'extérieur
 - zu* → dans l'extérieur ou dans l'intérieur
 - nach* → dans l'intérieur
- celui de l'expression de la relation locative ou directive :
 - bei* → que dans la relation locative
 - zu* → dans la relation locative et dans la relation directive
 - nach* → dans la relation directive

¹ Présence dans le même temps et dans le même plan que le repère, sans autre indication spatiale plus précise

	Avec marquant de cas (toujours datif)		Sans marquant de cas
	extérieur	intérieur	extérieur / intérieur
Relation locative	bei		zu (in)
Relation directive	zu	zu	nach

TABLEAU 9 : TABLEAU RECAPITULATIF DU SYSTEME 3

- 1° opposition *bei* locatif ~ *zu* directif : *zu* neutralise l'opposition intérieur ~ extérieur. Dans l'extérieur *bei* et *zu* neutralisent toutes les oppositions qui se trouvent dans le plan horizontal «avant ~ arrière», «à droite ~ à gauche»,
- 2° opposition *zu* locatif ~ *nach* directif : il ne s'agit plus ici que d'un reste des emplois de *zu* dans un système antérieur, rare dans la langue actuelle, il n'apparaît que dans un certain nombre de contextes déterminés :
- des lexies, c'est-à-dire des groupes qui ne sont aptes à recevoir aucun élargissement, ni aucune détermination : *ich bin zu Hause*,
 - quand le repère spatial est fourni par un nom de ville ou de pays, dans la langue actuelle *zu* est remplacé par *in* : *er wohnt zu / in Köln*.

nach est un substitut de marque de directif, surtout quand le repère spatial est donné par un nom de ville ou de pays sans article : *er fährt nach Köln*, par un adverbe de lieu : *nach hinten*. Il faut bien considérer que *nach* est le substitut de marque de relation directive quand le nom n'admet pas d'autres moyens de la marquer, essentiellement à cause de l'absence de tout adjectif grammatical : *ich fahre in die Schweiz*. Il faut bien faire la différence entre ce *nach* particulier et les deux *nach* préposé et postposé du sous-système de la poursuite : *ich fahre nach Deutschland* correspond à *ich fahre in die Schweiz*.

Origine	Relation locative zéro	Relation directive
von von hinten	hinten	nach nach hinten
↓	↓	↓
de derrière	derrière	vers l'arrière en arrière à la renverse

On ne peut pas représenter schématiquement les déterminations de la co-occurrence car les prépositions qui les constituent sont des substituts. Le lexique auquel on aboutit est d'une imprécision totale car le système français est d'une structure entièrement différente :

- * *bei* → à, sur, chez, près de + tout verbe de procès non orienté
- * *nach* → en / à + nom de ville ou pays + verbe de procès pouvant être orienté

- * *zu* locatif → à + nom de ville, dans les lexies : *zu beiden Seiten*
- * *zu* directif → à, chez, près de, sur, vers + tout verbe exprimant un procès orienté.

2.2.7.1.1.5. Cas particuliers

1° Les survivances de systèmes antérieurs :

- * *ob* + datif n'existe plus que dans quelques noms dont le plus connu est « Rothenburg ob der Trauer ». *ob* était autrefois le terme de neutralisation de toutes les prépositions en rapport avec la partie supérieure de l'axe vertical. *ob* pouvait remplacer toutes ces prépositions pour n'exprimer qu'une position « au-dessus » du repère sans autre détail.
- * *vor* + accusatif détermine une portion d'espace située avant le repère, exprime l'entrée de quelque chose qui vient du repère à l'intérieur de cette portion d'espace. La circomposition n'est utilisée que dans un nombre très restreint de contextes, pratiquement seules les différentes façons de regarder, rêver, faire entendre sa voix entrent en ligne de compte : *zuerst sah er still vor sich hin* → la traduction est difficile : d'abord il resta silencieux fixant le vide (ou : les yeux fixes). Dans des textes classiques, on trouve *für...hin* : *ich ging im Walde, so für mich hin* (Goethe). *vor ... hin* est plus récent, *vor* a pris la place de *für* quand celui-ci est devenu uniquement abstrait.

2° Prépositions en marge du domaine spatial pur :

On peut ranger dans cette rubrique des formations relativement récentes qui se rapportent à l'espace sans trouver place dans aucun des trois systèmes : *innerhalb* ~ *außerhalb* + génitif, *oberhalb* ~ *unterhalb* + génitif (de plus en plus accompagnés de *von*) sont des oppositions du sous-système Ubi du système 1, mais

- * *innerhalb* ~ *außerhalb* (à l'intérieur de ~ à l'extérieur de) ne déterminent qu'une opposition « intérieur ~ extérieur » sans autre indication spatiale,
- * *oberhalb* ~ *unterhalb* (au-dessus de ~ au-dessous de) représentent une détermination à une seule opposition et surtout géographique,
- * *diesseits* ~ *jenseits* (en deçà de ~ au-delà de) → la présence d'un repère et d'un observateur est obligatoire. C'est un petit système apparenté dans une certaine mesure au système 2 : les déterminations sont surtout relatives à deux participants

3° Emploi « faussement » spatial de *auf* :

Dans *ich gehe auf die Post*, il n'y a pas de détermination spatiale mais une détermination purement abstraite. Les déterminations spatiales, créées par notre comportement respectif, commenceront quand nous serons en vue du repère (bâtiment de la poste). On peut le trouver avec une relation locative : *er arbeitet auf der Post*.

En conclusion, on peut dire qu'il existe un rapport étroit entre les prépositions spatiales et les autres :

- * la succession *unde qua quo*, reconnue comme succession spatiale est aussi la succession temporelle à laquelle est attachée le destin de l'homme

- * l'espace est en trois dimensions et le temps est unilinéaire ; la structure fondamentale est différente mais il y a des parentés d'opposition (différence formelle : *durch* temporel est une postposition) :
 - espace : *durch* → intérieur ~ *über* → extérieur
 - temps : *durch* → relation entre un procès et une période incluant le procès, *über* → idem mais le procès est exclus de la période
- * *in* temporel est plus ambigu car il inclut aussi bien qu'il exclut le procès de la période
- * dans une large mesure, les prépositions abstraites sont des proches parentes des prépositions spatiales

2.2.7.1.2. Le système proposé par Schanen et Confais

La différence entre la relation locative et la relation directive est fondamentale en allemand : elle s'exprime non seulement par le choix de la préposition, mais aussi par l'opposition des cas accusatif / datif pour les neuf prépositions dites « mixtes ».

On peut classer les prépositions spatiales selon différents paramètres, relation locative / directive :

- * sans autre paramètre : *bei / zu*
- * paramètre « intérieur~extérieur »
- * paramètre « axes » (vertical + horizontal) avec ou sans autres paramètres (contact, pesanteur...)
- * autres paramètres

2.2.7.1.2.1. Sans autre paramètre

- * on utilise *bei* (+ datif) quand A est dans le même lieu qu'un repère B, ou *zu* (+ datif) quand A va dans le même lieu d'un repère B. B peut ou non être animé, la portion d'espace peut ou non être précisée (intérieur, extérieur, devant, derrière...) : *er ist bei seiner Tante, er geht zu seiner Tante.*
- * on a la même opposition dans des systèmes plus anciens ou concurrents :

LOCATIF ↔ DIRECTIF

zu ↔ nach	<i>er ist zu Hause, er geht nach Hause</i>
∅ ↔ nach	<u>devant les adverbes</u> : <i>er sitzt ∅ vorn, er sitzt nach vorn</i>
in ↔ nach	<u>devant les noms de localités et de pays sans article</u> : <i>er wohnt in Hamburg / in Deutschland, er fährt nach Hamburg</i>
in + D ↔ in + A	<u>devant les noms de pays avec article</u> : <i>er lebt in der Schweiz, ich fliege in die Schweiz</i>

2.2.7.1.2.2. Paramètre « intérieur~extérieur »

INTERIEUR (LOCATIF ET DIRECTIF)

in + D / A *er sitzt **im** Arbeitszimmer, er geht **ins** Arbeitszimmer*

EXTERIEUR (LOCATIF)

aus + D (rare) *mein Mann ist **aus** dem Haus*

außer + D (rare) *er ist **außer** Haus(e)*

außerhalb + G *er wohnt **außerhalb** der Stadt*

ESPACE ENVIRONNANT (LOCATIF)

um + A *er hat einen dicken Verband **um** den Kopf*

2.2.7.1.2.3. Paramètre « axes » (locatif = datif, directif = accusatif)

AXE AVANT~ARRIERE

vor + D / A ↔ *das Auto steht **vor der** Tür / **hinter der** Kirche*

hinter + D / A *ich stelle den Stuhl **vor die** Tür / **hinter die** Tür*

AXE DROITE~GAUCHE

neben + D / A *der Brief liegt **neben dem** Telefon, er setzte sich **neben** mich*

AXE VERTICAL

auf + D / A implique un contact à l'appui de la pesanteur : *das Geld liegt **auf dem** Tisch, ich lege das Geld **auf den** Tisch*

***auf** est aussi utilisé pour la notion d'intérieur (espace fermé) : **auf der Straße** mais in der Straße wohnen ; **auf** s'emploie aussi avec certains bâtiments ou institutions de la vie publique : **auf die Bank** gehen, **auf dem Markt**.*

über + D / A au-dessus (sans contact), sur (avec contact) : *die Lampe hängt **über dem** Tisch, ich hänge deinen Mantel **über** meinen*

unter + D / A au-dessous / sous (avec ou sans contact) : *der Schlüssel liegt **unter dem** Fußabstreifer, die Katze kriecht **unter das** Bett*

2.2.7.1.2.4. Autres paramètres

an + D / A contact sans que ni l'axe ni la position ne soient précisés : *das Bild hängt **an der** Wand, ich stelle das Fahrrad **an die** Wand*

gegen + A (directif) collision : *das Auto fuhr **gegen einen** Baum*
sens contraire : *er schwimmt **gegen den** Strom*

gegenüber + D (locatif)	<u>vis-à-vis, en face mais avec une séparation (route, fleuve, distance, espace vide...)</u> : <i>das Haus steht gegenüber der Kirche, er saß mir gegenüber</i> (postposition obligatoire avec un pronom)
DEPLACEMENT SIMULTANE	
neben + D [her]	<i>er lief neben dem Kind / zwischen den beiden Kindern her</i>
vor + D [her] ou D voran-	<i>der Hund lief vor dem Kind her</i> (déplacement de A devant B), <i>seine Leibwächter gingen ihm voran</i>
hinter + dat. [her] ou dat. nach	<i>der Hund ist die ganze Zeit hinter mir hergelaufen</i> (A est derrière B), <i>ich kenne den Weg nicht ! Fahr mir nach</i>
dat. voran signifie aussi « à la tête de » : <i>die europäischen Länder protestierten laut, allen voran Frankreich.</i>	
Selon le type de perspective (locative ou directive) on utilisera le datif ou l'accusatif, il faut toutefois faire attention aux verbes qui impliquent une perspective locative : <i>ankommen, anhalten, absteigen</i> (<i>er stieg in einem teuren Hotel ab</i>).	
Le choix de la préposition relative dépend du fait que le passage s'effectue par / sur / devant le repère visé par le membre du groupe prépositionnel.	
durch + A ou A (hin)durch	<u>passage à l'intérieur, c'est-à-dire par un espace fermé ou plein</u> : <i>er kam durch die Tür, ich höre durch die Wand [hindurch / durch]</i>
zu + D + hinaus/heraus	<i>er sah zum Fenster hinaus / heraus</i>
über + A + [hin / her / hinweg...]	<u>passage par l'extérieur et par le haut avec ou sans contact</u> : <i>er ging über die Straße, über die Brücke</i>
unter + D (ou A) + [hindurch / durch]	<i>er ging unter der Brücke [hindurch / durch]</i>
um + A + [herum]	<u>passage par l'extérieur et par le côté (contournement)</u> : <i>er lief um das Feuer [herum]</i>
an + D vorbei / vorüber	<u>passage par l'extérieur (la portion d'espace concernée peut être précisée par un adverbe)</u> : <i>wir gingen vorn / hinten am Haus vorbei</i>
an + D entlang ou A entlang ou längs + G	<u>passage de l'extérieur, parallèlement au repère (le long de)</u> <i>wir gingen am Fluß entlang, wir gingen den Fluß entlang</i>
zwischen + D ou acc. [durch / hindurch]	<u>passage entre deux repères</u> : <i>Karl-Heinz Rummenigge schoß den Ball zwischen den Beinen des Torhüters hindurch / durch</i>
bis + GP ou adverbe ou nom sans article	<u>marque le point d'arrivée après une relation de passage</u> : <i>er ging [durch das Zimmer] bis ans Fenster</i>

2.2.7.1.3. Le système proposé par Moilanen

En allemand, le système de prépositions statiques locatives est un système hétérogène. On peut le comprendre comme un conglomérat de plusieurs sous-systèmes, dont on retiendra les deux principaux. Markku Moilanen ¹ fait une distinction entre :

- * système de relations linéaires : relations grâce auxquelles deux individus sont reliés grâce à un axe. Ce système se divise en trois sous-systèmes :
 - relation horizontale : elle ne peut pas mettre en relation deux individus selon un axe vertical mais elle peut les mettre en relation selon un axe ascendant ; on utilisera les prépositions : *vor, hinter, gegenüber, neben, links, rechts*
 - relation verticale : relation inverse de la relation horizontale ; les prépositions employées seront les suivantes : *auf, über, unter, oberhalb, unterhalb*
 - relation *zwischen* : relation horizontale, verticale et selon l'orientation d'un axe,
- * système de relations non linéaires : axe de relations indifférentes ; on emploiera les prépositions : *an, um, in, innerhalb, außerhalb, bei*

2.2.7.2. DURCH

durch ² est une préposition de passage : on peut distinguer deux types de passage :

- le passage perfectif qui est terminé quand le trajector arrive à un point d'arrivée précis,
- le passage imperfectif dont l'étendue n'est pas déterminée.

durch, qui a le trait [+DIR] et qui gouverne l'accusatif, demande un passage continu (*Die Mauer durch Berlin, Der Zaun über den Berg*)

Dans certains cas, l'utilisation de *durch* est proche de celle de *in* : *er wandert durch den Wald / er wandert im Wald*. Cette relation entre *durch* et *in* est valable pour toutes les utilisations de *durch* et de *in* mais elle n'est pas symétrique : la situation de déplacement dans la forêt ne contient pas une situation de déplacement à travers la forêt.

L'objet localisé lui-même doit être l'argument de la relation de localisation LOC. En effet, *durch*, comme *in*, impose certaines conditions sur la dimensionnalité de l'objet de référence qui doit fournir au moins les dimensions fournies par l'objet localisé afin de pouvoir le contenir : un objet bidimensionnel ne peut contenir un objet tridimensionnel.

Les propriétés de dimensionnalité pertinentes sont déterminées par l'argument interne de la préposition. Pour des objets bidimensionnels tels que des cercles, des carrés..., l'inclusion ne peut faire référence qu'au bord du cercle.

$\lambda y \lambda x (\text{LOC}(x, \text{INT}(y)))$

¹ M. MOILANEN : *Statische lokative Präpositionen im heutigen Deutsch. Wahrheits- und Gebrauchsbedingungen*, Max Niemeyer Verlag, Tübingen, 1979.

² Cette préposition a été étudiée dans *op. cit.* [Zelinsky-Wibbelt93].

INT(y) doit être interprété comme l'intérieur de y, c'est-à-dire que x est entouré par les points intérieurs de y.

Les propriétés du passage sont : la linéarité, la continuité, la direction.

Les prépositions qui possèdent le trait [+DIR] et qui gouvernent ainsi l'accusatif peuvent être combinées avec des verbes causatifs tels que *stellen*, *legen* qui définissent un changement entre deux états de localisation et non pas un passage continu.

Dans une relation de passage, il y a toujours plus d'une région impliquée. Il existe deux types de relations entre deux régions : une relation d'ordre (passage discontinu orienté) et une relation de connexion (passage continu non orienté). Si on applique les deux relations, on obtient un passage continu orienté.

Dans une approche à deux niveaux, on utilise des paramètres qui sont des constantes dans la forme sémantique fixées par l'information conceptuelle.

On utilise le paramètre D(x) qu'on lit de la façon suivante : « une dimension à trouver par rapport à (l'argument externe de la préposition) x ». Si l'information conceptuelle associée au nom ne peut fournir un passage, le paramètre ne peut être instancié et l'expression est inacceptable (*der Weg durch den Wald, die Reise durch den Harz*). La Gestalt propriété d'un objet n'est manifestement pas suffisante pour déterminer l'acceptabilité (? *Der Teppich durch den Flur*).

Pour définir le passage, on n'utilisera pas toujours *durch* :

die Tuer in den Garten }
die Faehre nach Helgoland } concept de mouvement

der Blick durch den Flur }
das Fenster zum Hof } concept de perception

Der Schrei aus dem Wald => passage abstrait défini par une région source (la personne qui crie) et une région but (l'auditeur).

Dans *Die Mauer durch Berlin*, le passage ne comporte pas la direction mais est défini par rapport au « Gestalt » de l'objet localisé.

2.3. TRAITEMENT INFORMATIQUE DE LA LOCALISATION SPATIALE

Les travaux informatiques sur la localisation spatiale se multiplient : en 1978, Boggess a examiné les aspects d'une sémantique computationnelle des relations spatiales. En 1992, Briffault a développé un modèle informatique pour la description de la localisation spatiale en langue naturelle ; on pourra aussi noter les travaux de Aurnague & Vieu (1993), Borillo (1990), Briffault (1995), Freksa (1992).

2.3.1. Le système TOUR

Le modèle TOUR a été développé par Kuipers en 1978. Il se présente comme un modèle de la connaissance humaine de sens commun sur les environnements spatiaux de grande échelle. C'est l'un des premiers systèmes permettant la représentation

d'itinéraires complexes. Le modèle s'appuie sur la métaphore de la « carte cognitive », qui pourrait être vue comme l'équivalent d'une carte physique, dont elle partage les propriétés analogiques, même si elle n'a pas la même cohérence structurale que celle-ci. On peut la voir comme un réseau constitué de voies et d'intersections, c'est un répertoire de trajets, c'est-à-dire de procédures, correspondant généralement à des séquences ordonnées d'actions, permettant à l'individu de se déplacer d'un point à un autre.

Dans son modèle, Kuipers prend en considération cinq catégories de connaissance spatiale :

- a. les trajets sont des séquences d'actions permettant de se déplacer d'un lieu initial à un lieu terminal ; dans la simulation informatique du modèle, les trajets correspondent à des alternances d'instructions de type TURN (tourner) et GO-TO (aller tout droit),
- b. la structure topologique d'un réseau est représentée par des descriptions de chemins (PATHs) et des descriptions de lieux (PLACEs) ; la description d'un chemin inclut une description de l'ordre partiel des lieux sur ce chemin et possède donc une orientation unidimensionnelle définie par cet ordre,
- c. la position relative de deux lieux est définie par un vecteur relatif à un référentiel d'applicabilité limité, regroupant les informations d'orientation et de distance ; de plus, l'accessibilité d'un lieu à partir d'un autre est codée de manière indépendante des vecteurs d'orientation, dans la mesure où il est possible de connaître l'accessibilité mutuelle de deux lieux sans nécessairement connaître leurs positions relatives,
- d. l'utilisation de frontières définissant des régions fournit une connaissance qualitative partielle non vectorielle des positions,
- e. les régions, reliées par une relation d'inclusion, fournissent des niveaux d'abstraction utiles pour spécifier les relations entre leurs éléments.

Le modèle TOUR utilise trois classes de représentations :

- a. les représentations correspondant aux connaissances spatiales sur un environnement particulier, distribuées sur les catégories décrites ci-dessus,
- b. la description de la position courante d'un sujet en déplacement dans cet environnement ; une mémoire de travail, représentée par un pointeur de localisation, décrit la position courante en terme de lieu, de chemin, d'orientation unidimensionnelle sur ce chemin, de référentiel courant et d'orientation bidimensionnelle dans ce référentiel,
- c. des règles d'inférence permettant de manipuler les connaissances spatiales (règles comparant l'instruction courante, le pointeur de localisation et la description topologique de l'environnement ; des règles gérant l'orientation bidimensionnelle par rapport au référentiel courant ; des règles permettant de détecter des caractéristiques spécifiques de l'environnement ; des règles gérant la recherche de chemins).¹

¹ X. BRIFFAULT, M. DENIS : *Analyse d'un corpus de dialogues à bord d'un véhicule automobile*, LIMSI-CNRS, Orsay, 1995, p. 5.

En 1987, Habel a élaboré un formalisme de représentation à la fois propositionnel et visuo-spatial et a ainsi montré comment des représentations propositionnelles et des « images mentales » (utilisation de matrices bidimensionnelles) peuvent être utilisées conjointement pour effectuer des raisonnements spatiaux. De telles propositions ont connu un développement plus récent à travers la modélisation informatique des mécanismes d'imagerie mentale mis en œuvre à partir de descriptions spatiales (cf. Glasgow & Papadias, 1992).

Au cours de ces dernières années, on a vu se développer des modèles computationnels de l'acquisition et de la représentation des connaissances spatiales (cf. Gopal, Klatzky, & Smith, 1989 ; Gopal & Smith, 1990 ; Leiser & Zilbershatz, 1989).

2.3.2. VITRA

Le projet VITRA (VISual TRAnslator), financé par la *German Science Foundation* (DFG) depuis 1985, a pour but de contribuer à la recherche concernant la combinaison des systèmes de compréhension d'image et de traitement du langage naturel (Wahlster, 1989). Des études expérimentales ont été menées pour le développement de systèmes basés sur la connaissance et capable de traduire des informations visuelles en langue naturelle : courtes séquences d'un match de football (Herzog et al., 1989), différentes scènes de circulation avec des véhicules (Schirra et al., 1987), un piéton en déplacement (Herzog et Rohr, 1995)...

Les principaux thèmes de recherche incluent :

- * une sémantique référentielle des prépositions spatiales,
- * une représentation et une reconnaissance incrémentielle des événements de mouvement,
- * une reconnaissance incrémentielle et une verbalisation des plans, des intentions et des interactions de plans,
- * une description simultanée en langage naturel des figures de rhétorique dynamiques,
- * des descriptions incrémentielles et multimodales des itinéraires.

Les scénarios suivants sont à l'étude :

- * répondre à des questions sur les scènes de trafic observées,
- * générer des rapports de course pour de courtes sections de football,
- * décrire des itinéraires sur un modèle en trois dimensions du campus l'Université de la Sarre,
- * communiquer avec un robot mobile autonome.

2.3.2.1. CITY-TOUR

Le système CITYTOUR (André et al., 1985) est un système allemand de questions-réponses qui simule les aspects d'un tourisme fictif à travers une partie intéressante d'une ville particulière. CITYTOUR s'intéresse aux objets dynamiques et statiques projetés sur une surface bidimensionnelle. Dans ce système ont été implémentées des relations indépendantes de la position de l'observateur, de même que des relations pour lesquelles la position de l'observateur doit être prise en considération. Le point de

référence est donné soit par la position de l'observateur, soit par la face saillante de l'objet de référence.

Dans CITYTOUR, le centre d'intérêts des recherches se porte sur la reconnaissance et la verbalisation des relations spatiales statiques et dynamiques entre les objets dans une scène à temps variable. Un autre aspect intéressant, par rapport à la verbalisation des informations visuelles, est la reconnaissance des concepts de mouvement à un niveau supérieur et leur description en langue naturelle. Cet aspect est traité dans une seconde partie du projet VITRA, appelé SOCCER.

2.3.2.2. SOCCER

Quelques caractéristiques importantes différencient SOCCER de CITY-TOUR :

- * en plus des relations spatiales entre des sujets statiques ou dynamiques et des objets de référence statiques, des relations spatiales entre deux ou plusieurs objets dynamiques seront également verbalisés,
- * à côté des relations spatiales, des relations temporelles entre des mouvements (telles que la simultanéité et la succession) seront aussi verbalisées,
- * alors que dans CITYTOUR, on ne s'intéresse qu'aux trajectoires complètes des objets en mouvement, dans SOCCER, les mouvements seront reconnus et verbalisés (au présent) alors qu'ils sont encore en train de se réaliser,
- * en plus des concepts de déplacement visuels, on utilise également, dans SOCCER, des concepts non visuels tels que les intentions et les plans pour sélectionner une description en langue naturelle adéquate des mouvements observés, c'est-à-dire que la même trajectoire sera appariée avec des descriptions de déplacement complètement différentes, qui dépendent de l'intention imputée à l'acteur.

2.3.3. Système robotique KAMRO

Le but de ce travail est de fournir un accès en langue naturelle au robot mobile autonome à deux bras KAMRO¹ (*Karlsruhe Autonomous Mobile RObot*) qui est capable d'effectuer des tâches d'assemblage complexes. Pour permettre à ce robot d'être autonome, un système de caméras est utilisé pour percevoir l'environnement pendant l'exécution des tâches. Comme les expressions en langue naturelle doivent être interprétées en fonction de l'environnement du robot, il faut baser le traitement sur une sémantique référentielle perceptuellement ancrée. Les collaborateurs du projet expliquent comment les descriptions verbales, d'une part, et l'environnement physique, informations visuelles et géométriques, d'autre part, doivent être reliés entre eux.

La communication concernant les aspects spatio-temporels de l'environnement joue un rôle important dans l'interaction homme-robot. Par conséquent, l'interface en langage naturel ne doit pas uniquement prendre en considération la relation entre des expressions d'un langage de représentation des connaissances et des expressions en

¹ E. STOPP, K.-P. GAPP, G. HERZOG, T. LÄNGLE, T.-C. LÜTH : « Utilizing Spatial Relations for Natural Language Access to an Autonomous Mobile Robot » in B. Nebel, L. Dreschler-Fischer (eds) : *Advances in Artificial Intelligence*, Springer Verlag, Berlin / Heidelberg, 1994, pp. 39-50.

langage naturel, mais doit aussi reposer sur une définition de la sémantique référentielle perceptuellement ancrée.

2.3.3.1. ÉVALUATION DES RELATIONS SPATIALES

La classe des expressions de localisation se compose des prépositions dans leur sens spatial combinées avec des descriptions de lieux, d'un objet localisé (LO) et d'un objet de référence (REFO). La définition et la représentation de la sémantique des relations spatiales sont une condition essentielle pour la synthèse des expressions de référence spatiales en langage naturel. On peut définir les relations spatiales en spécifiant les conditions pour les configurations d'objets, comme la distance entre les objets, leur position relative...

2.3.3.2. INFORMATISATION DES RELATIONS SPATIALES ELEMENTAIRES

L'information de perception utilisable par le robot est encodée dans un modèle géométrique tridimensionnel accessible par un composant en langue naturelle. La signification des relations spatiales est définie par rapport aux représentations géométriques. Selon Landau et Jackendoff (1993), un algorithme approximatif est suffisant pour informatiser les relations spatiales car on ne tient pas compte des détails, on ne s'intéresse qu'aux propriétés physiques essentielles.

Trois classes de relations spatiales statiques sont prises en compte : les relations topologiques (*à* et *près de*), les relations projectives (*devant*, *derrière*, *à droite de*, *à gauche de*, *au-dessus de*, *en-dessous de* et *à côté de*), la relation *entre*, qui prend une place à part dans les relations spatiales.

2.3.3.3. L'UTILISATION COMPOSITIONNELLE DES PREPOSITIONS SPATIALES

Les relations prises en compte sont les relations topologiques *à* et *près de*, et les relations projectives *devant*, *derrière*, *à droite de*, *à gauche de*, *au-dessus de* et *en-dessous de*. Le tableau 1 donne un aperçu des compositions à deux places possibles des prépositions spatiales (en allemand). D'après ce tableau, les prépositions *devant*, *derrière*, *à droite* et *à gauche* sont réflexives dans leur utilisation compositionnelle, mais *à*, *près de*, *au-dessus* et *en-dessous* ne le sont pas, ces dernières occupant toujours la seconde position dans la composition.

Prépositions Spatiales	à, près de	devant, derrière	à droite, à gauche	au-dessus, en-dessous
à, près de	-	-	-	-
devant, derrière	X	-	X	X
à droite, à gauche	X	X	-	-
au-dessus, en dessous	-	-	-	-

2.3.3.4. L'IMPLEMENTATION DES RELATIONS SPATIALES COMPOSITIONNELLES

Pour l'implémentation des relations spatiales compositionnelles, les collaborateurs du projet n'ont pas besoin de savoir quelle relation occupe la première position. Ils définissent deux classes de relations spatiales compositionnelles à deux places : le

composé à deux places projectif projectif ($\text{Rel}_{\text{cpp}}^{2p}$), une composition de deux relations projectives élémentaires, et le composé à deux places projectif topologique ($\text{Rel}_{\text{cpt}}^{2p}$), qui est une combinaison d'une relation projective et d'une relation topologique.

L'applicabilité de ($\text{Rel}_{\text{cpp}}^{2p}$) et de ($\text{Rel}_{\text{cpt}}^{2p}$) nécessite une marque positive des deux relations spatiales élémentaires impliquées. Si on prend en compte la structure générale de la région d'applicabilité, il apparaît évident de devoir faire une importante différence entre les deux classes de compositions. Dans le cas de ($\text{Rel}_{\text{cpp}}^{2p}$) la combinaison des deux degrés d'applicabilité offre un degré d'applicabilité supérieur pour la composition, i.e., les deux relations élémentaires s'intensifient mutuellement. ($\text{Rel}_{\text{cpt}}^{2p}$) repose sur deux relations élémentaires indépendantes l'une de l'autre. La composition peut être vue comme une conjonction de deux conditions élémentaires, i.e., la faiblesse d'une relation ne peut pas être compensée par l'autre.

Pour l'implémentation de ($\text{Rel}_{\text{cpp}}^{2p}$) ils utilisent le calcul de combinaison des marques suivant (cf. Wahlster (1981)), qui garantit un degré d'applicabilité qui ne sera pas supérieur à 1 :

$$EV_{1\wedge 2} := EV_1 + EV_2 - EV_1 \cdot EV_2 \quad EV_1, EV_2 > 0$$

Le degré d'applicabilité $DA_{\text{Rel}_{\text{cpp}}^{2p}}$ peut être défini de la façon suivante :

$$DA_{\text{Rel}_{\text{cpp}}} := \begin{cases} DA_{\text{Rel}_1} + DA_{\text{Rel}_2} - |DA_{\text{Rel}_1} - DA_{\text{Rel}_2}| & : DA_{\text{Rel}_1}, DA_{\text{Rel}_2} > 0 \\ 0 & : \text{else} \end{cases}$$

avec $\text{Rel}_1 \in \{\text{devant}, \text{\`a droite}, \text{au-dessus}\}$, $\text{Rel}_2 \in \{\text{derrière}, \text{\`a gauche}, \text{en-dessous}\}$

ou $\text{Rel}_1 \in \{\text{derrière}, \text{\`a gauche}, \text{en-dessous}\}$, $\text{Rel}_2 \in \{\text{devant}, \text{\`a droite}, \text{au-dessus}\}$

Pour ($\text{Rel}_{\text{cpt}}^{2p}$), le degré d'applicabilité, $DA_{\text{Rel}_{\text{cpt}}^{2p}}$, est déterminé par le degré minimal des relations simples :

$$DA_{\text{Rel}_{\text{cpt}}} := \text{Min}(DA_{\text{Rel}_1}, DA_{\text{Rel}_2})$$

avec $\text{Rel}_1 \in \{\text{devant}, \text{derrière}, \text{\`a droite}, \text{\`a gauche}\}$, $\text{Rel}_2 \in \{\text{\`a}, \text{près de}\}$

2.3.3.5. L'INTERPRÉTATION DES T-UPLETS DE RELATION

Dans de nombreux cas, on peut penser que l'interprétation d'une proposition contenant une relation spatiale vérifie le t-uplet \diamond relationnel donné par rapport au modèle géométrique. Selon les arguments, des instantiations possibles du t-uplet relationnel sont implémentées et l'interprétation la plus plausible est sélectionnée. Cependant, il arrive parfois que ce type de demande ne soit pas possible, par exemple, si l'opérateur a fait référence à une partie de l'environnement que le robot ne peut pas percevoir de façon courante. Par conséquent, une sémantique référentielle pour les relations spatiales reste incomplète, si elle définit uniquement la façon dont peut être implémenté un t-uplet relationnel à partir des données sensorielles sous-jacentes.

Schirra et Stopp (1993) montrent comment un modèle géométrique cohérent peut être généré à partir d'un ensemble de propositions. Une interprétation adéquate est construite en recherchant la représentation la plus typique de la situation décrite par les propositions données. La distribution de typicalité correspondant à une certaine proposition est encodée dans le *Typicality Potential Field*, *champ potentiel de typicalité* (TyPoF), qui est une fonction qui apparie des localisations à des valeurs de typicalité. TyPoFs sont des exemples de schémas de typicalité associé à des relations spatiales. Chaque TyPoF prend en compte la dimensionnalité, la taille, et la forme des objets impliqués. Une valeur de typicalité associée à une expression spatiale correspond au degré d'applicabilité d'une relation spatiale pour une configuration d'objet donnée.

2.3.3.6. VERS UN DIALOGUE EN LANGUE NATURELLE

Un simple processus frontal en langage naturel ne permet pas assez de flexibilité pour communiquer avec un robot autonome situé dans un environnement dynamique complexe. Par conséquent, l'interface en langue naturelle devrait avoir accès à toutes les informations traitées et à la représentation de l'environnement. (cf. Lüth et al. (1994)).

Les commandes et les questions en langue naturelle de l'utilisateur représentent l'input pour le système d'accès en langue naturelle. L'analyse linguistique est menée par un parser syntactico-sémantique et traduit les expressions en langue naturelle en propositions. Le parser utilisé est une version modifiée de *Sb-Patr* (Harbusch (1986)) qui est basée une grammaire d'unification avec des informations sémantiques. Les propositions sont interprétées dans le composant d'évaluation, responsable de l'interprétation sémantique de référence. Il réalise l'interface avec le robot et les sources de connaissances partagées. Il est également responsable des réactions et réponses appropriées du système. Le composant de génération est responsable de la traduction de propositions sélectionnées en descriptions, explications et demandes en langue naturelle. Un générateur incrémentiel, basé sur les grammaires d'arbres adjoints (cf. Harbusch et al. (1991)), génère les structures de surface.

2.3.3.7. INTERPRÉTATION DES EXPRESSIONS DE LOCALISATION

En général, une proposition correspondant à l'analyse linguistique d'une expression de localisation ne peut pas fournir un t-uplet relationnel complètement instancié. Considérons une phrase du type *L'espace sur la gauche*, qui est employée pour dénoter un objet spécifique. La relation spatiale (*gauche*) et la restriction de type pour LO (*espace*) sont spécifiées. Ni le REFO, ni le cadre de référence ne sont donnés. Un système de référence est nécessaire puisque les relations projectives, de même que les prépositions correspondantes, dépendent de l'orientation. On peut distinguer les utilisations suivantes (cf. André et al. (1987), Retz-Schmidt (1988)) :

- * Utilisation intrinsèque : l'orientation est donnée par l'organisation inhérente au REFO, par exemple, par les organes de perception (humains, animaux), par la direction caractéristique du mouvement (véhicules, etc.), ou par les propriétés fonctionnelles (position des manipulateurs du robot, etc.). Suivant que l'on considère voir le REFO de l'extérieur (par ex. immeubles, bureaux, miroirs, etc.), ou de l'intérieur (par ex. chaises, vêtement, etc.) le vecteur arrière-avant

et le vecteur gauche-droite forment soit un système orthogonal droitier (*principe du miroir*), ou un système gaucher (*principe de coïncidence*).

- * Utilisation extrinsèque : l'orientation est déterminée par la position d'un observateur imaginaire possible, c'est-à-dire qu'elle est donnée par des facteurs contextuels, tels que l'accessibilité des objets de référence ou les objets situés dans son voisinage (par ex. *en regardant par la fenêtre, le levier est derrière le réservoir*). Si le REFO et l'observateur coïncident, l'orientation de l'observateur est transférée au REFO. Si l'observateur et le REFO sont spatialement séparés, l'orientation suit le principe du miroir.
- * Utilisation déictique : un cas spécifique d'utilisation extrinsèque où l'observateur coïncide avec le locuteur ou l'auditeur.

Dans l'exemple présenté, l'utilisation déictique semble être plausible et le robot joue le rôle de l'observateur et celui du REFO. Selon le contexte une interprétation différente de l'expression peut être plus appropriée. Pendant l'analyse référentielle, un type de recherche visuelle est mené afin de générer et de tester les candidats pour les arguments non instanciés du t-uplet relationnel. L'instanciation la plus plausible, par rapport au degré d'applicabilité et aux valeurs de champ dans la mémoire de dialogue, sera considérée comme l'interprétation correcte.

Si le LO et le REFO ne sont pas dans le champ visuel du robot (par ex. *le levier derrière vous*) il faut construire et ajouter une imagination plausible à la représentation de l'environnement. Des révisions ou ajustements pourront être nécessaires ultérieurement si l'interprétation est faussée par des données sensorielles. Ce mécanisme de génération d'une imagination est également requis si une expression fait référence à une situation hypothétique ou future (par ex. *mettez le levier derrière vous*).

2.3.3.8. SYNTHÈSE SUR LES EXPRESSIONS DE RÉFÉRENCE SPATIALES

Pour la synthèse et pour l'analyse d'une expression de localisation, il faut prendre en compte des aspects similaires :

- * Sélection des objets de référence : les critères suivants guident la sélection :
 - Distance entre le LO et le REFO, i.e., préfère les objets qui sont plus près de LO.
 - Saillance du REFO, dépend de facteurs tels que la forme, la taille, la couleur, etc.
 - Le contexte linguistique préfère les objets qui ont été mentionnés antérieurement.
- * Sélection des relations spatiales : on prend en compte non seulement le degré d'applicabilité d'un t-uplet relationnel, mais aussi la taille de la région d'applicabilité correspondante, par exemple une relation projective composée permettra une référence spatiale plus spécifique qu'une relation topologique binaire.
- * Sélection du cadre de référence : dans le cas de relations projectives, le système doit choisir une orientation. Cette sélection dépend des propriétés du REFO et du contexte situationnel.

2.3.3.9. RESUME

Dans ce travail, les auteurs se sont centrés sur le problème de la référence spatiale. Ils proposent d'utiliser les relations spatiales comme une représentation intermédiaire entre le niveau linguistique et le niveau sensoriel. Ces structures conceptuelles comblent le fossé entre les données géométriques et les concepts en langue naturelle, tels que les prépositions spatiales. Le modèle informatique est basé sur une graduation continue de l'applicabilité d'une relation spatiale et soutient à la fois, la synthèse et l'analyse de propositions contenant des relations spatiales. L'implémentation actuelle s'occupe des relations projectives et topologiques élémentaires et des relations spatiales compositionnelles. Les aspects importants de l'analyse et de la génération des expressions de localisation, tels que la détermination du cadre de référence et la sélection des objets de référence ont été considérés. Les idées présentées ici, doivent cependant encore être intégrées dans un prototype totalement utilisable d'une interface en langue naturelle pour le robot mobile autonome Kamro.

2.3.4. Les autres systèmes

Des versions simplifiées de relations spatiales statiques ont déjà été implémentées dans des systèmes tels que SWYSS (Hußmann & Schefe, 1984), LANDSCAN (Bajcsy et al., 1985), et des scènes à temps variable ont été traitées dans des systèmes tels que NAOS (Neumann, 1984), HAM-RPM (von Hahn et al., 1980). Un modèle computationnel de mémorisation pour les relations spatiales a été étudié dans le système MERCATOR (Davis, 1984).

2.4. CONCLUSION SUR LES PREPOSITIONS SPATIALES

On peut étendre le sens des prépositions spatiales à d'autres domaines tels que le temps, l'état, la manière, la cause, la circonstance... Le sens premier est dérivé et passe par des degrés d'abstraction croissante. On peut conclure que les langues ont conceptualisé les liens entre les domaines spatiaux et mentaux de l'expérience. A la base, on retrouve la conceptualisation de l'espace physique.¹

Scaliger, Madvig, Bréal et un grand nombre d'autres savants ont considéré que la signification fondamentale des prépositions était purement locale et que le sens temporel et tous les sens « modaux » en étaient dérivés par emploi « figuré ».

« Il ne fait pas de doute que l'image sensorielle est toujours la première à se présenter, qu'elle est souvent indispensable comme point de départ de la pensée, et que, dans les cas simples, il est commode de s'en servir comme illustration. Cependant il ne s'ensuit nullement que les assises de la langue soient des images des sens, spatiales ou autres ». ²

Les relations spatiales fournissent également de multiples expressions métaphoriques : *il est dans le cirage, il est dans les bras de Morphée*.³ Ces exemples

¹ Cf. *op. cit.* [Zelinsky93], pp. 76-96.

² Brøndal 1950, pp. 7-8 cité dans *op. cit.* [Cervoni91], p. 75.

³ Cf. F. LETOUBLON : « Sur les genoux des Dieux », in *Sur les compléments circonstanciels*, Langue Française 86, pp. 22-29, Larousse, Paris, 1990, p. 28.

sont traités dans notre étude s'ils s'intègrent dans les règles générales, sinon nous les avons laissés de côté, car ils nécessitent une traduction au cas par cas.

On peut noter des similitudes, concernant les prépositions, dans les deux langues :

- * au niveau morphologique, on constate que l'on peut regrouper les prépositions en groupes ou en classes très semblables : celles qui peuvent être dérivées à partir de bases nominales, adjectivales..., celles qui peuvent être composées par combinaisons de mots...,
- * toujours au niveau morphologique, il peut y avoir, dans certains cas, contraction de la préposition et de l'article,
- * au niveau syntaxique, les prépositions sont suivies d'un régime qui peut être de différentes natures : nom, pronom ... mais également d'un régime implicite (dans ce cas le régime n'est pas exprimé),
- * le choix de telle ou de telle préposition est souvent déterminé par les éléments situés à sa droite (on prendra en considération la valeur sémantique de cet élément mais on tiendra également compte de la présence d'un déterminant), mais il faut prendre en compte le verbe qui la précède et qui a un rôle important.

Il existe également des différences :

- * au niveau morphologique, s'il peut y avoir contraction entre la préposition et l'article, en français la contraction ne pourra se faire qu'avec des articles masculins ou pluriels, alors qu'en allemand, elle sera possible avec les 3 genres au singulier mais impossible au pluriel,
- * au niveau syntaxique, le français ne possède que des prépositions alors que l'allemand utilise également des postpositions et des circompositions,
- * en français, on a tendance à étudier les prépositions plus ou moins indépendamment les unes des autres alors qu'en allemand elles sont plutôt analysées en systèmes.

3. GENERALITES SUR LES MODELES

3.1. LES CAS

3.1.1. Généralités

Le français est une langue romane, c'est-à-dire qu'il est issu de l'évolution du latin. À l'origine, le latin était parlé à Rome puis s'est répandu essentiellement dans la partie occidentale de l'empire romain. Cette diffusion s'est faite, non pas tellement par les écoles, mais par les relations de personnes, et c'est ainsi que s'est développé le latin vulgaire, c'est-à-dire la langue parlée, développement d'autant plus libre que la force centralisatrice de Rome diminuait. Puis les invasions germaniques ont détruit l'unité romaine, et ont fait disparaître le latin, là où les Germains se sont établis en plus grand nombre (Belgique flamande, Alsace, Nord de la Suisse, Grande-Bretagne).

Vers l'an 800 (ancien français), le latin du nord de la Gaule a pris des caractères assez particuliers pour qu'il ne puisse plus se confondre avec le latin véritable, que la réforme des études à l'époque de Charlemagne avait d'ailleurs restitué comme langue de culture. D'importantes évolutions phonétiques se sont produites, séparant ainsi le français des autres langues romanes. Au niveau morphologique, l'ancien français comportait une déclinaison à deux cas : le cas sujet (pour le sujet et l'attribut), continuant le nominatif latin, et le cas régime (pour tous les compléments), continuant l'accusatif latin.

En moyen français (du milieu du XIV^e siècle à la fin du XVI^e), le cas sujet disparaît. Cette disparition est sans doute due au fait que l'ordre des mots est moins libre qu'en ancien français, la place normale du sujet est alors devant le verbe. D'autres évolutions ont également eu lieu : le pronom personnel sujet et l'article deviennent obligatoires, l'article partitif apparaît...

Entre le XVII^e et le XX^e siècle, la phonétique et la morphologie n'évoluent plus de façon notable ; le lexique s'enrichit, et le français devient, à partir du XIX^e siècle, la langue de la majorité. Le cas régime disparaît à son tour.¹

Dans des langues telles que le français, l'anglais c'est principalement l'ordre des constituants qui indique le rôle qu'ils jouent dans la phrase. Dans des langues telles que le latin, l'allemand, le russe, il existe un lien entre la forme morphologique d'un mot et son rôle dans la phrase : c'est la notion de cas de surface. Attention, il n'y a cependant pas toujours correspondance exacte entre les sens et les cas syntaxiques (un cas peut traduire différents sens : l'accusatif pour le COD et le lieu de destination en latin, des cas distincts peuvent exprimer des relations similaires : le gérondif au génitif accompagné d'un nom complément d'objet à l'accusatif peut être remplacé par l'adjectif verbal et le nom, tous deux au génitif, en latin). De même, il n'y a pas toujours correspondance entre les différentes langues : l'allemand et le latin n'ont pas le même

¹ Cf. *op. cit.* [Grévisse93], pp. 9-14.

nombre de cas, le datif ne recouvre pas les mêmes choses.¹ Dans certaines langues indo-européennes (notamment le russe), un même substantif peut, pour deux formes casuelles, avoir la même terminaison ; on distinguera ces formes uniquement par leur accentuation. En finnois, il existe plusieurs cas dit locaux :

INTERIEUR	Inessif	talossa	dans la maison
	Elatif	talosta	de (l'intérieur de) la maison (mouvement)
	Illatif	taloon	dans la maison (mouvement)
EXTERIEUR	Adessif	talolla	à la maison (ou près de)
	Ablatif	talolta	de (l'extérieur de) la maison (mouvement)
	Allatif	talolle	vers la maison (mouvement)
GENERAL	Essif	talona	
	Partitif	taloa	
	Translatif	taloski	

TABLEAU 10 : LES CAS DITS LOCAUX EN FINNOIS²

Un trait important de la théorie du cas (en latin et dans les langues indo-européennes) est qu'il n'y a jamais de distinction entre le nominatif et l'accusatif des noms neutres.

latin :	nominatif :	forum	russe :	nominatif :	поле
	accusatif :	forum		accusatif :	поле

Ceci conduit à penser que la catégorie du cas et la catégorie du genre sont interdépendantes dans ces langues. Dans la mesure où le genre a un fondement naturel, sexe et animé, cela implique deux distinctions : animé vs inanimé, féminin vs non féminin. La différence entre le masculin et le neutre au nominatif peut être interprétée comme une conséquence de l'importance plus grande de la distinction entre sujet et objet en ce qui concerne les noms animés. Il y a relativement peu de noms animés neutres. Il y a un rapport étroit entre le caractère animé et la position sujet.³

La casualisation (attribution de cas) situe une sémie par rapport à une catégorie casuelle, soit par actualisation d'un trait casuel, soit par propagation d'un tel trait. Les propagations casuelles diffèrent selon les langues (positions distributionnelles, marques ou flexions casuelles de surface...), il en est de même pour les sèmes inhérents : en guarani par exemple, les noms de lieu sont obligatoirement au locatif, tandis qu'en français ils peuvent parfaitement se voir afférer l'ergatif : *Les chutes du Niagara ont bouleversé la jeune mariée*. Des sèmes afférents socialement normés peuvent avoir une

¹ G. SABAH : *L'intelligence artificielle et le langage, Représentations des connaissances*, vol. 1, Hermès, Paris, 1988, 2ème édition, 1990, pp. 90-91.

² Ce tableau est extrait de D. CREISSELS : *Éléments de syntaxe générale*, Presses Universitaires de France, Paris, 1995, pp. 188.

³ J. LYONS : *Linguistique générale, Introduction à la linguistique théorique*, Larousse, Paris, 1970, pp. 226-227.

prédisposition casuelle : un syntagme ayant pour trait /humain/ sera rarement au locatif ou à l'instrumental mais plutôt à l'ergatif. Certains traits casuels sont associés à des traits génériques dimensionnels : *des taches s'étalent sur son smoking*, 'des taches' prend le trait /ergatif/, et son trait /inanimé/ devient latent, voire virtualisé.¹ La primitive casuelle locatif (LOC) peut être spécifiée selon qu'il y a contact ou non contact, position supérieure ou inférieure relativement à un repère (cf. *sur* vs *sous*, *au-dessus* vs *au-dessous*). L'ergatif peut être spécifié selon qu'il commande un procès qui comporte ou non une transmission (cas destinataire par opposition au datif ou destinataire et au bénéfactif).

Il n'est généralement pas possible de substituer une flexion casuelle (ou une préposition) abstraite à une autre, sans opérer de changements ailleurs dans la phrase. La différence entre la variation flexionnelle et l'emploi des prépositions n'est qu'une différence assez peu importante de structure de surface.²

Beaucoup de linguistes ont eu besoin d'élaborer une théorie générale afin de répartir en deux types les marques casuelles : distinction entre cas « directs » et cas « obliques », ou cas « syntaxiques » et cas « sémantiques ».

3.1.2. La grammaire des cas

3.1.2.1. BRØNDAL

Brøndal souligne l'importance de la différence qu'il y a entre des mots autonomes qui « peuvent remplir plusieurs fonctions syntaxiques » et des « formes » qui ne sont que des « modifications de mots » (*Théorie des prépositions*, p. 11). Il fait une distinction entre plusieurs cas : cas proches parents de la notion générale de préposition (locatif, génitif, accusatif), cas n'ayant qu'une parenté indirecte (datif, nominatif) et cas qui ne présentent avec les prépositions qu'un rapport de parenté très éloigné (ablatif, instrumental, vocatif), (*Théorie des prépositions*, p. 16).³

3.1.2.2. GUILLAUME

Guillaume ne fait pas la même utilisation de la notion de cas que les « grammaires casuelles ». Un cas n'est pas un fait indépendant de la morpho-syntaxe. L'étude des cas se rattache « au procès de grammaticalisation du mot » (Guillaume 1964, pp. 108-109), procès dont l'arrêt plus ou moins précoce donne à la morphologie de chaque langue sa physionomie particulière. À un arrêt précoce correspond une morphologie légère comme le nom en français ou en anglais (le nom anglais ne prend pas la marque du genre (Guillaume 1964, p. 111)), à un arrêt plus tardif une morphologie plus chargée (en latin, au cours de sa phase de grammaticalisation, un mot reçoit un genre, un nombre et un des cas de fonction = nominatif, accusatif...).⁴

¹ F. RASTIER, M. CAVAZZA, A. ABEILLE : *Sémantique pour l'analyse, De la linguistique à l'informatique*, Masson, Paris / Milan / Barcelon, 1994, pp. 134, 138.

² Cf. dans *op. cit.* [Lyons70], pp. 232-233.

³ Cité dans *op. cit.* [Cervoni91], p. 93.

⁴ *Ibid.*, p. 185.

« La déclinaison a longtemps joué un rôle capital : morphème à double effet, elle fixe la « partie du discours » et assigne prévisionnellement au sémantème des limites d'emploi (nominatif : sujet et attribut, accusatif : complément d'objet...) (...) La déclinaison est encore suppléée par des morphèmes à simple effet, exclusivement chargés d'exprimer une certaine fonction dans la phrase : les prépositions et l'ordre des mots, dont le registre, à l'examen, ne se révèle pas identique ».

La désinence dite « casuelle » est donc appelée par G. Guillaume morphème à double effet : d'une part elle est la marque du cas d'emploi (terme plus général que cas de fonction) ; d'autre part, elle joue dans le mot le rôle de déterminant de la partie du discours (Guillaume 1964, p. 103). En cela, elle s'oppose à la préposition :

« quant à la préposition, elle est un morphème à simple effet. Elle sert à exprimer une certaine fonction du nom dans la phrase, mais n'intervient à aucun degré comme déterminant de la partie du discours, obtenue dans le mot sans que la préposition ait eu à se produire » (*ibid.*).¹

La morphologie intérieure, « inséparable du mot » est moins complète en français qu'en anglais qu'en latin. Pour jouer les mêmes rôles fonctionnels, elle doit se compléter et fait appel à une « morphologie grammaticale extérieure séparable » avec, notamment, un recours aux prépositions. Cela ne signifie pas que le français ne possède pas de cas, il possède un cas synthétique :

« La langue française, qui n'a pas de cas apparent, est une langue qui a construit, en pensée, un cas nouveau – celui que nous nommons, dans notre enseignement, le cas synthétique du français – lequel cas, en langue, totalise en les indiscrétant, les fonctions adversatives de sujet et d'objet, et la fonction moyenne d'attribut, alors qu'en discours, et sans l'aide d'aucun signe qui soit un mot, les fonctions indiscrétées en langue se séparent » (Guillaume 1974, p. 34).

Il y a une séparation qui se produit dans la transition de la langue au discours, il faut opérer un choix entre les trois fonctions puisque un mot ne peut être en même temps sujet et objet, sujet et attribut ou objet et attribut.

« Il n'existe aucune possibilité de neutralisation de ces fonctions. Le cas synthétique « se dénoue », « s'ouvre » pour « livrer une partie de son contenu » ; « il se résout (...) en un cas analytique ». Faute de cette résolution, il est inemployable ».²

La préposition mais aussi les lexèmes qu'elle relie participent à l'expression du cas profond. En fait, la préposition a un rôle d'adjuvant : elle n'apparaît que si le contenu lexical des éléments n'est pas suffisant pour exprimer les rapports casuels présents dans les schèmes d'entendement (on retrouve la base de ce point de vue dans la syntaxe génétique selon laquelle l'intervention de la préposition est tardive). La lexigénèse des mots est importante : dans *Félicité pleure son perroquet*, la mise en contact de *pleurer* et de *perroquet* suffit à exprimer le cas causal ; mais dans *Félicité pleure à cause d'une piqûre*, piqûre est non animée, on a donc besoin d'une préposition ; dans les deux énoncés le contenu lexigénétique est différent : dans le premier *pleurer* signifie « déplorer la perte de quelqu'un », dans le second, c'est un sens subduit par rapport à « verser des larmes ».³

3.1.2.3. FILLMORE

Fillmore (1968-1969) décrit les prépositions, dans le cadre de la Grammaire des cas, comme des variantes d'unités non flexionnelles. Il distingue trois types de préposition :

¹ Cf. *op. cit.* [Cervoni91], p. 186.

² *Ibid.*, p. 186.

³ *Ibid.*, p. 211.

celles avec un contenu lexical, celles qui dépendent du verbe, et celles qui expriment des relations de cas sous-jacentes.

Selon Fillmore (1968),

« on peut identifier un ensemble de cas sémantiques permettant de mettre en évidence, à la manière des cas syntaxiques, les relations de sens qui existent entre les noms et le verbe dans une phrase simple ».

Pour Fillmore, la représentation profonde d'une phrase se compose d'une modalité¹ et d'une proposition² ou, plus simplement, d'un verbe relié à un groupe syntaxique par des relations sémantiques (les cas).

Selon lui, la structure profonde doit être conçue comme contenant déjà des indications sur le rôle sémantique que jouent les syntagmes nominaux dans les phrases ». Pour que celle-ci puisse contenir des indications telles que valeur d'agent, d'instrument, de datif, de lieu..., Fillmore propose de la concevoir comme un ensemble comprenant, associée à un constituant « Modalité », une « Proposition » constituée d'un verbe et d'un ou plusieurs noms, caractérisés par la relation ou « cas profond » qui les rattache au verbe.

Dans son livre page 24, Fillmore a fait une première proposition de liste de cas : agentif, instrumental, datif, factitif, locatif, objectif.³

- agentif l'instigateur animé d'une action identifiée par le verbe (sauf dans le cas de forces naturelles (vent, mer...) ou d'appareils mécaniques)
- instrumental la force inanimée ou l'objet affecté
- datif l'animé affecté par l'action
- factitif l'objet résultant de l'action
- locatif le lieu ou l'orientation
- objectif le reste ; c'est le cas sémantique le plus neutre. Il ne faut pas confondre ce terme avec la notion d'Objet Direct ou avec le nom du cas de surface synonyme de l'accusatif.⁴

Ex. : Jean ouvre la porte avec sa clef
 agent objet instrument

La Bretagne est pluvieuse
 lieu

En 1971, dans l'article intitulé « Quelques problèmes posés à la grammaire casuelle » la liste des cas dont Fillmore se dit « content », bien qu'il reconnaisse que ses critères « sont relativement vagues » et « donnent rarement des résultats parfaitement nets » (*op. cit.*, p. 70) est sensiblement différente de celle de *The case for case*. On a maintenant 8

¹ Information sur la négation, le temps, le mode, l'aspect.

² Structure indépendante du temps qui permet l'identification du verbe.

³ Cas cités dans *op. cit.* [Cervoni91], pp. 173-175.

⁴ Cf. G. SABAH : *L'intelligence artificielle et le langage, Représentations des connaissances*, vol. 1, Paris, 1990, p. 92 et J.-T. PLATT : *Grammatical Form and Grammatical Meaning, a tagmemic view of Fillmore's deep structure case concepts*, North-Holland Publishing Company Amsterdam, London, 1971, p. 19.

cas : agent, expérier (qui remplace le datif pour les verbes d'événement psychologique ou d'état mental), instrument, objet, source, but (qui remplace le factitif), lieu, temps. ¹

- agent	l'instigateur d'un événement
- contre-agent	la force contre laquelle l'action est exécutée
- objet	l'entité qui bouge, change, dont la position ou l'existence est en question
- résultat	l'entité créée par l'action en question
- instrument	le stimulus ou la cause physique causant l'événement
- source	le lieu de départ de quelque chose qui bouge
- but	le lieu d'arrivée de quelque chose qui bouge
- patient	l'entité qui reçoit, accepte ou subit les effets d'une action

Ex : Jean a conduit son bateau, malgré le courant du large vers la plage
 agent objet contre-agent source but

Le courant a fait chavirer le bateau que Jean a fabriqué avec du chêne
 instrument objet resultat agent patient

La première ligne concerne le verbe «chavirer» et la seconde le verbe «fabriquer».

Il y a deux grandes idées importantes dans sa théorie :

- * un cas n'a qu'une réalisation (dans une phrase donnée, pour un verbe donné un seul nom peut être lié à ce verbe par un cas sémantique donné)
- * on peut attacher une ossature (*case-frame*) à un verbe, c'est-à-dire indiquer pour ses divers arguments le type du lien qui le lie à tel ou tel argument.

Des constituants ayant des rôles syntaxiques identiques peuvent avoir des rôles sémantiques différents :

Jean casse la branche avec une pierre

La pierre casse la branche

La branche casse

ex. : casser [(Agent), (Instrument), Objet] : a priori 3 cas sémantiques sont possibles, mais seul l'objet est obligatoire

Le système des cas choisi a des incidences importantes sur les phénomènes d'ambiguïté.

A quoi servent les cas ? ²

- * l'ossature explicite une hiérarchie linéaire simple entre les cas sémantiques qui sont ordonnés. L'instrument ne peut être sujet que si l'agent est absent, et l'objet que si l'agent et l'instrument sont absents (*casser [(Agent), (Instrument), Objet]*),
- * si un verbe a plusieurs sens, il aura plusieurs ossatures (*voler [agent]*, *voler [objet]*, *voler [agent, objet]*). Grâce aux cas, on peut faire une distinction entre

¹ Cas cités dans *op. cit.* [Cervoni91], pp. 173-175.

² Cf. *op. cit.* [Sabah90], pp. 94-95.

des phrases qui ont la même structure (*Cette veste est chaude (instrument), cette journée est chaude (temps [période]), cette pièce est chaude (lieu), son front est chaud (patient)*),

- * une conjonction de coordination ne peut relier que des groupes nominaux jouant le même cas sémantique par rapport au verbe (*Jean et Paul cassent la branche, *Jean et la pierre cassent la branche*).

Mais qu'en est-il de toutes les prépositions ? Si les cas indiquent les fonctions sémantiques fondamentales des noms, comment l'appareil casuel intervient-il dans la détermination du choix des différentes prépositions ? Fillmore cherche « un moyen de soutenir que les prépositions qu'on aimerait intuitivement regrouper ensemble sont les marqueurs d'un même cas » (*ibid.*). L'emploi de prépositions différentes pour un même cas est imputable à des différences dans « la structure sous-jacente des SN correspondants » (*ibid.*), ce qui signifie que :

« les noms qui figurent dans les expressions locatives peuvent se voir imputer des propriétés comme celle d'être un point ou une surface ou un volume, ou celle d'être une partie d'une surface ou d'un volume, ou celle d'être un point ou une zone au-dessus ou au-dessous ou à l'arrière ou à l'avant ou à côté d'un certain objet, etc. » (*ibid.*).¹

La découverte des affinités qui peuvent exister entre les cas et les prépositions exige que l'on s'appuie non seulement sur des définitions précises des cas mais aussi des propositions aussi explicites que possible concernant le signifié des signes qui leur correspondent (les prépositions).

3.1.2.4. SIMMONS

Simmons (1974) s'intéresse, quant à lui, aux réseaux sémantiques. Ceux qu'il a développés comportent un ensemble de cas sémantiques profonds permettant de relier les concepts nominaux aux verbes.



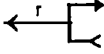
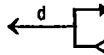


* acteur	l'animé qui effectue une action
* thème	ce qui reçoit l'effet de l'action
* source	l'endroit ou l'état au début de l'action
* but	l'endroit ou l'état à la fin de l'action
* instrument	un objet ou un processus qui contribue à l'action
* lieu ²	l'endroit où se situe l'action
* temps	le moment de l'action

3.1.2.5. SCHANK

Schank (dépendances conceptuelles, 1972) a développé un système dans lequel la représentation du sens est basée sur les relations conceptuelles (indépendantes de la langue) entre les objets et les actions. Il a établi onze primitives à partir desquelles on peut décrire toutes les actions. Les rôles conceptuels qui peuvent lier les primitives aux concepts (les signes ont été utilisés par Schank dans les schémas conceptuels) et les cas sémantiques de Schank sont les suivants :²

¹ Cité dans *op. cit.* [Cervoni91], pp. 198-199.

² Cas cités dans *op. cit.* [Sabah90], p. 102.

Acteur	l'animé qui produit une action	
Objet	ce sur quoi porte l'action	
Bénéficiaire	celui qui reçoit un objet (le donneur est aussi exprimé)	
Direction	le point de départ et le point d'arrivée d'un objet déplacé	
État	valeur d'un attribut	
Instrument	élément qui contribue à l'action	

La dépendance conceptuelle implique que deux phrases ayant une signification équivalente doivent avoir la même représentation interne, même si elles ont des structures syntaxiques très différentes, toute information implicite doit être explicite dans la représentation, toute action s'exprime en termes de primitives, chacune d'entre elles possédant un schéma associé qui devra être « instancié » et rempli lors du processus de compréhension.

On représente la signification d'une phrase grâce à un schéma de dépendances conceptuelles qui regroupent les catégories suivantes :

- * les PPs (*Picture Production*) : les noms,
- * les ACTs (*Action*) : les verbes ou groupes de verbes,
- * les PA (*Picture Aider*) : les modificateurs des PPs comme les adjectifs,
- * les AA (*Action Aider*) : les modificateurs des ACTs comme les adverbes.

Les dépendances peuvent être à double sens (le plus souvent entre un PP et un ACT), à sens unique entre un ACT et un PP ou entre un PP et un PA, à sens unique entre deux PPs.

La représentation en dépendances conceptuelles fait apparaître les relations causales, contrairement à la représentation en grammaire casuelle. Pour représenter la signification d'une phrase, il faut trouver les primitives qui correspondent aux actions et compléter la structure associée. Il en existe quinze, parmi lesquelles on peut noter :

- * ATRANS : elle désigne le transfert abstrait de quelque chose à quelqu'un (*donner*), de quelque chose à soi-même (*prendre*), le transfert simultané de plusieurs objets (*acheter*),
- * PTRANS : elle désigne le transfert physique de soi-même quelque part (*aller*), d'un objet quelque part (*mettre*),
- * SPEAK : elle indique une production de sons (*parler, chanter, siffler*),
- * ATTEND : elle indique le fait de consacrer un organe des sens à un stimulus (*écouter, voir*).

3.1.2.6. POTTIER

Pottier fait une distinction entre relation endocentrique¹ à laquelle il fait correspondre la voix attributive et exocentrique² à laquelle il fait correspondre la voix active. Le principe de la distinction est la puissance ou la non-puissance de l'entité. Chacune des deux voix se définit par le cas dont se revêtent les actants : voix attributive = nominatif (cas de l'actant auquel le comportement est attribué [la base du « schème »]), c'est le cas de la non-puissance mais aussi de la non-différence de potentiel entre les actants ; voix active = ergatif (= puissant ; cas de l'actant correspondant au support), accusatif (= non-puissant ; cas de l'actant contenu dans l'apport). Ces trois cas représentent le noyau qui structure l'ensemble du système. Sur l'axe d'actance (de part et d'autre du noyau) se trouvent deux zones : à l'avant du noyau, la « zone de puissance » (causal, agentif, instrumental), après le noyau, la « zone de non-puissance » (datif, bénéfactif, final). A côté de ce premier axe, il existe un axe de dépendance sur lequel se trouvent les entités n'ayant pas de rapport direct avec l'événement : le locatif, dont les applications, pour Pottier, sont spatiales, temporelles ou « notionnelles » ; le sociatif, cas marginal qui « permet d'associer un actant à n'importe quel autre » (Pottier 1976a, p. 133). Pottier regroupe onze cas en trois zones.

En ce qui concerne l'analyse sémantico-logique des lexèmes verbaux, Pottier propose une grille sémantique composée de six voix (l'existential, l'équatif, le situatif, le descriptif, le possessif, le subjectif) auxquelles peuvent s'appliquer trois statuts (statif, évolutif, causatif).

La fluidité entre les zones casuelles a un rôle important chez Pottier, elle est liée à deux autres notions : la dominance casuelle et le transfert casuel. La dominance casuelle est l'affinité spéciale d'un lexème avec un cas conceptuel : le contenu lexical du mot *clé* le prédispose au rôle d'instrumental (il en est de même pour quasiment tous les substantifs dont la définition pourrait commencer par « instrument, objet, qui sert à... »), les noms de personne ont une affinité avec l'ergatif.

On peut faire quelques observations :

- * elle est une réalité socio-culturelle d'étendue variable : l'affinité casuelle n'est pas toujours aussi nette qu'avec le mot *clé*. L'affinité casuelle peut varier en fonction de l'utilisation d'un mot dans un domaine particulier, de la catégorie d'utilisateur, du domaine d'expérience...
- * elle est une réalité d'ordre statistique, c'est-à-dire une réalité que des statistiques peuvent mettre en évidence. « Il se trouve que tous les exemples du mot *porte* dans le dictionnaire de Robert sont au nominatif ou à l'accusatif, et jamais à l'ergatif » (Pottier 1974, p. 119).
- * le terme « dominance » implique une idée de hiérarchie, il y a un ordre de dominance décroissant (*grue* est essentiellement un instrumental mais peut aussi être un ergatif (*la grue soulève la voiture*) ou accusatif (*le chef utilise la grue*)... On aura alors pour le mot *grue* : grue {ins., erg., acc., ...}).

¹ Relation orientée vers l'entité.

² Relation orientée vers le comportement.

La notion de transfert casuel complète celle de dominance :

« Le terme s'applique aux emplois où un élément joue, dans une phrase, un rôle sémantique qui ne correspond pas à sa ou à ses dominance(s) casuelle(s), par exemple à *clé*, dans *la clé a ouvert la porte* » (*Linguistique générale*, p. 119).

Le transfert consiste à donner à *clé* une place normalement réservée à un ergatif (idem dans *À midi on ne doit plus prendre son petit déjeuner*, Pottier commente :

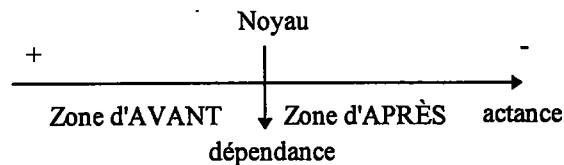
« Ainsi à *midi* est-il banalement un locatif (temporel). Mais on peut le faire fonctionner sur l'acte d'actance comme causal »).

Le nominatif correspond au non-événement, sans différence de potentiel : *Pierre est un homme*, *Pierre est heureux*, et l'ergatif/accusatif à l'événement avec différence de potentiel : *Pierre casse l'œuf*.

Ce sont les trois cas du noyau : ¹

O ~ NOM
O ⁺ ERG -> O ⁻ ACC

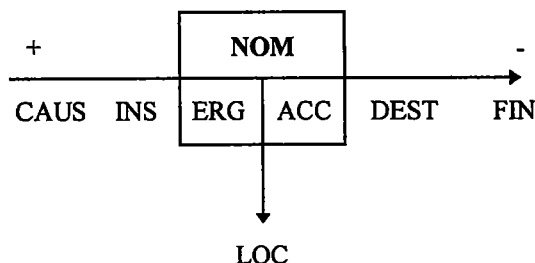
Les actants peuvent avoir un rôle sur l'axe d'actance ou sur l'axe de dépendance (position de marginalité ; cet axe correspond au cas conceptuel LOCATIF qui s'applique à trois champs : l'espace, le temps, le notionnel).



AVANT : zone de puissance (+) = cause, origine, condition, instrument, agent secondaire

APRÈS : zone de puissance (-) = destination, but, finalité, conséquence.

Les 8 cas conceptuels : ²



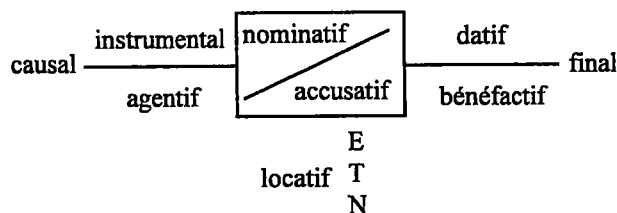
Le cas conceptuel appartient à la sémantique profonde, indépendante de la solution linguistique. Le causatif introduit toujours un actant puissant agissant sur un actant non-puissant, objet de cette action.

¹ B. POTTIER : *Théorie et analyse en linguistique*, Hachette Supérieur, Paris, 1992, p. 123.

² *Ibid.*, p. 125.

« Toutes les langues ont un nombre n de cas linguistiques, liés à une morphologie distinctive. La linguistique générale se préoccupe de mettre en relation ces cas linguistiques avec les zones et sous-zones casuo-conceptuelles ».

Le cas linguistique est issu de la prédication (sémantique seconde, ou dépendante). Les cas linguistiques du français sont : ¹

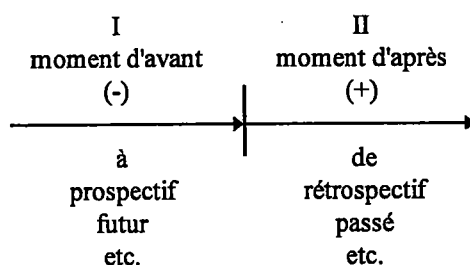


Il se peut qu'un même cas conceptuel puisse être manifesté à travers différents cas linguistiques : ERGATIF → nominatif, causatif, locatif, ACCUSATIF → accusatif, nominatif, BÉNÉFACTIF → datif, nominatif, locatif. ²

Chez Pottier, on retrouve des caractéristiques communes entre la représentation du système casuel et celle des micro-systèmes prépositionnels : orientations cinétiques, seuils séparateurs, positions d'avant et d'après. Il existe une correspondance entre certains cas et certaines prépositions : *par/pour* => cette opposition reflète deux groupes de cas : causal, instrumental, agentif et datif, bénéfactif, final (Pottier 1974, p. 132). ³

Sur l'axe d'actance, *pour* occupe la position d'avant (comme *à*), et *par* la position d'après (comme *de*) ; les cas causal, instrumental et agentif la position d'avant (ce sont des sous-classes de « l'ergatif »), les cas datif, bénéfactif et final la position d'après : l'avant du système casuel correspond à l'après du système prépositionnel et vice versa ; on peut trouver une explication dans la distinction que fait Pottier entre chronologie et chrono-expérience.

a) Schéma d'après la chronologie : ⁴



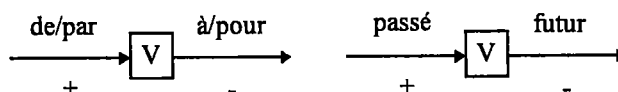
¹ Cf. *op. cit.* [Pottier92], p. 127.

² *Ibid.*, p. 128.

³ Cité dans *op. cit.* [Cervoni91], pp. 199-200.

⁴ Note 71 du livre de Cervoni : La « chrono-logie » est du même ordre que la « chronologie de raison », cette logique qui, selon Guillaume, tend à se projeter dans l'esprit en une sorte de chronologie abstraite.

b) Schéma d'après la chrono-expérience ; la vision est fixée par le locuteur, qui est pris comme point de repère, et le schéma s'inverse.



3.1.2.7. BOGURAEV/ SPARK-JONES

Ils se sont basés sur les travaux de Wood (1979) sur l'utilisation des prépositions. Leur analyse part de 421 phrases où figurent 86 prépositions, ils mettent 27 cas en évidence :

Accompagnement	Direction	Possesseur
Activité	État	Quantité
Agent	Force	Raison
Après	Instrument	Récipient
Attribut	Lieu	Source
Avant	Localisation (abstrait)	Source (abstrait)
But	Manière	Sujet
Destination	Objet	Temps (période)
Destination (abstrait)	Objet-mental	Temps (situation) ¹

3.1.2.8. GRIMES

Grimes (1972) veut obtenir un degré d'abstraction plus important en réunissant les cas qui regroupent des concepts voisins, ce qui lui permet de distinguer différents rôles (ce modèle s'applique à l'analyse des discours) :

* rôles d'orientation :	objet	la chose dont la position où le mouvement est décrit
	source	la position origine de l'objet
	but	la position finale de l'objet
	chemin	l'endroit traversé durant un mouvement
	véhicule	ce qui sert à transporter l'objet
* rôles de changement :	patient	l'élément qui change ou dont l'état est décrit
	matériel	l'élément qui change, avant son changement
	résultat	l'élément qui change, après son changement
	référent	le champ ou l'objet qui définit les limites de changement
* rôles spécifiques :	agent	le responsable d'une action
	instrument	l'outil utilisé pour réaliser une action
	force	la cause non volontaire d'une action
	bénéficiaire	la personne (ou l'objet) sur qui l'action a un effet secondaire.

¹ Cas présentés dans *op. cit.* [Sabah90], p. 103.

Grimes cherche à mettre en évidence une structure fondée sur les liens et les similarités entre ces différents rôles, les rôles de base figurent au centre du schéma : ¹

ORIENTATION		→	UNIFICATION	←	CHANGEMENT
AGENT INSTRUMENT			FORCE		
VEHICULE	→	VEHICULE			
OBJET	→	PATIENT	←	PATIENT	
SOURCE	→	PRECEDENT	←	MATERIEL	
BUT	→	SUIVANT	←	RESULTAT	
CHEMIN	→	CHEMIN	←	REFERENT	
BENEFICIAIRE					

3.1.2.9. FEUILLET

Feillet (1980) essaie de déterminer un ensemble de fonctions sémantiques profondes qui devraient être sous-jacentes à tout un ensemble de cas : les fonctions processives, les fonctions descriptives, les fonctions situatives. À partir de là, il fait ressortir 7 fonctions sémantiques profondes (statif, agentif, réceptif, objectif, médiatif, descriptif et situatif).

3.1.2.10. DE BOER

Selon de Boer il existe six rapports casuels syntaxiques : les cas du sujet, du régime, de la localisation, du point de départ, de la direction et des circonstances accompagnantes ; les six cas sont : nominatif, accusatif, locatif, ablatif, directif et instrumental-sociatif. Le datif allemand est remplacé par le directif : directif final (*travailler pour vivre*), directif d'intérêt, un supin directif (*donner à lire*), un directif de temps et de lieu (*remettre à demain, aller à Paris*), de même les emplois du génitif sont placés sous l'accusatif : il renferme l'accusatif de l'objet (*je le vois*), les « datifs de relation » introduits par *à* (*le chapeau à papa*), certains syntagmes introduits par « de » en position écrasée (*charge de cavalerie*), génitif de l'objet (*peur de la mort*), génitif de possession (*le livre de Paul*), apposition (*la ville de Paris*), matière (*table de bois*), partitif (*trois de ces hommes*), compléments de certains verbes ou adjectifs (*féliciter de, content de*), emplois de la préposition de « fin de mot » (*cela sert de table*).

3.1.2.11. TABLEAU RECAPITULATIF

(voir page suivante)

¹ Cas présentés dans *op. cit.* [Sabah90], p. 100.

Brondal	Fillmore	Simmons	Schank	Pottier	Grimmes	Feuillet	Boguraev Spark-Jones	De Boer
----------------	-----------------	----------------	---------------	----------------	----------------	-----------------	---------------------------------	----------------

<i>nominatif</i> ¹	agent	acteur	acteur	agentif	agent / force	<i>agentif</i>	agent	<i>nominatif</i>
	contre-agent							
<i>accusatif</i>	objet		objet		objet	<i>objectif</i>	objet	<i>accusatif</i>
	résultat							
<i>instrumental</i>	instrument	instrument	instrument	instrumental	instrument		instrument	<i>instrumental</i>
	source	source	direction		source		source	<i>ablatif</i>
<i>locatif</i>	but	but		final	but	but		<i>directif</i>
					chemin			
		lieu		locatif			lieu	<i>locatif</i>
		temps		L T N			temps	
<i>accusatif</i>	patient	thème			patient			
<i>datif</i>			bénéficiaire	bénéfactif	bénéficiare	<i>réceptif</i>		
					référent			

TABLEAU 11 : TABLEAU RECAPITULATIF SUR LES CAS

¹ Cas en italique = cas de fonction, cas en normal = cas profond

3.1.2.12. CONCLUSION

On peut noter une différence entre le point de vue de Fillmore et celui de Simmons : chez Simmons plusieurs groupes distincts d'une même phrase peuvent jouer le même rôle sémantique.

Il existe également quelques différences entre le modèle de Schank et celui des autres :

- * les cas sémantiques relient des entités conceptuelles (et non des éléments syntaxiques tels qu'un GN, ou un GP) à une des onze actions conceptuelles de base (et non aux verbes de surface),
- * un seul cas sémantique peut lier des entités différentes,
- * les cas liés à une action conceptuelle de base sont tous obligatoires, il faudra les retrouver par inférence s'ils ne sont pas présents dans une phrase.

En revanche, on constate des similitudes entre Grimmes et Feuillet, il y a une certaine convergence mais pas de recouvrement des deux conceptions :

FEUILLET	GRIMMES
Statif	Patient
Agentif	Agent et Force
Réceptif	Bénéficiaire
Objectif	Objet
Médiatif	Instrument et véhicule
Descriptif	Référent
Situatif	Source, but et chemin

3.1.3. Cas & prépositions

3.1.3.1. GENERALITES

Au départ, les prépositions n'existaient pas. La flexion nominale était assez riche pour exprimer à elle seule tous les rapports possibles entre le nom et les autres éléments de la phrase. Cependant, si la flexion nominale ne suffisait pas, on ajoutait un adverbe déterminant le sens du verbe. Ce phénomène a été de plus en plus fréquent et la cooccurrence de l'adverbe et de la désinence casuelle a eu pour effet de transformer l'adverbe rattaché au verbe en préposition régissant la forme nominale. Beaucoup de prépositions sont issues des adverbes. On a assisté à une augmentation du nombre des prépositions et à une baisse du nombre des cas. En effet, les prépositions ont pris en charge une partie des rapports à exprimer et de ce fait, le rôle des désinences casuelles s'est amoindri. Des synapses¹ ont pu alors se produire.

« Plus une préposition est « dématérialisée », « abstraite », moins la correspondance cas ↔ préposition se présente avec des contours nets et faciles à établir. Plus une

¹ Fusion, sous morphologie identique, de cas psychiques différents ayant entre eux certaines affinités ; par exemple, l'ablatif en latin est vu comme une synthèse de l'instrumental, du locatif et de l'ablatif indo-européens selon les comparatistes.

préposition est riche en substance notionnelle et d'un usage étroitement limité, plus elle a vocation à ne correspondre qu'à un cas bien précis, à être monocasuelle ». ¹

La fonction forme un écran entre la marque dite « casuelle » et le cas profond ; elle ne permet pas d'identifier la relation casuelle dont un nom est porteur ; elle est multicasuelle et peut être vue comme une interposition entre le cas et la « marque casuelle ».

La polysémie des fonctions et la polysémie des relateurs (ex. *de*) (les deux polysémies ne se superposent pas) entraînent des complications pour établir une théorie de la relation entre les cas conceptuels et les prépositions.

La question du rapport entre les cas et les prépositions est d'une grande complexité. Fillmore associe chaque préposition à chacun des cas profonds.

« Dans l'analyse des prépositions la notion de cas apparaît comme la plus adéquate quand il s'agit de donner consistance à ce « schème de compréhension » auquel il faut remonter pour interpréter les rapports prépositionnels ». ²

Françoise Létoublon parle de polyvalence des prépositions et de polyvalence des cas pour le grec et note le phénomène parallèle pour l'allemand. Une préposition peut s'agglutiner au verbe. ³

Selon Jürgen Weissenborn, les conditions d'emploi syntaxiques d'une préposition concernent surtout des problèmes de rection dans les langues casuelles.

Benvéniste écrit :

« Monsieur Hjelmslev, quoiqu'occupé uniquement des cas, n'a pu se dispenser de considérer en même temps, au moins d'une manière latérale, les prépositions ; et à bon droit, si étroit est le rapport fonctionnel entre les deux catégories ». ⁴

On peut lire dans Hjelmslev :

« Pott a introduit dans la linguistique l'hypothèse générale d'un rapport intrinsèque et d'une transmission possible entre le système des cas et celui des prépositions. [...] il paraît que les cas recouvrent la même catégorie sémantique que les prépositions, et que la différence réside plutôt dans la subdivision de cette catégorie (c'est-à-dire le nombre de ses membres et la façon dont ils sont agencés au point de vue paradigmatique) que dans sa délimitation ». ⁵

Il (Hjelmslev, 1971, p. 155) estime que :

« Les limites entre la concordance et la rection sont flottantes et ne jouent aucun rôle au point de vue structural. Il y a dans la préposition un élément casuel et dans la conjonction un élément modal qui régissent un élément identique renfermé dans leur régime. De la sorte la rection se réduit à une concordance ».

¹ Cf. *op. cit.* [Cervoni91], p. 205.

² *Ibid.*, p. 213.

³ Cf. *op. cit.* [Létoublon90], p. 23.

⁴ E. BENVENISTE : *Le système sublogique des prépositions en latin*, Travaux du Cercle Linguistique de Copenhague, 1949, pp. 177-184, cité dans [Pottier62], p. 291.

⁵ HJELMSLEV : *La catégorie des cas*, étude de grammaire générale, Copenhague, 1935, cité dans [Pottier62], p. 291.

Quant à Guillaume, il fait une distinction entre le *cas de langue* (déclinaison) et le *cas de discours* (préposition) :

« Ce qui fait l'originalité du cas de déclinaison par comparaison avec la préposition, c'est peut-être un *morphème à double effet* qui, d'une part, assigne prévisionnellement au nom un certain emploi, à l'exclusion d'autres ; et d'autre part, joue dans le mot le rôle de déterminant de la partie du discours... Quant à la préposition, elle est un *morphème à simple effet*. Elle sert à exprimer une certaine fonction du nom dans la phrase, mais n'intervient à aucun degré comme déterminant de la partie du discours, obtenue dans le mot sans que la préposition ait eu à se produire ». ¹

3.1.3.2. EN ALLEMAND

En allemand, on constate l'existence de neuf classes de lexèmes transitifs (en rapport avec leurs caractéristiques de rection ^o) : ²

1. Nominatif, accusatif	à, per, pro, je
2. Seulement génitif	während, kraft...
3. Génitif, von-GN	östlich, infolge...
4. Génitif, datif	a) dank, wegen... b) trotz, statt... c) zufolge
5. Génitif, von-GN, datif	innerhalb, außerhalb...
6. Seulement datif	aus, seit...
7. Datif, accusatif	a) auf, vor... b) ab c) über
8. Seulement accusatif	für, um
9. Génitif, datif, accusatif	a) entlang b) außer

Il est important de faire une distinction sémantique animé /inanimé, notamment en allemand où le cas sera différent : localisation d'un objet immobile (préposition + datif), localisation d'un objet en mouvement (préposition + accusatif).

La rection de certaines prépositions est à ramener à la valence des constructions verbales : *sich beziehen auf* ↔ *in bezug auf*, introduit un objet prépositionnel qui ne peut être qu'à l'accusatif.

Quand le membre obligatoire d'un groupe prépositionnel est un groupe nominal, celui-ci est impérativement marqué d'un cas. La rection de la préposition peut changer en fonction de sa position.

Nous donnons ci-dessous une liste des prépositions spatiales en fonction du cas qu'elles régissent (selon la place qu'elles occupent). Notons qu'il peut y avoir combinaison de plusieurs positions, et qu'une préposition peut régir plusieurs cas. ¹

¹ G. GUILLAUME : *Esquisse d'une théorie psychologique de la déclinaison*, pp. 172-173, , cité dans [Pottier62], p. 291.

² N. FRIES : *Präpositionen und Präpositionalphrasen im Deutschen und im Neugriechischen*, Max Niemeyer Verlag, Tübingen, 1988, p. 145.

P(G)+ correspond aux prépositions qui régissent le génitif en pré-position

P+(G) correspond aux prépositions qui régissent le génitif en post-position

P+(G)+ correspond aux prépositions qui régissent le génitif en circom-position

P _{+(G)}	=	abseits, angesichts, diesseits, inmitten, jenseits, längsseits, seitlich
P _{+(D)}	=	aus, bei, von
P _{(D)+;(D)}	=	entgegen, gegenüber, nach, zu
P _{+(D)+}	=	an ++ vorbei, von ++ auf, an ++ entlang
P _{+(A)}	=	gegen, per, via, um
P _{(A)+;(A)}	=	durch
P _{+(A)+}	=	auf ++ zu, um ++ herum
P _{+(D/A)}	=	an, auf, hinter, in, neben, über, unter, vor
P _{+(D/A);+(D/A)+und+(D/A)}	=	zwischen
P _{+(G causal/D local)}	=	ob
P _{+(G/D)}	=	außerhalb, binnen, innerhalb, oberhalb, unfern, unweit, vis-à-vis
P _{+(N/A)}	=	je, frei*
P _{(D/A)+;(G/D)}	=	entlang**

TABLEAU 12 : LES PREPOSITIONS ALLEMANDES : POSITIONS ET RECTIONS

* Selon la liste de Engels, les seules prépositions qui peuvent montrer une rection au nominatif sont *je* et *frei* : *Lieferung frei angegebener Bahnhof*

** À l'origine *entlang* demandait le datif, maintenant la pré-position et les rections génitif et accusatif sont possibles. L'accusatif n'est possible qu'en post-position et le génitif (rare) qu'en pré-position.

3.1.4. Conclusion

Les grammaires de cas présentent divers avantages :

- * elles offrent un modèle de la structure profonde d'une phrase où la sémantique joue un rôle essentiel et où elle reste facile à utiliser ; la notion de verbe et de son modèle casuel rappelant l'application du prédicat logique à des arguments.
- * elles tendent vers un mécanisme d'analyse purement sémantique fondé sur les restrictions de sélection issues des ossatures des verbes, mécanisme qui, de plus, n'exclut pas l'utilisation de contraintes syntaxiques. Ce type de théorie est un des rares à autoriser des traitements automatiques dans le cas de phrases non normées, c'est-à-dire des phrases qui ne respectent pas la syntaxe habituelle.

¹ R. DIRVEN, R.-A. GEIGER : *A User's Grammar of English : Word, Sentence, Text, Interaction*, Part B, The Structure of Sentences, Peter Lang Verlag, Frankfurt am Main, 1989, pp. 96-97.

Peut-on dire qu'il y a redondance entre prépositions et cas ? Si on examine les phrases suivantes : 1) *ich gehe in die Stadt*, 2) *ich bin in der Stadt*, on constate qu'il y a une opposition de cas. Le choix des cas permet donc d'exprimer une différence de sens (mouvement ou absence de mouvement). En fait, il n'y a pas redondance avec la préposition mais avec le sens du verbe qui, dans la première phrase, exprime déjà le mouvement.

3.2. L'UNIFICATION

3.2.1. Qu'est-ce que l'unification ?

3.2.1.1. DEFINITION

L'unification est l'opération fondamentale pour de nombreux formalismes grammaticaux contemporains qui utilisent, comme représentations syntaxiques, des structures de traits contenant des informations partielles sur les constituants. Elle permet de tester, de propager ou de combiner l'information contenue dans ces représentations.

« L'unification de deux structures de traits A et B (notée $A \cup B$) est la structure minimale qui est à la fois une extension de A et de B ». Si une telle structure n'existe pas, l'unification « échoue » (ce qui est noté \perp).

L'unification vérifie la compatibilité entre deux structures de traits et produit une structure résultante qui est la plus petite structure qui contient toute l'information contenue dans la première et dans la deuxième structure. Elle consiste donc à combiner l'information des deux structures de traits afin d'obtenir une nouvelle structure qui inclut toute l'information qu'elles portent à elles deux.

3.2.1.2. PROPRIETES

L'unification a différentes propriétés : ¹

- elle ajoute de l'information :

$$[\text{cat:SN}] \cup [\text{accord:}[\text{nombre:singulier}]] = \begin{bmatrix} \text{cat:SN} \\ \text{accord:}[\text{nombre : singulier}] \end{bmatrix}$$

- les variables sont l'élément d'identité de l'unification :

$$[] \cup \begin{bmatrix} \text{cat:SN} \\ \text{accord:}[\text{nombre:singulier}] \end{bmatrix} = \begin{bmatrix} \text{cat:SN} \\ \text{accord:}[\text{nombre:singulier}] \end{bmatrix}$$

- ses effets diffèrent selon que les valeurs sont semblables ou identiques :

$$\begin{bmatrix} \text{accord:}[\text{nombre:singulier}] \\ \text{sujet : accord:}[\text{nombre:singulier}] \end{bmatrix} \cup [\text{sujet:accord:}[\text{personne:troisième}]] = \begin{bmatrix} \text{accord:}[\text{nombre:singulier}] \\ \text{sujet:accord:} \begin{bmatrix} \text{nombre:singulier} \\ \text{personne:troisième} \end{bmatrix} \end{bmatrix}$$

$$\begin{bmatrix} \text{accord:}\diamond 1[\text{nombre:singulier}] \\ \text{sujet:}[\text{accord : } \diamond 1] \end{bmatrix} \cup [\text{sujet:accord:}[\text{personne:troisième}]] = \begin{bmatrix} \text{accord:}\diamond 1 \begin{bmatrix} \text{nombre:singulier} \\ \text{personne:troisième} \end{bmatrix} \\ \text{sujet:accord:}[\diamond 1] \end{bmatrix}$$

¹ P. MILLER, T. TORRIS : *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Hermès, Paris, 1990. pp. 40-41.

L'unification est une opération :

- idempotente : $A \cup A = A$
- commutative (le résultat ne dépend pas de l'ordre des éléments) : $A \cup B = B \cup A$
- associative : $A \cup (B \cup C) = (A \cup B) \cup C$
- déclarative : si $A = A'$ et $B = B'$ alors $A \cup B = A' \cup B'$
- monotone : $A \cup B \supset A$ et $A \cup B \supset B$; si $A \supset B$ alors $\forall C, A \cup C \supset B \cup C$

Les exemples ci-dessous montrent comment fonctionne l'unification :

(1) $[\text{Cat}=\text{D}] \cup [\text{Cat}=\text{N}] = \perp$ (échec)

\Rightarrow l'unification ne peut réussir car les deux structures de traits ont une valeur différente pour le trait *cat*.

(2) $[\text{Cat} = \text{D}] \cup [\text{Accord} = [\text{Num} = \text{sing}]] = \left[\begin{array}{l} \text{Cat} = \text{D} \\ \text{Accord} = [\text{Num} = \text{sing}] \end{array} \right]$

\Rightarrow l'unification réussit, la structure résultante comprend toute l'information contenue dans les deux structures de départ

(3) $\left[\begin{array}{l} \text{Cat} = \text{N} \\ \text{Accord} = [\text{Num} = \text{sing}] \end{array} \right] \cup [\text{Accord} = [\text{Genre} = \text{masc}]] = \left[\begin{array}{l} \text{Cat} = \text{N} \\ \text{Accord} = \left[\begin{array}{l} \text{Genre} = \text{masc} \\ \text{Num} = \text{sing} \end{array} \right] \end{array} \right]$

\Rightarrow l'unification réussit, la paire attribut-valeur [*genre=masc*] est ajoutée aux valeurs du trait *accord*

(4) $\left[\begin{array}{l} \text{Cat} = \text{N} \\ \text{Accord} = [\text{Genre} = \text{fem}] \end{array} \right] \cup [\text{Accord} = [\text{Genre} = \text{masc}]] = \perp$

\Rightarrow l'unification échoue car les deux structures de traits ont une valeur différente pour le trait *genre*.

3.2.2. Les formalismes basés sur l'unification

3.2.2.1. HISTORIQUE

Les formalismes basés sur l'unification sont le résultat de recherches distinctes en linguistique informatique, en linguistique formelle et en traitement automatique du langage. (...) Plusieurs courants de recherche, à l'origine indépendants, ont donc convergé vers l'idée de l'unification comme moyen de contrôle du flux d'informations.¹

La Grammaire lexicale fonctionnelle (LFG, *Lexical Functional Grammar*) se développe en s'inspirant du concept des réseaux de transition enrichis^o (ATN, *Augmented Transition Networks*) et des travaux de Bresnan qui prônent une linguistique non transformationnelle et fortement lexicaliste. Dans le même temps Kay élabore une grammaire fonctionnelle appelée maintenant Grammaire d'unification fonctionnelle (FUG, *Functional Unification Grammar*). C'est également à cette époque que Joshi et al. élaborent la grammaire d'arbres adjoints (TAG, *Tree Adjoining Grammar*).

¹ Cf. *op. cit.* [Miller-Torris91], pp. 28-29.

Indépendamment, Colmerauer travaille sur des outils de traitement automatique du langage naturel (système Q, formalisme des grammaires de métamorphose, Prolog). Ses travaux seront notamment utilisés par Pereira et Warren qui créent les Grammaires à clauses définies (DCG, *Definite Clause Grammar*). Celles-ci serviront de base à d'autres travaux de programmation pour développer des formalismes basés sur l'unification (grammaires d'extraposition, grammaires à cases, grammaires de discontinuité).

Un courant de recherche parallèle basé sur les analyses linguistiques non transformationnelles apparaît. C'est dans ce courant que s'inscrivent la Grammaire syntagmatique généralisée (GPSG, *Generalized Phrase Structure Grammar*) élaborée par Gazdar & Pullum, puis la Grammaire syntagmatique guidée par la tête (HPSG, *Head Phrase Structure Grammar*) définie par Pollard à partir de GPSG.

Au sein de la nouvelle génération de modèles grammaticaux, on trouve la grammaire d'unification fonctionnelle (UCG, *Unification Categorical Grammar*) qui se base, dans une large mesure, sur l'information codée dans le lexique.

3.2.2.2. GENERALITES

L'objectif des grammaires d'unification est de caractériser l'ensemble des phrases bien formées en s'appuyant sur des représentations linguistiques déclaratives, monotones et aussi proches que possible des phénomènes de surface. Elles sont basées sur des définitions mathématiques rigoureuses. Elles s'inscrivent dans la lignée du programme de grammaire générative de Chomsky, elles sont nées d'une critique de la grammaire générative transformationnelle :

- 1) il y a une opposition entre les partisans d'une sémantique « générative » (s'appuyant sur une structure profonde) et les partisans d'une sémantique « interprétative » (s'appuyant sur une structure de surface),
- 2) les psycholinguistes ont un doute sur la plausibilité du modèle transformationnel,
- 3) le modèle mal adapté à la construction de programmes informatiques d'analyse automatique de textes (problèmes d'implémentation).

Les formalismes basés sur l'unification présentent quelques divergences mais ils ont de nombreux liens qui permettent des progrès communs, aussi bien sur le plan des outils de description et de traitement, que sur celui des analyses linguistiques, générales et particulières.

Ils exigent que le formalisme grammatical soit : ¹

- surfaciste : il doit donner une caractérisation directe de l'ordre de surface réel des éléments de la phrase,
- informatif : il doit associer aux chaînes linguistiques de l'information provenant d'un certain domaine d'information,

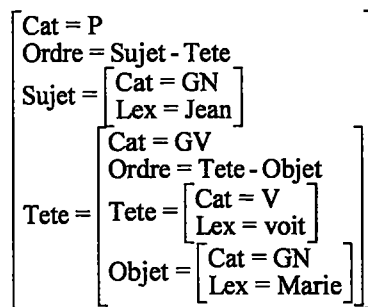
¹ Cf. *op. cit.* [Miller-Torris91], p. 31.

- inductif : il doit définir l'association entre chaînes et éléments d'information de manière récursive, de sorte que de nouvelles associations puissent être dérivées en combinant, selon des opérations préétablies, d'une part des chaînes partielles et d'autre part les informations associées à ces dernières,
- déclaratif : l'association entre chaînes et éléments d'information doit être définie par des assertions et non par des procédures.

Il faut caractériser les éléments d'information comme étant basés sur des complexes de traits (structures de traits) qui associent aux traits des valeurs tirées d'un ensemble bien défini, et éventuellement structuré.

Les points communs les plus représentatifs entre les grammaires d'unification sont les suivants : ¹

- la réhabilitation des descriptions de « surface » : la décomposition des catégories en ensembles de traits permet d'ajouter des contraintes explicites sur les traits aux règles de réécriture proprement dites (accord en genre et en nombre, informations contextuelles (<V sous-cat> = \emptyset), les règles de réécriture sont réinterprétées comme des descriptions de sous-arbres permettant de filtrer ceux qui sont compatibles avec la grammaire considérée,
- le renouvellement des descriptions syntaxiques par la définition de traits « complexes », avec l'utilisation des structures de traits à la place des structures arborescentes. Les traits sont enchâssables dans des structures complexes (appelées descriptions fonctionnelles, combinées à l'aide de l'unification). Ce style d'analyse syntaxique plus « concret » permet de limiter le recours à des éléments « vides » et de restreindre le nombre d'étapes intermédiaires dans la production d'une phrase. *ex. Jean voit Marie* :



- la définition de principes généraux de bonne formation des énoncés, parmi lesquels on trouve notamment les deux suivants :
 - * le principe de cooccurrence pour définir quels ensembles de traits des éléments du cadre forment des catégories linguistiquement cohérentes,
 - * le principe de propagation des traits pour définir quelles répartitions de traits entre catégories appartenant à un même constituant sont prévisibles,
- l'intégration du lexique, de la syntaxe et de la sémantique ; il y a une vérification systématique de la compatibilité entre informations lexicales, syntaxiques et sémantiques grâce à l'unification.

¹ A. ABEILLE : *Les nouvelles syntaxes : grammaires d'unification et analyse du français*, Colin, Paris, 1993, p. 19.

« Certaines grammaires d'unification reprennent des modèles sémantiques construits indépendamment (la grammaire de Montague pour GPSG, la sémantique situationnelle pour HPSG), d'autres modèles développent leur propre mode de représentation sémantique (structure sémantique et structure argumentale en LFG, arbres sémantiques en TAG). »¹

Ce type de grammaire est très intéressant pour les représentations qui font l'objet d'implémentations informatiques (utilisation des traits sémantiques, possibilité de hiérarchiser les contraintes sémantiques et les contraintes morphosyntaxiques, distinction entre sèmes inhérents et afférents, possibilité de description des effets de contexte sémantiques au sein du syntagme).²

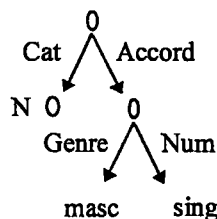
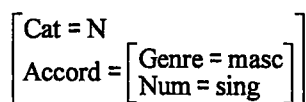
3.2.2.3. LES PRINCIPALES NOTIONS UTILISEES

3.2.2.3.1. Trait

Un trait est un couple attribut-valeur. L'attribut représente le nom du trait, la valeur peut être une variable. Si le trait est (à valeur) atomique, la valeur est un symbole non décomposable (par ex. [genre=féminin]). Si le trait est complexe (ou à valeur non atomique), la valeur est une liste ou un ensemble de traits (par ex. [accord=[nombre=singulier,genre=féminin]]). Enfin, si le trait est binaire (ou booléen), la valeur est égale à + ou -, et peut être calculée à l'aide de connecteurs logiques (par ex. [humain=+]).

3.2.2.3.2. Structure de traits

Une structure de traits est un ensemble de traits. Il existe deux types de représentation entre crochets ou sous forme d'arbres, comme le montrent les deux schémas ci-dessous pour le mot « chien » ; il n'y a pas d'ordonnement, on peut commencer aussi bien par le trait Cat que par le trait Accord.



Une structure de traits est :

- vide quand elle ne comporte aucun trait,
- cyclique quand elle comprend au moins un cycle³,

¹ Cf. *op. cit.* [Abeillé93], p. 26.

² F. RASTIER, M. CAVAZZA, A. ABEILLE : *Sémantique pour l'analyse, De la linguistique à l'informatique*, Masson, Paris / Milan / Barcelone, 1994, p. 139.

³ Un graphe (ou une structure de traits) comporte un cycle quand il existe un chemin non vide entre un nœud et lui-même.

- réursive quand la valeur d'un attribut peut elle-même être une structure d'information,

$$\left[\begin{array}{l} \text{cat : sn} \\ \text{acc : } \left[\begin{array}{l} \text{pers : 3} \\ \text{nb : sg} \end{array} \right] \end{array} \right]$$

- réentrante quand deux attributs sont susceptibles de partager la même valeur. On représente la valeur réentrante au moyen d'un index (noté <1>),

$$\left[\begin{array}{l} \text{suj : acc : } \langle 1 \rangle \\ \text{sv : acc : } \langle 1 \rangle \left[\begin{array}{l} \text{pers : 3} \\ \text{nb : sg} \end{array} \right] \end{array} \right]$$

- non réentrante quand deux attributs comportent des valeurs identiques mais qu'une valeur unique ne se trouve pas effectivement partagée par les deux attributs,

$$\left[\begin{array}{l} \text{suj : acc : } \left[\begin{array}{l} \text{pers : 3} \\ \text{nb : sg} \end{array} \right] \\ \text{sv : acc : } \langle 1 \rangle \left[\begin{array}{l} \text{pers : 3} \\ \text{nb : sg} \end{array} \right] \end{array} \right]$$

- complexe quand elle se compose de traits à valeur non atomique.

Quand on a des structures complexes, on peut faire une distinction entre : ¹

- * les structures identiques : ces structures sont identiques mais ne le resteront pas après des modifications ultérieures : soient la structure A : $\left[\begin{array}{l} \text{Det} = \left[\text{Accord} = \left[\text{Num} = \text{sing} \right] \right] \\ \text{Nom} = \left[\text{Accord} = \left[\text{Num} = \text{sing} \right] \right] \end{array} \right]$ et la structure C : $\left[\text{Det} = \left[\text{Accord} = \left[\text{Genre} = \text{masc} \right] \right] \right]$

l'unification de A et C aura pour résultat :

$$C \cup A : \left[\begin{array}{l} \text{Det} = \left[\text{Accord} = \left[\text{Num} = \text{sing}, \text{Genre} = \text{masc} \right] \right] \\ \text{Nom} = \left[\text{Accord} = \left[\text{Num} = \text{sing} \right] \right] \end{array} \right]$$

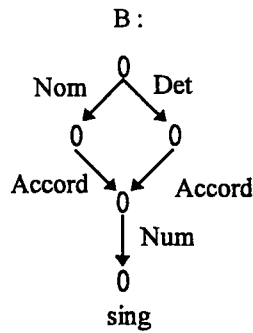
- * les structures à valeurs partagées (ou réentrantes) : ces structures sont identiques mais le resteront après des modifications ultérieures : soient la structure B : $\left[\begin{array}{l} \text{Det} = \left[\text{Accord} = \left[\text{Num} = \text{sing} \right] \right] \\ \text{Nom} = \left[\text{Accord} = \left[1 \right] \right] \end{array} \right]$, et la structure C : $\left[\text{Det} = \left[\text{Accord} = \left[\text{Genre} = \text{masc} \right] \right] \right]$ l'unification de B et de C donnera le résultat suivant:

$$C \cup B : \left[\begin{array}{l} \text{Det} = \left[\text{Accord} = \left[\text{Num} = \text{sing}, \text{Genre} = \text{masc} \right] \right] \\ \text{Nom} = \left[\text{Accord} = \left[1 \right] \right] \end{array} \right]$$

après unification, le trait Accord enchâssé sous l'attribut Nom a également un trait Genre spécifié dans le cas de $C \cup B$ mais pas dans celui de $C \cup A$.

¹ Cf. *op. cit.* [Abeillé93], pp. 29-31.

La représentation, sous forme de graphes, d'une structure réentrante telle que B sera la suivante :



3.2.2.3.3. Extension entre structures de traits

Une structure de traits A est une extension¹ d'une structure de traits B (ce qui se note $A \supset B$) si et seulement si :

- tous les traits à valeur atomique présents dans B sont présents dans A avec la même valeur
- pour tout trait <f> ayant une valeur non atomique, la valeur de <f> dans A est une extension de la valeur de <f> dans B

$$\left[\begin{array}{l} \text{Cat} = \text{N} \\ \text{Accord} = \left[\begin{array}{l} \text{Genre} = \text{masc} \\ \text{Num} = \text{sing} \end{array} \right] \end{array} \right] \supset \left[\begin{array}{l} \text{Cat} = \text{N} \\ \text{Accord} = [\text{Genre} = \text{masc}] \end{array} \right]$$

3.2.2.3.4. Subsumption

C'est une relation d'ordre sur les structures de traits qui mesure la compatibilité et la spécificité relative de l'information qu'elles contiennent. C'est la relation inverse de l'extension. De façon intuitive, une structure de traits plus générale subsume une structure de traits comprenant des informations plus spécifiques. A subsume B si et seulement si B est une extension de A. Toute structure se subsume elle-même. La structure vide subsume toutes les autres.

3.2.2.3.5. Généralisation

La généralisation² de deux structures de traits A et B, notée AGB, est la structure de traits maximale qui subsume les deux, c'est-à-dire telle que A soit une extension de AGB et B aussi. Contrairement à l'unification, elle ne peut pas « échouer ».

$$\left[\begin{array}{l} \text{Cat} = \text{N} \\ \text{Accord} = \left[\begin{array}{l} \text{Genre} = \text{masc} \\ \text{Num} = \text{sing} \end{array} \right] \end{array} \right] \text{G} \left[\begin{array}{l} \text{Cat} = \text{A} \\ \text{Accord} = \left[\begin{array}{l} \text{Genre} = \text{masc} \\ \text{Num} = \text{sing} \end{array} \right] \end{array} \right] = \left[\begin{array}{l} \text{Accord} = \left[\begin{array}{l} \text{Genre} = \text{masc} \\ \text{Num} = \text{sing} \end{array} \right] \end{array} \right]$$

¹ Cf. *op. cit.* [Abeillé93], pp. 31-33.

² *Ibid.*, pp. 35-36.

Pour prendre une analogie tirée de l'arithmétique, on peut comparer le résultat de l'unification au calcul du Plus Petit Commun Multiple (PPCM) entre deux nombres, et celui de la généralisation au calcul de leur Plus Grand Diviseur Commun (PGDC).

Selon L. Hjelmslev, si un objet O1 a une propriété donnée, et un O2 a cette même propriété ainsi que d'autres caractéristiques, le principe de généralisation consiste, en mettant entre parenthèses les déterminations spécifiques de O2, à appliquer au second objet la propriété du premier. Donc si un énoncé admet une lecture et un second en admet deux, l'isotopie retiendra, par généralisation, la lecture qui leur est commune.

3.2.2.3.6. Union prioritaire

Cette opération (notée A/B) est proposée par R. Kaplan (1987), pour les constructions coordonnées avec ellipse. C'est une opération « non commutative (A/B ≠ B/A), qui ajoute à A les traits de B qui ne sont pas déjà dans A (sans comparer leurs valeurs pour les attributs communs), sans changer B et sans prendre en compte les traits incompatibles ». Elle ne peut pas « échouer », elle permet de résoudre le problème de la coordination entre masculin et féminin.

3.2.2.4. PRESENTATION DES FORMALISMES

On peut scinder les grammaires d'unification en deux sous-groupes : les grammaires de type « outils » (DCG, FUG et UCG) et les grammaires de type « théorie » (LFG, GPSG, HPSG et TAG). Nous présenterons ces sept grammaires les unes après les autres en nous attachant un peu plus sur les quatre dernières.

3.2.2.4.1. Grammaire à clauses définies (DCG, *Definite Clause Grammar*)

Les DCG utilisent une variété d'unification basée sur les structures de termes. Les termes diffèrent des structures de traits par deux propriétés importantes : ¹

- * l'ordre : au lieu d'identifier les valeurs en les associant avec un nom de trait, les termes recourent à l'ordre linéaire dans lequel elles sont disposées,
- * l'arité : le nombre d'éléments dans une structure de termes joue un rôle significatif pour l'unification.

Un prédicat est suivi par une série de termes (constantes ou variables) moins complexes placée entre parenthèses.

Les inconvénients des structures de termes sont les suivants :

- * l'encodage de l'unification est laborieux car chaque trait de tête doit être mentionné séparément,
- * l'interprétation des valeurs est difficile en raison de l'absence de noms de trait étiquetant ces valeurs,

¹ Cf. *op. cit.* [Miller-Torris91], pp. 65-66.

- * comme le grammairien doit tout spécifier, cela mène à des grammaires contenant de très nombreuses variables, disséminées un peu partout dans les règles.

En revanche, ces grammaires écrites en DCG peuvent être directement compilées et produisent des analyseurs rapides, ce qui représente un avantage important.

3.2.2.4.2. Grammaire d'unification fonctionnelle (FUG, *Functional Unification Grammar*)

La Grammaire d'unification fonctionnelle est élaborée en 1983 par Kay. La concaténation est la seule opération de formation des constituants. Les structures fonctionnelles, qui constituent les entités d'information, forment un système traits-valeurs généralisé.

FUG utilise exclusivement des structures de traits réentrantes. Une règle est donc simplement une structure de traits comportant des traits partagés.¹

$$\left[\begin{array}{l} \text{cset:} \{ \diamond 3 \ \diamond 4 \} \\ \text{schema:} \langle \diamond 3 \ \diamond 4 \rangle \\ \text{cat:} P \\ \text{tete:} \diamond 1 \text{ [sujet:} \diamond 2 \text{]} \\ \text{sujet:} \diamond 3 \left[\begin{array}{l} \text{cat:} SN \\ \text{tete:} \diamond 2 \end{array} \right] \\ \text{pred:} \diamond 4 \left[\begin{array}{l} \text{cat:} SV \\ \text{tete:} \diamond 1 \end{array} \right] \end{array} \right]$$

Le trait *cset* permet d'identifier les constituants dans une structure. Sa valeur est un ensemble de structures de traits désignant les sous-structures de traits qui doivent être associées à des constituants immédiats.

Le trait *schema* permet de savoir comment les chaînes associées aux constituants immédiats sont ordonnées pour former le constituant supérieur. L'ordre de la suite correspond à l'ordre de concaténation de la chaîne des constituants immédiats.

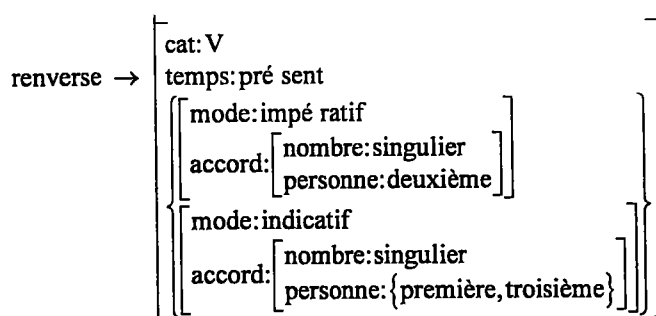
En ce qui concerne les structures fonctionnelles, FUG utilise plutôt des spécifications de chemins que des boîtes de coïndexation pour les traits réentrants, le signe d'égalité à la place des deux points, les parenthèses pour noter les ensembles et les suites.

Ou plus simplement

$$\left[\begin{array}{l} \text{cset} = (\langle \text{sujet} \rangle \ \langle \text{pred} \rangle) \\ \text{schema} = (\langle \text{sujet} \rangle \ \langle \text{pred} \rangle) \\ \text{cat} = P \\ \text{tete} = [\text{sujet} \ \langle \text{sujet tete} \rangle] \\ \text{sujet} = [\text{cat} = SN] \\ \text{pred} = \left[\begin{array}{l} \text{cat} = SV \\ \text{tete} \ \langle \text{tete} \rangle \end{array} \right] \end{array} \right]$$

$$\left[\begin{array}{l} \text{cset} = (\langle \text{tete sujet} \rangle \ \langle \text{pred} \rangle) \\ \text{schema} = (\langle \text{tete sujet} \rangle \ \langle \text{pred} \rangle) \\ \text{cat} = P \\ \text{tete} = [\text{sujet} = [\text{cat} = SN]] \\ \text{pred} = \left[\begin{array}{l} \text{cat} = SV \\ \text{tete} \ \langle \text{tete} \rangle \end{array} \right] \end{array} \right]$$

¹ Cf. *op. cit.* [Miller-Torris91], pp. 59-61.



La disjonction appelée alternative en FUG est représentée par des accolades. Si on met un ensemble de structures fonctionnelles entre accolades, cela signifie qu'une seule de ces structures doit être unifiée. La valeur ANY est une valeur atomique et non un trait. Elle peut être unifiée avec toute valeur, exactement comme une variable, mais contrairement à elle, elle ne peut apparaître dans une structure fonctionnelle finale bien formée. Chacune des valeurs ANY doit obligatoirement être unifiée avec une autre sous-structure. Cette valeur est utilisée notamment dans les cas de sous-catégorisation. Un système comprenant des valeurs ANY permet à une structure fonctionnelle mal formée de devenir bien formée par suite d'unifications postérieures.¹

3.2.2.4.3. Grammaire d'unification catégorielle (UCG, *Unification Categorical Grammar*)

3.2.2.4.3.1. Naissance, motivations de départ, objectifs

En 1929, Leniewski élabore l'esquisse d'une grammaire catégorielle, et Ajdukiewicz, un de ses élèves, va la développer en 1935 pour tendre vers la grammaire catégorielle dite classique. C'est en 1953 que Bar-Hillel forge le concept de grammaire catégorielle. En 1958, Lambek va développer le formalisme existant, d'un point de vue mathématique et en 1971, Geach développera la forme classique, indépendamment de Lambek, et donnera à ses règles un fondement plus fortement orienté vers l'utilisation linguistique.

Leniewski a réuni dans son travail les trois idées essentielles sur lesquelles se fondent les grammaires catégorielles :

- * les catégories de sens de E. Husserl : on lui a attribué la définition du terme « catégorie » suivante :
 - « tous les mots qui sont interchangeable dans un contexte invariable, sans que celui-ci ne devienne insensé, appartiennent à une catégorie ». ²
- * la fonctionnalité du langage de G. Frege. Frege remplace la division traditionnelle de la structure syntaxique en sujet-prédicat par une relation syntaxique unique, en utilisant, pour des expressions linguistiques, le concept de fonctions qu'il a établi pour des expressions mathématiques. On lui a

¹ Cf. *op. cit.* [Miller-Torris91], p. 63.

² Cité dans K. VON HEUSINGER : *Kategoriale Unifikationsgrammatik*, Arbeitspapier 37, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1991, p. 3.

attribué un principe sans qu'il ne l'ait lui-même explicitement formulé et que l'on appelle « Principe de Frege » ou « Principe de Fonctionnalité », qui dit que :

« La signification d'une expression composée est une fonction des sens de ses constituants et de la manière dont ils sont syntaxiquement reliés ». ¹

- * la hiérarchie de la langue de Russel : elle est introduite sous la forme de la théorie des types de Russel, et utilisée pour éviter les antinomies ² (ou paradoxes).

Bar-Hillel occupe une position-clé dans le développement de la grammaire catégorielle. Il s'est inspiré du formalisme de la grammaire catégorielle classique et l'a enrichi afin que l'on puisse l'employer pour décrire le langage naturel de façon plus adéquate. En 1958, Lambek a cherché un algorithme qui permettrait de distinguer les phrases des non-phrases. Il a utilisé les travaux de Bar-Hillel et Ajdukiewicz et a essayé d'établir une arithmétisation plus importante du formalisme sans pour autant perdre de vue le concept de langage naturel. Il a développé ce que l'on appelle le « Calcul de Lambek ».

Les grammaires catégorielles représentent le type de grammaires formelles le plus ancien, tout au moins, plus ancien que les grammaires à contexte libre, les grammaires transformationnelles et les grammaires de dépendance. Au départ, elles avaient une syntaxe peu sophistiquée. Bien que leur pouvoir expressif ait été augmenté, elles n'ont pas pris en compte l'une des évolutions fondamentales des formalismes grammaticaux. En effet, les modèles basés sur des grammaires syntagmatiques ont montré une forte tendance à encoder de plus en plus d'information dans des symboles non terminaux complexes et très structurés. Les catégories sont conçues comme des structures de traits complexes.

K. von Heusinger affirme que « nous pouvons concevoir une grammaire catégorielle comme une catégorie linguistique très générale, simple et claire, qui est très marquée par la sémantique mais soumise à peu de restrictions syntaxiques ».

Cependant il se pose un problème avec les catégories construites selon des critères sémantiques et qui ne peuvent pas appréhender toutes les finesses syntaxiques. Par exemple, pour décrire avec succès des phénomènes de congruence et de rection, il faut multiplier les catégories et les règles selon des critères de nombre, de genre, de cas, de finitude... On peut résoudre ce problème en développant la grammaire catégorielle vers une grammaire d'unification catégorielle.

La grammaire d'unification catégorielle intègre à la sémantique des traits syntaxiques, dont la propagation peut être notée en indexant les symboles catégoriels par des alphas et des betas. Elle hérite de l'unification le concept d'information partielle et introduit ce type d'information dans les graphes catégoriels. Contrairement à la grammaire catégorielle classique où la catégorie foncteur encode seulement le type de fonction (la fonction particulière elle-même est supposée être purement sémantique), en

¹ Cf. *op. cit.* [Heusinger91], p. 4.

² LOG. Contradiction entre deux idées, deux principes, deux propositions. - Contradiction à l'intérieur d'une théorie déductive.

grammaire d'unification fonctionnelle, on peut considérer le foncteur comme encodant une fonction syntaxique et une fonction sémantique d'un type particulier. ¹

Les grammaires d'unification catégorielles font partie d'une nouvelle génération de modèles grammaticaux très empreints pour leur utilisation en linguistique informatique. Elles ont également des caractéristiques linguistiques très intéressantes. Elles se basent, dans une large mesure, sur l'information codée dans le lexique. Ceci permet de coder les informations syntaxiques et sémantiques dans un seul endroit et de faciliter un travail parallèle. La syntaxe et la sémantique utilisent des principes généraux et n'ont pas besoin de règles très spécialisées. ²

Les premières grammaires d'unification catégorielles ont été écrites en PATR-II, dont la notation est très bien adaptée à l'émulation d'autres formalismes, et testées sur les systèmes implémentés au centre de recherche SRI.

3.2.2.4.3.2. Notations utilisées dans UCG

Dans la grammaire catégorielle, les éléments linguistiques sont rangés dans des catégories. Les noms et les phrases sont les catégories de base, les autres éléments sont des catégories de foncteurs, qui, selon leur utilisation, sont construites sur d'autres catégories.

Catégorie	Notation	Exemple
Phrase	s	die Rose blüht
Nom	n	Rose, Sokrates
Prédicat à un argument	s/n	blüht, fliegt
Prédicat à deux arguments	s/nn	liebt
Attribut	n/n	schön, groß
Article	n/n	die, eine
Adverbes de phrase	s/s	nicht
Conjonctions	s/ss	und, oder, aber
Adverbes	(s/n)/(s/n)	stark, gut
Adverbes	((s/n)/(s/n))/ ((s/n)/(s/n))	sehr

TABLEAU 13 : LES NOTATIONS UTILISEES DANS UCG ³

¹ Cf. *op. cit.* [Miller-Torris91], p. 201.

² Cf. *op. cit.* [Heusinger91], p. 1.

³ Ce tableau est extrait de *op. cit.* [Heusinger91], p. 13.

Dans l'exemple suivant ¹ les catégories de foncteurs sont orientées vers la droite :

und	sehr	stark	duftet	der	Flieder	blüht	die	Rose
s/ss	$((s/n)/(s/n))$	$((s/n)/(s/n))$	$(s/n)/(s/n)$	s/n	n/n	n	s/n	$\frac{n/n \quad n}{n}$
					$\frac{n/n \quad n}{n}$	n	$\frac{s/n \quad n}{s}$	
	$\frac{((s/n)/(s/n)) \quad ((s/n)/(s/n)) \quad (s/n)/(s/n)}{(s/n)/(s/n)}$		s/n			n		
	s/n					n		
s/ss		s					s	

3.2.2.4.3.3. Le formalisme UCG

3.2.2.4.3.3.1. Description du formalisme

UCG s'inspire des idées des grammaires catégorielles et des grammaires d'unification. Ces deux types de grammaire présentent chacun des avantages : les grammaires catégorielles sont claires au niveau conceptuel et permettent des liens étroits entre syntaxe et sémantique, les grammaires d'unification, quant à elles, se répandent rapidement et se prêtent à des implémentations rapides, mais efficaces. ² Elle utilise quatre catégories (spécifiées par des traits) comme primitives, ces dernières étant de deux types :

- * les primitives basiques regroupent les noms (noun) et les phrases (sent) ; il existe des entrées lexicales pour ce type de catégories,
- * les primitives non basiques regroupent les syntagmes nominaux (NP) et les syntagmes prépositionnels (PP) ; « pour ce type de catégories il existe seulement des entrées lexicales dont la catégorie est une catégorie complexe formée à partir de ces catégories non basiques. ³ Ces catégories dérivées sont des fonctions d'une catégorie vers une autre.

Les catégories se définissent de la façon suivante :

- a. Toute catégorie primitive (y compris les traits qui lui sont associés) est une catégorie
- b. Si A est une catégorie, et B est un signe, alors A/B est une catégorie.

« On appelle B la partie active de la catégorie, et également du signe entier dont la catégorie est A/B. L'opérateur « / » est associatif à gauche. Lorsqu'un A/B se combine avec un B, on dit que A/B est le foncteur et B l'argument ».

¹ L'exemple est tiré de *op. cit.* [Heusinger91], p. 16.

² Cf. *op. cit.* [Miller-Torris91], p. 200.

³ K. BASCHUNG : *Grammaires d'unification à traits et contrôle des infinitives en français*, Adosa, Clermont-Ferrand, 1991.

On peut dissocier l'application fonctionnelle en deux opérations distinctes :

- * l'instanciation : S_3 est l'instanciation de S_1 par rapport à S_2 si S_3 est S_1 après unification de la partie active de S_1 avec S_2 . L'instanciation a pour effet de spécifier plus avant la sémantique.
- * la réduction : étant donné un signe S_1 de catégorie A/B, le résultat de la réduction de S_1 est un signe S_2 identique en tous points à S_1 sauf que sa phonologie est la concaténation des phonologies de S_1 et de B, et que sa catégorie est réduite à A. « Comme l'unification peut échouer, il peut y avoir beaucoup de signes par rapport auxquels il n'est pas possible d'instancier un signe donné S_1 . » C'est alors qu'on utilise la réduction.

La règle d'application fonctionnelle se définit alors de la façon suivante : soient S_1 et S_2 des signes bien formés, la réduction de l'instanciation de S_1 par rapport à S_2 a pour résultat un signe bien formé.¹

3.2.2.4.3.3.1.1. Le signe

C'est l'objet de base de UCG. Il est composé de traits, c'est-à-dire d'attributs et de valeurs représentables sous la forme de graphes acycliques orientés[◊] (ou DAGs). Il est formé d'un quadruplet dont le premier élément représente la phonologie (W), le second la catégorie syntaxique (C), le troisième la sémantique (S) et le quatrième l'ordre (O) et se note soit verticalement :

W
C
S
O

ou horizontalement : W : C : S : O

Une entrée lexicale utilise, en règle générale et par commodité, une combinaison des deux notations.

ex. : regarder
sent/W1:np:X:pre/W2:np:Y:post
[e] regarder (e,X,Y)
O

La catégorie de cet exemple est complexe, sa partie active est un signe dont la phonologie est la variable W2, dont la catégorie est np, dont la sémantique est la variable Y et dont l'ordre est post. Pour simplifier la notation, il est admis de ne pas noter les valeurs non spécifiées d'un signe (à moins que la variable en question n'ait une autre occurrence ailleurs dans le même signe) :

ex. : regarder
sent/:np:X:pre/:np:Y:post
[e] regarder (e,X,Y)

¹ Cf. *op. cit.* [Baschung91], p. 165.

3.2.2.4.3.3.1.2. Les règles

Une règle est composée d'une tête et d'un corps. La tête est une règle de réécriture non contextuelle et le corps un graphe acyclique orienté.

tête $X_0 \rightarrow X_1 X_2$

corps en notation d'unification

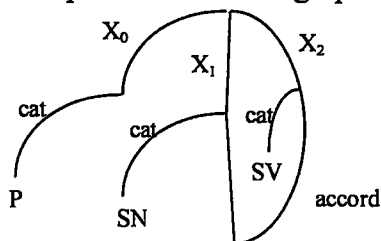
$\langle X_0 \text{ cat} \rangle = P$

$\langle X_1 \text{ cat} \rangle = SN$

$\langle X_2 \text{ cat} \rangle = SV$

$\langle X_1 \text{ accord} \rangle = \langle X_2 \text{ accord} \rangle$

corps en notation de graphe



3.2.2.4.3.3.1.3. Les catégories

Les catégories primitives sont sent, np et pp. Elles sont augmentées de traits de type morphologique essentiellement. On trouve par exemple dans la version française FDP¹ :

Catégorie	Traits	Morphologie
sent	fin	forme verbale finie
	inf	forme verbale infinitive
	cfin	forme verbale finie avec complémenteur (e.g. complétive).
	prp	participe présent
	psp	participe passé
np	subj	sujet
	obj	objet
pp	P	préposition de sous-catégorisation P
	adj	syntagme prépositionnel adjoint

TABEAU 14 : LES CATEGORIES DANS UCG (VERSION FRANÇAISE FDP)

Le fait de définir les catégories complexes comme C/Signe permet de travailler avec une notion très simple d'application fonctionnelle, tout en faisant passer l'information présente dans le signe argument au signe qui résulte de l'application. Cela est rendu possible par la présence de variables identiques dans le signe foncteur et dans la partie active de la catégorie. L'information transmise peut concerner la sémantique, les traits, l'ordre et/ou la catégorie syntaxique de l'expression argument. L'information se transmet dès lors qu'il y a unification, et comme l'unification est commutative, elle se transmet dans n'importe quel sens.

3.2.2.4.3.3.1.4. L'ordre

Dans UCG, il y a une forte restriction à ne combiner que des éléments adjacents. *post* signifie que le foncteur suit l'argument dans l'application fonctionnelle et *pre* signifie que le foncteur le précède.

¹ Cf. *op. cit.* [Baschung91], p. 167.

3.2.2.4.3.3.1.5. La sémantique

Au niveau sémantique, UCG utilise un dérivé de la *Discourse Representation Theory* mais avec une différence importante : chaque formule est assortie d'une variable spéciale, son index, dont la valeur est la chose décrite ou postulée par l'expression, et du fait que les variables peuvent être typées, il est possible d'exprimer de l'information sur le type d'entité que l'expression décrit. Une variable typée est un ensemble de traits associés à une variable particulière ou à une expression référentielle. Il est possible d'unifier des variables typées, ce qui permet à la fois d'exprimer les contraintes de sélection et d'enrichir progressivement l'information liée à la variable. Les termes sont des variables (typées), des constantes et des formules.¹

On peut définir l'ensemble des formules de la façon suivante :

- a. Si x est une variable, P un prédicat à arité n et t_1, \dots, t_{n-1} des termes du type requis par P , alors $P(x, t_1, \dots, t_{n-1})$ est une formule (atomique) à index x .
- b. Si x est une variable et A et B des formules alors $[x] [A, B]$ est une formule (conjonction).
- c. Si x est une variable et A et B des formules alors $[x] [A \Rightarrow B]$ est une formule (implication).
- d. \perp est une formule (l'absurde).

3.2.2.4.3.3.2. *FDP (French Dialog Parser) - un fragment UCG pour le français*

3.2.2.4.3.3.2.1. Le signe

Il comporte six champs et prend la forme suivante :

catégorie^traits(/signe1/.../signen)
 :transitions morphologiques
 :sémantique
 :occurrence
 :ordre
 :phonologie

ex. pour le verbe aller :

aller:
 [sent^[inf,_,_]/(np^[subj,_,_]:_X:(_:occ:[_,_,subj,_,Occ]):pre:_)
 :(nneg,nque)
 :((0:3:1:1:_:go:_):go[X])
 :Occ
 :O
 :_]
 :morpho(être,[aller],[v,all,i,aill,alla]).²

¹ Cf. *op. cit.* [Baschung91], p. 171.

² Les exemples qui suivent ont été donnés dans *op. cit.* [Baschung91], p. 174 sqq.

Le premier champ comporte la catégorie de base (sent) à laquelle est associée une liste de trois traits : la forme morphologique (ici, verbe à l'infinitif), l'information sur les transitions entre clitiques (non encore spécifiée), l'information sur les valeurs d'accord (non encore spécifiée).

Le second champ s'intéresse aux transitions morphologiques relatives à la négation et à l'interrogation.

Le troisième champ est celui de la sémantique dont la syntaxe a la forme ternaire suivante Index:Prédicat:Liste argumentale (ici le prédicat est *go*, la liste argumentale ne comporte qu'un argument non encore spécifié (le sujet) mais unifié à la sémantique du groupe nominal sujet attendu). L'index est un objet complexe composé de huit cases dont les six premières servent à exprimer des informations « ontologiques » sur l'objet (ici il s'agit d'un objet temporel de type locatif exprimant le mouvement d'une source vers une destination, de type événementiel), la septième opère un renvoi à un prédicat de la base de connaissance et la huitième comporte un nombre, identificateur propre de l'objet.

Le quatrième champ gère l'occurrence des syntagmes nominaux pour la résolution des pronoms.

Le cinquième champ est celui de l'ordre, il comprend les valeurs *pre*, *post*, *opt* (élément optionnel : *Jean mange / Jean mange une pomme*) et *jump* (élément permutable : *Jean apporte un livre à Marie / Jean apporte à Marie un livre*).

Enfin, le sixième champ concerne la phonologie ; cette information peut se trouver au début d'une entrée lexicale pour plus de lisibilité. Une entrée lexicale complète comporte également une information morphologique ; comme le français est une langue très riche de ce point de vue, on ne stocke pas dans le lexique une entrée par forme fléchie mais on ne met que les formes de base.

3.2.2.4.3.3.2.2. Les règles

En UCG, il y a deux types de règles classiques d'application fonctionnelle en avant (>*apply*) et en arrière (<*apply*), dont la représentation est la suivante :

```
'>apply'⇒
C:MT:S/Occ:O:(['>apply',W1,W2],_)
→
(C/#1):MT:S:Occ:O:(W1,_),
1#(C1:MT1:S1:Occ1:pre:(W2,_)).
```

```
'<apply'⇒
C:MT:S/Occ:O:(['<apply',W1,W2],_)
→
1#(C1:MT1:S1:Occ1:post:(W1,_)).
(C/#1):MT:S:Occ:O:(W2,_),
```

3.2.2.4.3.4. Le traitement des prépositions dans UCG

Il existe des règles qui permettent d'insérer les adjoints dans la liste catégorielle d'un verbe, d'un nom ou d'une phrase. Comme la règle d'optionnalité, ce sont des règles unaires, qui introduisent des compléments non obligatoires dont la catégorie est *pp*, et pour lesquels on a la constante *adj* comme première valeur de la liste des traits. Elle traitent respectivement :

- * les modifieurs de syntagmes verbaux (ppv) : *Jean envoie Marie à Paris*
- * les modifieurs de noms communs (ppn) : *La fille de Paris s'appelle Marie*
- * les modifieurs de phrases (pps1, pps2) : *A Paris, Jean a donné un livre à Marie.*

ppv⇒

```
2#(sent^[fin or inf or pspe or pas]_)
      /(np^[subj]_:_:Occs:pre:_)
      /(pp^[adj]_:_:I:(_:occ:[_,_],adj,_,Occs):post:_)
:M:1#(I:_:):Occ:O:([ppv,W],_)
→
#2:M:(#1):Occ:O:(W,_)
```

ppn⇒

```
2#(noun^[_,(nil or adj),_])/(pp^[adj]_:_:I:(_:occ:[_,_],adj,_,Occ)
      :post:lex)
_:1#(I:_:):Occ:O:([ppn,W],_)
→
#2:_(#1):Occ:O:(W,_)
```

pps1⇒

```
2#(sent^[fin]_)
/(pp^[adj]_:_(T:O:M:A:B:C:KB:V):( _:occ:[_,_],adj,_,_):pre:lex)
/(lex:_:_:post:(virg,_)
:NW:1#((T:L:M:A:B:C:KB:V:_:):):Occ:O:([pps1,W],_)
→
#2:NW:(#1):Occ:O:(W,_)
```

pps2⇒

```
2#(sent^[fin]_)
/(pp^[adj]_:_(T:O:M:A:B:C:KB:V):( _:occ:[_,_],adj,_,_):pre:(lexint or print))
:NW:1#((T:L:M:A:B:C:KB:V:_:):):Occ:O:([pps2,W],_)
→
#2:NW:(#1):Occ:O:(W,_)
```

3.2.2.4.3.5. Le formalisme PATR-II

PATR-II a joué un rôle très important dans les développements expérimentaux de la grammaire d'unification catégorielle. Le but était de fournir un outil permettant d'écrire, de tester et de comparer des grammaires de types différents. PATR-II a donc été utile pour écrire des grammaires, concevoir des formalismes grammaticaux et explorer des classes de formalisme.

PATR-II est un formalisme informatique qui permet l'implémentation et l'utilisation de grammaires basées sur l'unification. C'est également un formalisme linguistique basé sur l'unification qui favorise la définition de grammaires à traits et leur utilisation pour l'analyse. Les quatre caractéristiques principales de PATR-II sont les suivantes :

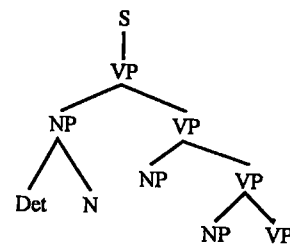
- * il sert à décrire la structure de surface d'une phrase,
- * il permet d'associer à chaque chaîne de mots analysée une structure d'information, c'est un langage informationnel,
- * il permet de définir de façon inductive la manière dont les éléments d'une phrase et leur structure de traits sont combinées, c'est un langage inductif,
- * il détermine les combinaisons possibles plutôt que d'indiquer la manière dont ces combinaisons se réalisent effectivement, c'est un langage déclaratif.

Dans PATR-II, les données ne font pas référence à une catégorie mais à une structure de traits complexe, qui se compose de paires <attribut-valeur> où l'attribut est un élément simple, alors que la valeur peut être simple ou complexe.

PATR-II repose sur deux principales opérations :

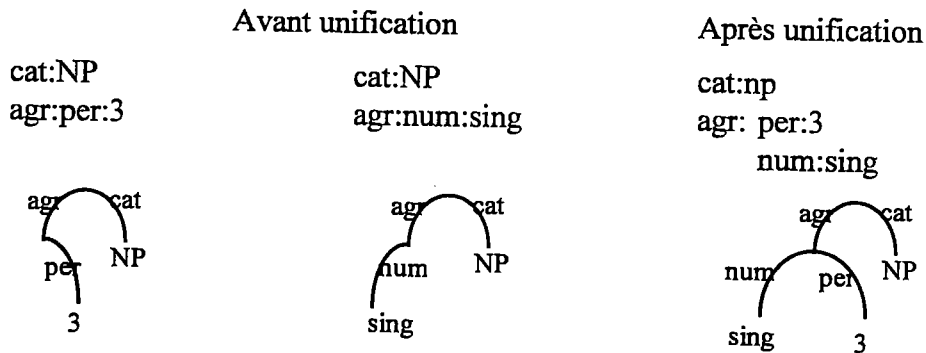
- * la concaténation : elle permet de combiner deux éléments de la chaîne à analyser, ce qui garantit que la grammaire résultante est une grammaire de surface. Elle est décrite dans une règle syntagmatique et représentée dans un arbre ¹.

$S \rightarrow VP$
 $VP \rightarrow NP VP$
 $NP \rightarrow Det N$



- * l'unification : elle permet de combiner deux structures d'information, ce qui autorise le développement de grammaires à la fois déclaratives, inductives et informationnelles. Il y a unification de deux constituants quand ils n'ont pas de valeurs différentes pour un même trait. Elle est particulièrement bien adaptée pour les structures partiellement spécifiées, c'est-à-dire les structures qui ne précisent pas toutes les valeurs et tous les traits.

¹ Le graphique est issu de U. EGLI, K. VON HEUSINGER : *Das Patr-II Format in Anwendung auf das Deutsche*, Arbeitspapier 42, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1992, p. 2.



Il y a trois composantes principales dans une règles PATR-II :

1. une règle de structure écrite sous forme d'une règle syntagmatique et représentée sous forme d'un arbre,
2. des équations de tests qui vérifient si la concordance des traits et donc si les deux structures de traits sont unifiables,
3. des équations de construction qui transmettent les traits et valeurs aux nouveaux constituants du nœud-mère.

- ex. : 1. structure : NP → Det N
 2. test : Nagr = Detagr
 3. construction : Npagr = Nagr

La plupart des grammaires écrites en PATR-II utilisent également des graphes de constituants pour encoder l'information sémantique. Chaque constituant a un attribut appelé *trad* ou *sémantique*, dont la valeur contient obligatoirement la structure sémantique fonction-argument propre à ce constituant, mais elle peut aussi encoder d'autres informations sémantiques. L'encodage séparé de la sémantique permet une sémantique compositionnelle, même dans des constructions qui ont des structures syntaxiques et sémantiques divergentes.

3.2.2.4.4. Grammaire lexicale fonctionnelle (LFG, *Lexical Functional Grammar*)

3.2.2.4.4.1. Naissance, motivations de départ et objectifs

Ce formalisme est élaboré à la fin des années 70 par Joan Bresnan et Ronald Kaplan. Tous deux visent la réhabilitation des descriptions de surface et le remplacement de la composante transformationnelle par un niveau « lexical-fonctionnel ». ¹ Leurs motivations de départ sont l'intégration des travaux en psycholinguistique en abandonnant l'appareil transformationnel jugé peu plausible et la réhabilitation des notions fonctionnelles pour permettre l'expression d'universaux linguistiques pour des langues « non configurationnelles ». Encore à l'heure actuelle, un des objectifs principaux de la recherche en LFG est la découverte de principes de correspondance

¹ Cf. *op. cit.* [Abeillé93], p. 41.

entre les structures c (structures de constituants) et les structures f (structures fonctionnelles) entre les langues : de langues aborigènes australiennes hautement « non-configurationnelles » (Austin & Bresnan 1994, Nordlinger 1993), à des langues qui ont un ordre des mots libre plus limité tel que le russe (King 1993), le Hindi/Urdu (Mohanani 1994, Butt 1993) et le Tagalog (Kroeger 1993), et finalement à des langues relativement configurationnelles telles que l'anglais (Bresnan 1994).

3.2.2.4.4.2. Organisation du modèle

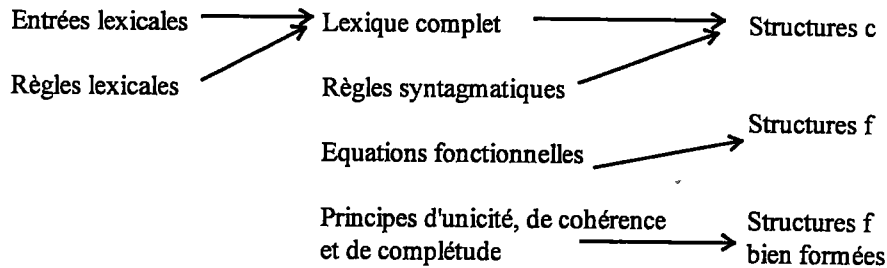


FIGURE 9 : ORGANISATION DU MODELE LFG ¹

3.2.2.4.4.3. Notations utilisées dans LFG

- ⊃ : comprend (comme élément)
- ↓ : ensemble des traits de la catégorie désignée
- [] : structure de traits (structure fonctionnelle)
- ↑ : ensemble des traits de la catégorie immédiatement dominante
- ⇒ : produit d'une règle lexicale
- =_c : équation contrainte
- ◊ : liste de fonctions grammaticales sous-catégorisées et recevant un rôle sémantique

3.2.2.4.4.4. Le formalisme LFG

3.2.2.4.4.4.1. Description du modèle

LFG est une théorie lexicale de la grammaire fonctionnelle. Ce formalisme insiste sur l'importance de la réalité psychologique de la théorie grammaticale. La compétence linguistique est une notion importante, mais la composante essentielle reste le lexique.

LFG est composée de deux niveaux : ²

- * le premier niveau est celui de la structure externe ou structure des constituants, (structure c, *constituant structure* ³) : on obtient des arbres de dérivation à partir de règles de réécriture hors contexte qui correspondent à des règles syntagmatiques ou à des entrées lexicales. Ces arbres étiquetés décrivent

¹ Le graphique est extrait de *op. cit.* [Abeillé93], p. 42.

² *Ibid.*, pp. 44-49.

³ Nous utiliserons la notation française structure c, et structure f.

l'agencement superficiel des éléments de la phrase. Le principe de variabilité établit que les structures externes varient entre les langues.

- * le second niveau est celui de la structure interne ou fonctionnelle (structure *f*, *feature structure*), structure de traits minimale qui satisfait la description fonctionnelle de l'arbre syntagmatique. Des structures hiérarchiques de couples « attribut-valeur » permettent de décrire des fonctions comme sujet, prédicat... ainsi que les caractéristiques morpho-sémantiques et lexicales qui leur sont attachées. Le principe d'universalité établit que les structures internes ne varient pas entre les langues.

Le postulat de principes reliant les structures *f* à la structure argumentale représente un développement important (Bresnan & Kanerva 1989, Joshi 1993, Alsina 1993, Mchombo 1993, Austin 1993, Zaenen 1993). La structure argumentale encode l'information concernant les rôles joués par les différents participants dans une situation décrite par une phrase. Un classement des rôles détermine la fonction grammaticale des arguments du verbe dans la phrase, et ainsi la structure *f* pour la phrase.

La structure *c* et la structure *f* ont en commun le lexique et la description de fonctions. La structure *c*, qui indique l'ordre des mots mais apporte peu d'informations, a un rôle moins important que la structure *f*, au niveau de laquelle sont représentées les fonctions primitives.

En ce qui concerne les équations fonctionnelles, certains enrichissements notationnels ont été ajoutés :

- * les équations avec « chemin » fonctionnel : un chemin est un attribut enchâssé sous un autre notamment pour les phénomènes de dépendance entre deux catégories qui ne sont pas de dominance immédiate,
- * les contraintes « existentielles » : ce sont des contraintes sur la présence (ou l'absence) d'un certain attribut sans préciser sa valeur,
- * les équations « contraintes » (notée =_c) : elles sont utilisées pour vérifier qu'un trait a une valeur donnée dans une structure fonctionnelle.

De nombreuses règles de réécriture sont utilisées mais elles ne sont pas dérivées l'une par rapport à l'autre ; il n'y a pas de transformation comme chez Chomsky.

Le lexique a un rôle très important notamment pour filtrer les structures correctes ; la multiplicité des structures *c* correspond à la multiplicité des entrées lexicales qui comportent des informations fonctionnelles et des indications concernant certains phénomènes dits de « contrôle fonctionnel ». Ce n'est qu'après l'insertion lexicale qu'interviennent les structures *f*.

Bresnan a mené une réflexion théorique approfondie sur le statut du lexique et en particulier sur les conditions de bonne formation d'une entrée lexicale. Elle travaille dans une perspective cognitive : il y a toujours la distinction entre « compétence [◊] » et « performance [◊] » mais « elle cherche à construire un modèle de la compétence ayant une vraisemblance psychologique plus forte que celui de Chomsky ». ¹

¹ C. FUCHS, P. LE GOFFIC : *Les linguistiques contemporaines, Repères théoriques*, Hachette Supérieur, Paris, 1992, pp. 94-96.

Dans LFG, on associe à chaque règle syntagmatique une ou plusieurs règles (ou équations) contraignant l'unification des structures de traits des différentes catégories.

Les règles lexicales (notion centrale dans LFG, reprises dans HPSG et TAG) relient les équations fonctionnelles caractéristiques des entrées lexicales. Elles peuvent modifier les fonctions grammaticales sous-catégorisées dans la liste du trait Pred.

La sous-catégorisation \diamond , qui correspond à la valence en grammaire de dépendance, indique quel type de complément prend un prédicat donné. En LFG, elle s'exprime en termes de fonctions grammaticales (notées entre crochets, en plus de la valeur sémantique du prédicat, comme valeur de l'attribut Pred) et non en termes de catégories syntagmatiques ; de plus la sous-catégorisation des sujets et celle des compléments se fait de la même façon.

Les fonctions grammaticales sont considérées comme des primitives de la théorie linguistique, indépendantes de leurs réalisations syntagmatiques qui varient considérablement selon les langues. Elles peuvent être codées dans les règles syntagmatiques et dans les entrées lexicales. Parmi elles, on peut distinguer :

- * les fonctions sous-catégorisables : Sujet, Objet, Comp (complétives), Xcomp (infinitives et participiales), Prep-Obj pour les compléments prépositionnels (aussi appelés Obliques),
- * les fonctions non sous-catégorisables : Ajout, Xajout = fonctions auxquelles il manque un élément (en général le sujet) pour être saturées,
- * Focus et Topic (utilisées pour les phénomènes d'extraction).

FONCTIONS GRAMMATICALES		
Sous-catégorisables		Non sous-catégorisables
Non restreintes	Restreintes	ADJ (contrôlée de l'intérieur)
SUBJ	OBL $_{\theta}$	XADJ (contrôlée de l'extérieur)
OBJ	COMP (contrôlée de l'intérieur)	
OBJ2	XCOMP (contrôlée de l'extérieur)	
	TOPIC	
	FOCUS	

TABLEAU 15 : LES FONCTIONS GRAMMATICALES DANS LFG

Les relations grammaticales sont des associations de fonctions grammaticales avec des rôles thématiques ou des valeurs non thématiques. Ces associations sont encodées dans le lexique, où chaque verbe est représenté comme une forme lexicale composée d'une structure argument prédicat (liste des arguments pour lesquels il y a des restrictions de sélection) et d'une assignation de fonctions grammaticales (liste des fonctions syntaxiquement sous-catégorisées).

1. structure prédicat argument : break (agent,theme)
2. assignation de fonctions grammaticales : ((SUBJ),(OBJ))
3. forme lexicale : 'break agent theme'

SUBJ OBJ

Les listes d'assignation de fonctions grammaticales servent de cadres de sous-catégorisation. La sous-catégorisation est contrôlée dans la structure fonctionnelle par la complétude, qui vérifie que tous les arguments sous-catégorisés sont présents dans la structure fonctionnelle, et par la cohérence, qui réduit les occurrences des fonctions grammaticales sous-catégorisables à celles précisées dans la forme lexicale du verbe (Kaplan & Bresnan, 1982).

Plusieurs types d'attributs sont employés :

- * les attributs à valeur atomique : traits morphologiques (ou fonctionnels) tels que genre, nombre... L'attribut Pred représente un cas particulier : il apparaît dans les entrées lexicales et prend pour valeur la « forme sémantique » (la valeur sémantique + le cadre de sous-catégorisation éventuel) de l'item en question. Il n'est JAMAIS unifiable,
- * les attributs ayant pour valeur une structure fonctionnelle : Suj (sujet), Obj (objet) ; ils ont pour valeur une structure fonctionnelle qui peut correspondre à un syntagme en structure c, mais qui peut aussi ne pas avoir de réalisation syntagmatique, ou être répartie en plusieurs constituants discontinus,
- * les attributs ayant pour valeur un ensemble de structures fonctionnelles, ensemble dont le nombre de structures f n'est pas fixé.

LFG emploie des métavariabes syntaxiques notées par des flèches : \uparrow renvoie à la structure f supérieure et \downarrow à la structure f inférieure du constituant immédiat auquel l'unification est associée.¹

P → SN SV
 (↑ tête sujet)=(↓ tête) (↑ tête)=(↓ tête)

En utilisant une spécificité de LFG, il n'y a pas besoin de distinguer le trait *cat* des traits *tête* ; les structures f ne sont censées inclure aucune information concernant la catégorie des constituants auxquels elles sont associées ; elles ne renferment aucune information catégorielle (le problème est que cette information ne pourra pas être utilisée pour la sous-catégorisation).

P → SN SV
 (↑ sujet)=(↓) (↑)=(↓)

Mais LFG comporte beaucoup d'outils qui facilitent différentes analyses syntaxiques :²

- * les équations de contrainte qui sont utilisées pour garantir une valeur à un trait, sans spécifier cette valeur,
- * les valeurs de type ensemble qui permettent à certains traits de prendre comme valeur des ensembles de structures de traits,
- * la disjonction,

¹ Cf. *op. cit.* [Miller-Torris91], p. 67.

² *Ibid.*, p. 69.

- * les expressions régulières qui apparaissent dans le squelette non contextuel,
- * les métavariabes de longue distance qui sont utilisées dans l'analyse du déplacement de qu- et d'autres dépendances non bornées,
- * les règles lexicales qui expriment des généralisations lexicales concernant des mots apparentés.

LFG est également limitée par une série de contraintes formelles :

- * l'analysabilité en temps différé qui interdit les dérivations circulaires dans le squelette des règles non contextuelles,
- * la localité fonctionnelle qui bannit des règles les chemins longs de plus de deux traits.

3.2.2.4.4.2. Principes et théorèmes

Les trois principes auxquels doivent répondre les structures fonctionnelles sont : ¹

- * le principe d'unicité : un même attribut ne peut apparaître deux fois dans une même sous-structure f ; ce principe est pris en charge par l'unification.
- * le principe de cohérence : toutes les sous-structures doivent être localement cohérentes : les fonctions sous-catégorisables doivent toutes être gouvernées par un prédicat local, il permet le contrôle de la sous-catégorisation.
- * le principe de complétude : toutes les sous-structures doivent être localement complètes : elles doivent comporter toutes les fonctions gouvernées par un prédicat local (permet le contrôle de la sous-catégorisation).

Les fonctions grammaticales jouant un rôle sémantique doivent comprendre obligatoirement un trait Pred, elles ne peuvent pas être réalisées par un élément lexical sémantiquement vide. En revanche une fonction grammaticale ne jouant pas de rôle thématique peut comprendre ou non un prédicat sémantique : *Jean semble dormir* => la fonction sujet de *semble* ne joue pas de rôle thématique et pourtant lui correspond (en vertu du principe de contrôle) la même structure f que celle du sujet de *dormir* avec un trait Pred = 'Jean'.

3.2.2.4.4.3. Tableau comparatif notation traditionnelle / notation LFG

Catégories traditionnelles	Représentation en LFG
Sujet	SUBJ
Objet	OBJ
Complétives	COMP
Infinitives et participiales	XCOMP
Compléments prépositionnels	Prep-Obj ou OBL _θ

TABLEAU 16 : NOTATION TRADITIONNELLE VS NOTATION LFG

¹ Cf. *op. cit.* [Abeillé93], pp. 61-62.

3.2.2.4.4.5. Le traitement des prépositions dans LFG

Il se pose un problème pour la sous-catégorisation des compléments prépositionnels. Il y a deux raisons à cela :¹

- * en LFG, les syntagmes prépositionnels (SP) introduits par des prépositions différentes ont des fonctions grammaticales différentes : un SP introduit par *à* aura comme fonction A-Obj (ou Oblà), un SP introduit par *de* De-Obj (ou Oblde),
- * « la valeur lexicale de la préposition dépend du verbe principal, tout en étant introduite par une autre règle syntagmatique que celui-ci ». Comme la valeur de la préposition n'apparaît pas directement dans la règle où figure le SP, on utilise l'attribut Pcas, noté dans l'entrée lexicale de chaque préposition (pouvant introduire un complément sous-catégorisé) qui prend pour valeur le nom de la fonction grammaticale du SP correspondant.

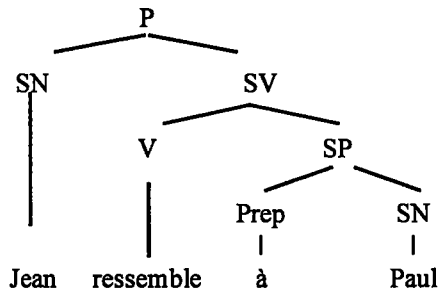
Ex. : Jean ressemble à Paul

Prep → à

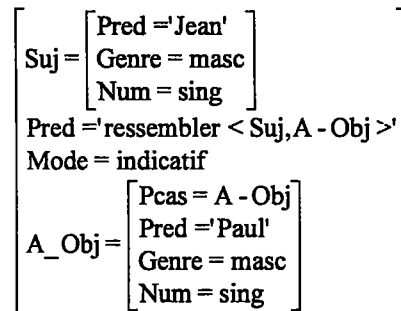
(↑Pcas)=A-Obj

Pred = 'ressembler<Suj,A-Obj>'

Structure c :



Structure f :



Exemple d'équation de contrôle du trait Cas :

$$\begin{aligned}
 P &\rightarrow \text{SN} && \text{V} \\
 &(\uparrow\text{Suj}) = \downarrow && \uparrow = \downarrow \\
 &(\downarrow\text{Cas}) \text{ nominatif} \\
 \text{ou} &(\uparrow\text{Suj Cas}) = \text{nominatif}
 \end{aligned}$$

¹ *Ibid.*, pp. 57-58.

On peut faire une distinction selon que la préposition est suivie d'un SN ou d'une infinitive. On a alors deux solutions : ¹

- on ajoute les catégories concernées dans la règle syntagmatique du SP (le symbole / indique la disjonction entre catégories) :

$$\begin{array}{l} \text{SP} \rightarrow \text{Prep} \quad \text{SN/SV} \\ \quad \uparrow = \downarrow \quad \uparrow = \downarrow \\ \quad \sim(\downarrow \text{Pred}) \end{array}$$

- on dédouble la règle du SP en assignant une fonction Obj au SN et une fonction Vcomp au SV interne, et on réduit en conséquence la valeur du trait Pcas au nom de la préposition :

$$\begin{array}{l} \text{SP} \rightarrow \text{Prep} \quad \text{SN} \\ \quad \uparrow = \downarrow \quad (\uparrow \text{Obj}) = \downarrow \\ \text{SP} \rightarrow \text{Prep} \quad \text{SV} \\ \quad \uparrow = \downarrow \quad (\uparrow \text{Vcomp}) = \downarrow \\ \quad \quad \quad (\downarrow \text{Mode}) = \text{inf} \\ \text{Prep} \rightarrow \quad \quad \text{à} \\ \quad \quad \quad (\uparrow \text{Pcas}) = A \end{array}$$

Exemple : Paul rêve de ce voyage/de faire ce voyage

$$(\uparrow \text{Pred}) = \text{'rêver<Suj,De-Obj/De-Vcomp>}'$$

$$(\uparrow \text{De-Vcomp Mode}) = \text{inf}$$

Il n'y a pas d'isomorphisme [◇] entre structure c et structure f : en structure c, les nœuds occupés par des marqueurs syntaxiques, c'est-à-dire ceux qui n'ont pas de fonction grammaticale et dont les traits sont fondus avec ceux d'une autre catégorie n'ont pas de structure f associée qui leur soit propre. C'est le cas pour les prépositions marqueurs d'arguments (resp. des déterminants) dont les traits sont simplement unifiés avec ceux du SN (resp. du N) qui les suivent.

Une conférence sur LFG sera organisée par C. Manning et J. Simpson, du 30 juin au 02 juillet 1998, à l'Université de Queensland en Australie.

3.2.2.4.4.6. Conclusion

Les hypothèses architecturales de LFG ont conduit à rechercher et à découvrir de nouveaux moyens informatiques pour traiter des expressions en langue naturelle. De nouveaux algorithmes de résolution des contraintes disjonctives (Maxwell & Kaplan, 1989) ont été développés, de même que des stratégies d'analyse plus efficaces (Maxwell & Kaplan, 1993) qui tirent avantage du parcellement de l'information linguistique dans différents niveaux de représentation.

¹ Cf. *op. cit.* [Abeillé93], pp. 58-59.

3.2.2.4.5. Grammaire syntagmatique généralisée (GPSG, *Generalized Phrase Structure Grammar*)

3.2.2.4.5.1. Naissance, motivations de départ et objectifs

Ce formalisme est élaboré au début des années 80 par G. Gazdar, E. Klein, G. Pullum, I. Sag. Leur motivation de départ est la construction d'un modèle équivalent à une grammaire hors contexte. Leurs objectifs sont de définir un modèle syntaxique sans transformations, de montrer que l'on peut construire une grammaire (pour une langue naturelle) appartenant formellement à la classe des grammaires hors contexte, à partir du constat qu'il n'y a pas (ou fort peu) de propriétés dans les langues qui nécessitent un modèle plus puissant.

C'est un modèle syntagmatique, c'est-à-dire basé sur les règles de réécriture des grammaires hors contexte, généralisé, donc utilisant des règles volontairement sous-spécifiées auxquelles s'appliquent des principes généraux, et de compétence neutre, par conséquent indépendant de toute considération sur la production par le locuteur ou sur la compréhension par l'auditeur. Il doit pouvoir servir aussi bien en synthèse qu'en analyse.

Il évolue en 1985 : ce n'est plus un modèle procédural ° mais un modèle déclaratif °, se présentant comme un ensemble ordonné d'assertions indépendantes les unes des autres. La syntaxe prend la place centrale.

3.2.2.4.5.2. Organisation du modèle

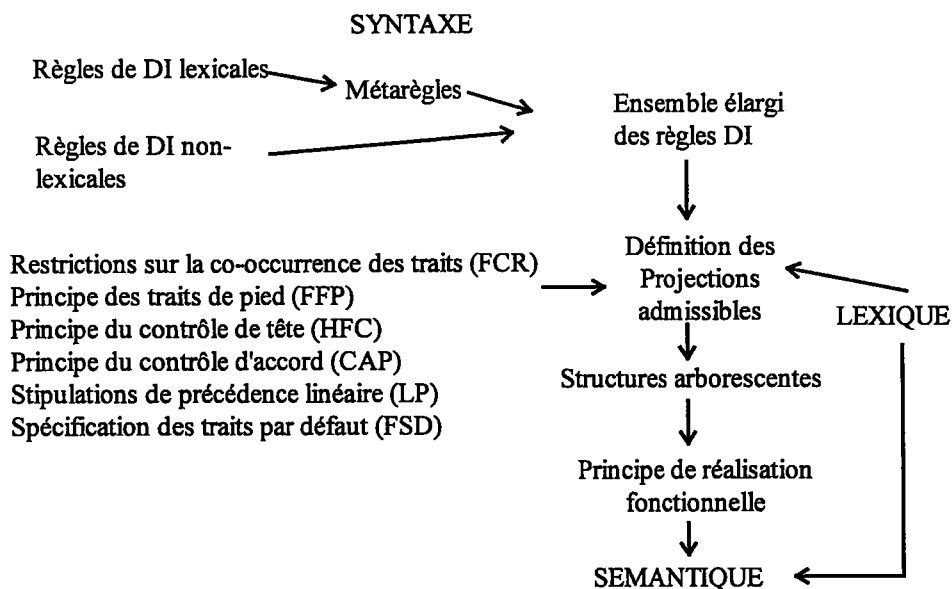


FIGURE 10 : ORGANISATION DU MODELE GPSG ¹

¹ Cette figure est une combinaison des figures proposées dans *op. cit.* [Miller-Torris91], p. 93 et *op. cit.* [Baschung91], p. 104.

3.2.2.4.5.3. Notations utilisées dans GPSG

X ² : syntagme	N ² : syntagme nominal
⇒ : produit d'une métarègle	/ : trait SLASH
{ } : ensemble de traits (catégories)	< : précède
T : catégorie tête	A ² : syntagme adjectival
Prep ² : syntagme prépositionnel	V ² : syntagme verbal

3.2.2.4.5.4. Le formalisme GPSG

3.2.2.4.5.4.1. Description du modèle

C'est un formalisme explicite et restrictif du point de vue formel, tout en ayant une large couverture empirique. Les catégories syntaxiques sont décrites grâce à des structures de traits. Il définit de manière explicite le rapport entre la syntaxe et une composante de sémantique formelle inspirée des travaux de Montague (sémantique compositionnelle qui associe une règle d'interprétation sémantique à chaque règle syntaxique de la grammaire), il est basé sur la logique intensionnelle.¹

C'est un modèle à un seul niveau de représentation syntaxique. Il existe un lien étroit entre l'analyse syntaxique et l'interprétation sémantique : « Chaque structure syntaxique est directement associée à une interprétation théorique du modèle et définit de manière déterministe les compositions de fonctions dans la sémantique ». ²

GPSG repose sur une base syntagmatique non contextuelle, la concaténation est utilisée uniquement pour former les chaînes linguistiques de surface. Le système traits/valeurs est restreint. Les traits peuvent avoir une valeur atomique ou une valeur complexe.

3.2.2.4.5.4.1.1. Les catégories

Pour GPSG, les catégories sont entièrement réductibles à des ensembles de traits (couples attribut-valeur). Les traits sont utilisés pour coder des informations linguistiques, ils sont distingués par leur type de valeur (atomique (les plus utilisés) ou catégorielle (notamment Slash utilisé pour les phénomènes d'extraction et AGR qui règle les phénomènes d'accord et de contrôle)) et par leur mode de propagation (traits de Tête, traits de Pied, ou traits de Contrôle). Plus le nombre de traits est grand dans une catégorie (catégorie = ensemble de traits non ordonnés notés entre accolades), plus la description est précise et inversement ; l'ensemble vide est la catégorie la moins spécifiée qui s'unifie avec toutes les autres.³

GPSG est une grammaire à traits. Une catégorie syntaxique est un ensemble de paires <trait, valeur>, où le trait est un symbole atomique et où la valeur est soit un symbole atomique soit une catégorie. GPSG adopte une version à deux niveaux de la théorie X-barre : les catégories majeures N, V, A & P sont définies en fonction des traits primitifs [N] et [V] ; NP se représente sous la forme : {<N,+>, <V,->, <BAR,2>}, S et VP se

¹ Cf. *op. cit.* [Miller-Torris91], p. 24.

² Cf. *op. cit.* [Baschung91], p. 103.

³ Cf. *op. cit.* [Abeillé91], pp. 101-103.

distinguent par rapport au trait [SUBJ] et se représentent respectivement de la façon suivante : $\{ \langle N, - \rangle, \langle V, + \rangle, \langle \text{BAR}, 2 \rangle, \langle \text{SUBJ}, + \rangle \}$ et $\{ \langle N, - \rangle, \langle V, + \rangle, \langle \text{BAR}, 2 \rangle, \langle \text{SUBJ}, - \rangle \}$. Les catégories mineures sont des catégories non spécifiées pour le trait [BAR] noté $\sim[\text{BAR}]$.¹

Les catégories syntaxiques ne sont pas traitées comme des atomes mais comme des ensembles de traits syntaxiques. La métavariable *tête* désigne la tête d'un syntagme et reflète son statut de primitif syntaxique.

Il y a deux relations importantes entre les catégories :

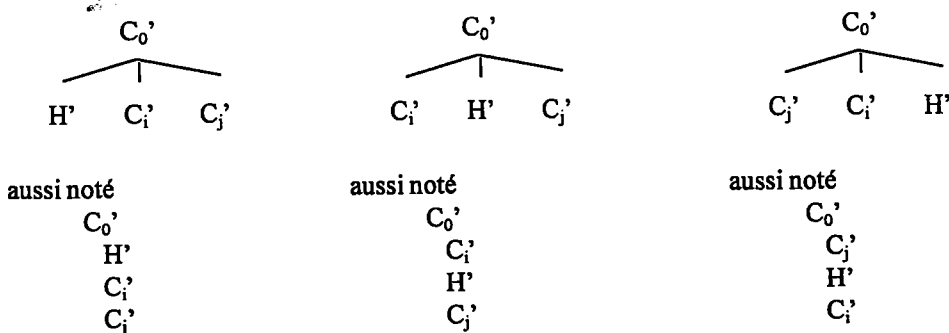
- * l'extension : une catégorie A est une extension de la catégorie B ($B \subseteq A$) si et seulement si :
 - a. les valeurs atomiques de B sont toutes dans A, et
 - b. pour tout trait à valeur catégorielle, la valeur de t dans A est une extension de la valeur de t dans B.
- * l'unification : l'unification d'un ensemble de catégories K ($\cup K$) est la plus petite catégorie qui est une extension de tous les membres de K.

3.2.2.4.5.4.1.2. Les règles

Ce sont des règles de réécriture qui opèrent sur des catégories, elles se divisent en règles de dominance immédiate² (*Immediate Dominance Rules*, règles ID) et en règles de précedence linéaire (*Linear Precedence Statements*, règles LP).

3.2.2.4.5.4.1.2.1. Les règles ID

Les règles complexes sont composées d'arbres locaux simples constitués chacun d'une catégorie mère et d'une ou plusieurs catégories filles, et chacun est associé à une règle ID différente. La catégorie située à gauche de la règle est appelée catégorie mère, celles situées à droite catégories filles. Les règles ID indiquent seulement les relations de dominance entre la catégorie mère (notée C_0) et les catégories filles (notées C_i, \dots, C_n), dont une est notée H (la tête) et a des propriétés particulières. Les catégories filles ne sont pas ordonnées et on obtient donc plusieurs arbres locaux ; $C_0 \rightarrow H, C_i, C_j$ donne, par exemple, les arbres suivants :³



¹ Cf. *op. cit.* [Baschung91], pp. 104-105.

² K. Baschung les appelle *règles de domination immédiate*.

³ Cf. *op. cit.* [Baschung91], p. 106.

Les règles ID lexicales sont des règles dont la tête H va dominer un symbole terminal (une entrée du lexique), c'est-à-dire qui ont pour tête une catégorie lexicale. Une règle est dite « non lexicale » quand aucune tête lexicale n'apparaît sur son côté droit.

Une entrée lexicale est représentée par un quadruplet dont le premier élément contient l'information phonologique, le deuxième l'ensemble des traits syntaxiques, le troisième l'information morphologique et le quatrième la forme sémantique. Une entrée lexicale a toujours le trait [BAR,0] et a toujours une valeur pour le trait [SUBCAT] (pour *subcategorization*, sous-catégorisation). Cette valeur est un nombre et fonctionne comme un pointeur sur une règle, celle-ci fournit, en tant que règle lexicale, les arguments pour lesquels sa tête doit être sous-catégorisée.

Les propriétés des règles sont les suivantes : ¹

- * les règles ID sont des règles qui ne peuvent référer au contexte,
- * ce sont des règles de dominance pure, indépendantes des règles de précedence linéaire,
- * elles sont limitées à la dominance immédiate et expriment donc des relations strictement locales entre catégories mères, filles et sœurs,
- * elles mentionnent des catégories syntaxiques partiellement spécifiées et conçues comme des ensembles de traits syntaxiques complexes,
- * elles sont considérées comme des conditions d'acceptabilité sur les structures,
- * elles forment un ensemble structuré par de puissants mécanismes d'abréviation, tels que les métarègles,
- * les métarègles s'appliquent seulement aux règles de DI lexicales ».

3.2.2.4.5.4.1.2.2. Les règles LP

Les règles de précedence linéaire ² appartiennent à une sous-composante grammaticale distincte, réfèrent à un trait syntaxique, et vont ordonner les catégories filles dans les arbres locaux. Elles sont de la forme : $Cat_1 < Cat_2$.

« Les règles OL doivent s'appliquer aux catégories (en partie droite) de toutes les règles DI d'une grammaire donnée, c'est-à-dire qu'elles spécifient un ordre entre catégories de même niveau syntagmatique [à savoir, entre catégories sœurs]. Elles n'expriment pas directement l'ordre observé sur les catégories préterminales ou les mots des phrases ». ³

Les deux principales sont les suivantes :

- a. [SUBCAT] < ~ [SUBCAT] signifie qu'une tête lexicale doit toujours précéder ses compléments, un complémenteur doit toujours précéder la phrase dont il est la sœur.
- b. [+N] < P₂ < V₂ règle l'ordre entre les syntagmes dans VP.

¹ Cf. *op. cit.* [Miller-Torris90], p. 100.

² K. Baschung les appelle *stipulations de précedence linéaire*, et A. Abeillé *règles d'ordre linéaire*.

³ Cf. *op. cit.* [Abeillé91], p. 113.

3.2.2.4.5.4.1.2.3. Le format ID /LP

On appellera les règles ID et les règles LP le format ID/LP de la grammaire.

De :	on passe à (en GPSG)
P → SN SV Adv	P → SN, T{Suj=-}, Adv [règle DI]
P → Adv SN SV	SN<SV (< signifie « précède ») [règle OL]
P → SN Adv SV	ex. : <i>Jean viendra probablement</i>

Un sous-arbre local est une projection d'une règle DI si tous ses nœuds sont des extensions des catégories correspondantes de la règle et respectent les règles LP de la grammaire. Si les sous-arbres qui sont une projection d'une règle DI obéissent aux principes généraux d'instanciation de traits, alors ils sont les arbres bien formés de la grammaire.

On trouve deux types de traits dans un arbre :

- * les traits « hérités » : ceux que la règle DI correspondante mentionne explicitement,
- * les traits « instanciés » : ceux qui sont ajoutés aux nœuds par les entrées lexicales ou selon des principes généraux.¹

On peut noter des différences entre les règles GPSG et les règles syntagmatiques classiques :

- * les relations structurelles de dominance et les relations de précedence linéaire sont dissociées et traitées par deux types de règles distincts ; elles sont donc encodées séparément dans la grammaire,
- * les catégories syntaxiques ne sont plus traitées comme des atomes mais décomposées en ensembles de traits syntaxiques définis par trois couples attribut-valeur <N,->, <V,-> et <SOUSCAT,n>,
- * toutes les catégories sont des ensembles de traits et non des symboles,
- * toutes les règles ont dans leur partie droite une catégorie considérée comme la tête (notée T) de la catégorie en partie gauche (la notion de tête est héritée de la théorie X-barre), pour GPSG, le SV est la tête de la phrase et le verbe la tête du SV, on le note de la façon suivante :

$$P \rightarrow SN, T\{Suj=-\}$$

$$SV \rightarrow T\{Sous-cat=2\}, SN$$
- * les catégories de la partie droite ne sont pas ordonnées (d'où leur notation avec des virgules).²

GPSG code la sous-catégorisation dans les règles et pas dans le lexique. En GPSG, la sous-catégorisation est un phénomène syntaxique qui ne concerne que les compléments ; elle s'exprime en termes de catégories, dans un sous-arbre local. Les verbes intransitifs ont le trait Sous-cat=1, et les verbes transitifs Sous-cat=2.

¹ Cf. *op. cit.* [Abeillé91], p. 115.

² *Ibid.*, pp. 109-110.

Les éléments mobiles (adverbes par exemple) posent problème pour les règles de réécriture classiques (contextuelles ou hors contexte) car il faut écrire autant de règles que l'adverbe peut avoir de position.

3.2.2.4.5.4.1.2.4. Les métarègles

Pour exprimer certaines généralisations syntaxiques la notion de métarègles est introduite pour relier un ensemble de règles ID à un autre ensemble de règles ID. Les métarègles sont utilisées pour le passif, les constructions impersonnelles... mais elles sont inutilisables pour l'accord, la « montée du sujet », la « montée de l'objet ». Les métarègles sont des mécanismes qui s'appliquent aux règles de la grammaire pour en produire de nouvelles. Elles sont des fonctions partielles de l'ensemble des règles DI lexicales vers lui-même. Exemple de métarègle pour le passif : $SV \rightarrow W, SN \Rightarrow SV[PASS] \rightarrow W, (SP[par])$.¹ Les métarègles opèrent uniquement sur des règles ID et permettent d'en obtenir de nouvelles à partir de celles qui existent déjà, elles expriment une généralisation sur des ensembles de règles dont le domaine d'application est la sous-catégorisation lexicale.²

3.2.2.4.5.4.2. Principes et théorèmes

La projection est réglée par des principes universaux d'instanciation qui conditionnent les relations entre les catégories d'une règle et les catégories correspondantes dans un arbre local, et les relations entre catégories à l'intérieur d'un même arbre.

Les projections admissibles régissent l'interaction des règles et des principes d'instanciation. Quand tous les principes sont satisfaits, on peut établir l'admissibilité d'un arbre par rapport à l'ensemble des règles ID d'une grammaire, selon la définition suivante :

Étant donné un ensemble R de règles ID, un arbre t est admissible pour R si, et seulement si :

- a. t est terminé (un arbre est terminé si toutes ses feuilles sont des éléments du vocabulaire terminal),
- b. tout sous-arbre local dans t est soit un arbre terminé, soit une projection localement admissible d'une règle r appartenant à R .³

Il faut exprimer les contraintes d'unification de façon plus globale qu'en LFG grâce à des principes généraux d'instanciation des traits, qui sont de deux types :

- * soit ils expriment des rapports d'équivalence, d'implication ou d'incompatibilité entre traits d'une même catégorie (contraintes de bonne formation sur les catégories),

¹ Cf. *op. cit.* [Miller-Torris90], p. 97.

² Cf. *op. cit.* [Baschung91], p. 108.

³ Cf. *op. cit.* [Miller-Torris90], p. 112.

- * soit ils expriment des relations entre les traits appartenant aux différentes catégories d'une même règle (contraintes de bonne formation sur les arbres associés aux règles syntagmatiques de la grammaire).

En GPSG, il existe trois classes de traits FOOT, HEAD & CONTROL pour assurer la percolation adéquate des traits dans les arbres.¹ Les principes généraux d'instanciation des traits sont :

- * le principe des traits de Pied (*Foot Feature Principle, FFP*) : il « assure le partage des informations entre la catégorie racine et un élément dominé situé à un niveau de profondeur quelconque ». Les traits de pied hérités ne sont pas concernés par ce principe.

$$\text{FOOT} = \{\text{WH, SLASH, RE}\}$$

où WH indique si le constituant est interrogé ou relativisé,

SLASH indique si le constituant est omis,

RE indique si le constituant est réflexif ou réciproque.

Notation du principe : $\phi(C_0) \mid \text{FOOT} \sim C_0 \cup \phi(C_i) \mid \text{FOOT} \sim C_i$

$$1 \leq i \leq n$$

ϕ représente l'instanciation.

Signification : « Les traits de Pied instanciés dans la catégorie mère d'un arbre local doivent être identiques à l'unification des traits de Pied instanciés dans les catégories filles ».

- * le principe des traits de Tête (*Head Feature Convention, HFC*) : le but est de faire partager des traits entre la catégorie dominante (ou catégorie mère) en partie gauche des règles, et une catégorie fille privilégiée désignée comme Tête de la règle.² Il assure la percolation des traits entre la catégorie mère d'une règle et la catégorie fille qui en est la tête.

$$\text{HEAD} = \{\text{N, V, PLU, PER, VFORM, PFORM, AUX, ADV, SUBJ, INV, PAST, PRD, SLASH, AGR, SUBCAT, BAR, LOC}\}^3$$

Notation simplifiée du principe : $\phi(C_0) \mid \text{HEAD} = \phi(C_h) \mid \text{HEAD}$

Signification : « Les traits de Tête de la catégorie mère (dans l'arbre local) doivent être identiques aux traits de Tête de la catégorie fille qui en est la tête ».

Comme le dit Karine Baschung :

« Il faut assurer cette identité des traits dans les limites de la bonne formation des catégories, ce qui est obtenu au moyen d'une définition des traits dits libres : Soit C_i une catégorie dans une règle r , et

¹ Cf. *op. cit.* [Baschung91], p. 111 sqq.

² Cf. *op. cit.* [Abeillé91], p. 104.

³ Cf. *op. cit.* [Baschung91], p. 113.

soit Ψ_r l'ensemble des projections de r ; l'ensemble des traits libres de C_i par rapport à Ψ_r est $\Psi(C_i, \Psi_r)$, où $\Psi(C_i, \Psi_r) = \{ \langle t, v \rangle : \exists \phi \subset \Psi_r [\langle t, v \rangle \in \phi(C_i)] \}$; ce qui signifie que les traits libres d'une catégorie sont susceptibles d'apparaître dans une extension de celle-ci, c'est-à-dire en satisfaisant les FCR, FFP et CAP ». ¹

Notation du principe : a. $(\phi(C_0) \cap \Psi(C_h, \Phi_r) \mid \text{HEAD} \subseteq \phi(C_h) \mid \text{HEAD}$

b. $(\phi(C_h) \cap \Psi(C_0, \Phi_r) \mid \text{HEAD} \subseteq \phi(C_0) \mid \text{HEAD}$

Φ_r représente l'ensemble des projections de r .

Signification : a. « Les traits de la Tête de la catégorie sont une extension du résultat de l'intersection des traits de Tête de la mère avec les traits de Tête libres de la Tête »,

b. « Les traits de Tête de la catégorie mère sont une extension du résultat de l'intersection des traits de Tête de la catégorie Tête avec les traits de Tête libres de la mère ».

* le principe du contrôle et de l'accord ² (*Control Agreement Principle, CAP*) : il assure l'accord morphologique sujet-verbe, fait en sorte que les éléments extraits correspondent à la position d'où ils sont extraits, et traite le contrôle des infinitives. Ce principe entraîne une forte interaction entre la syntaxe et la sémantique. Rappelons que GPSG utilise une version de la sémantique compositionnelle et vériconditionnelle de Montague (où on retrouve un parallélisme strict entre la syntaxe et la sémantique). ³

Principe général de l'application fonctionnelle : Si α est une expression de type $\langle b, a \rangle$ et b une expression de type $\langle b \rangle$, $\alpha(b)$ est une expression de type $\langle a \rangle$.

Les types⁹ sémantiques devant déterminer les relations de contrôle sont les types associés avec les spécifications χ des catégories.

$\chi(\phi(C_i)) = \phi(C_i) \mid ((\text{HEAD-FOOT}) \cup (\text{DOM}(C_i) \mid \text{FOOT}))$

Les spécifications χ d'une catégorie dans un arbre local sont ses traits de Tête moins ceux qui sont aussi des traits de Pied (c'est-à-dire SLASH) augmentés des traits de Pied hérités pouvant comprendre SLASH le cas échéant).

Contrôle : si ϕ est une projection de r , où $r = C_0 \rightarrow C_1, \dots, C_n$, alors une catégorie $\phi(C_i)$ contrôle une catégorie $\phi(C_j)$ en ϕ , $1 \leq i, j \leq n$, si et seulement si :

(i) $\text{TYP}(\chi(\phi(C_j))) = \langle \text{TYP}(\chi(\phi(C_i))), (\chi(\phi(C_0))) \rangle$, ou bien

(ii) $\text{TYP}(\chi(\phi(C_j))) = \text{TYP}(VP)$ et l'un des types associés à la tête de r est : $\langle \text{TYP}(VP), \langle \text{TYP}(\chi(\phi(C_i))), \text{TYP}(VP) \rangle \rangle$.

¹ *Ibid.*

² Traduction donnée par Miller et Thoris dans *op. cit.* [Miller-Torris90], p. 354.

³ Cf. *op. cit.* [Baschung91], p. 114 sqq.

Le sous-ensemble des traits relatifs au contrôle est : CONTROL = {AGR, SLASH}.

Trait de contrôle : Soient C_i une catégorie dans une règle r telle que $C_i(\text{BAR}) \neq 0$ et ϕ une projection de r ; un trait t est le trait de contrôle de $\phi(C_i)$ si et seulement si :

- (a) $t = \text{SLASH}$ et $t \in \text{DOM}(C_i)$, ou
- (b) $\text{SLASH} \in \text{DOM}(C_i)$ et $t = \text{AGR}$.

Notation et signification du principe : Soit Φ_r , l'ensemble des projections de r , où $r = C_1, \dots, C_n$. $\phi \in \Phi_r$ est satisfaisable CAP par rapport à r si et seulement :

- 1- si $\phi(C_j)$ contrôle $\phi(C_i)$ alors $\phi(C_i)(t_i) = \chi(\phi(C_j)) \cup \phi(C_j) \mid \{t_i\}$ où t_i est le trait de contrôle de $\phi(C_i)$ ce qui signifie que si une catégorie C_j « contrôle » une catégorie C_i , alors le trait de Contrôle de C_i a pour valeur C_j (ceci est utilisé pour les phénomènes d'accord, pour les phénomènes de contrôle où c'est l'objet qui « contrôle » le sujet du groupe infinitif),
- 2- s'il existe une catégorie prédicative sans contrôleur, alors $(C_i)(t_i) = \phi(C_0)(t_0)$, où t_i et t_0 sont les traits de contrôle de $\phi(C_i)$ et de $\phi(C_0)$ respectivement, ce qui signifie que si C_i est une catégorie prédicative sans contrôleur, son trait de Contrôle a même valeur que celui de sa catégorie mère (ceci est utilisé pour régler le cas des verbes à contrôleur sujet, souvent des verbes dont le seul complément est une infinitive (espérer, vouloir, pouvoir, sembler)),
- 3- le trait de Contrôle est le trait Slash s'il est hérité dans la règle, sinon c'est le trait AGR (le trait Slash n'est considéré comme « hérité » que dans les règles qui le mentionnent explicitement, sinon c'est AGR qui est le trait de contrôle) ».

Les deux principes qui règlent la bonne formation des catégories sont : le principe de restriction sur la cooccurrence des traits (*Feature Co-occurrence Restrictions, FCR*) et le principe de spécification de traits par défaut (*Feature Specification Default, FSD*).

- * le principe de restrictions sur la cooccurrence des traits permet d'exprimer le fait que certains traits sont incompatibles entre eux alors que d'autres sont nécessairement liés. Ces restrictions restreignent la classe des catégories possibles dans une langue donnée. Elles sont par exemple de la forme $[+A] \supset [-B]$, ce qui signifie que toute catégorie qui comporte le trait A avec la valeur + comporte obligatoirement le trait B avec la valeur -. Certaines FCR sont présumées universelles (cf. ci-dessous), d'autres sont à paramétrer en fonction de la langue qu'il faut décrire.

FCR 2 : [VFORM] \supset [-N,+V]

FCR 3 : [NFORM] \supset [+N,-V]

FCR 4 : [PFORM] \supset [-N,-V]

FCR 7 : [BAR 0] \equiv [N] & [V] & [SUBCAT]

FCR 8 : [BAR 1] \supset \sim [SUBCAT]

FCR 9 : [BAR 2] \supset \sim [SUBCAT]

où \supset représente l'implication, \sim la négation et \equiv signifie « si et seulement si ».¹

- * les spécifications de traits par défaut expriment des contraintes absolues ou relatives (conditionnelles) sur la valeur des traits dans une catégorie (ex. $\{V=+,N=-,Barre=0\} \Rightarrow \{AGR=SN\}$ = un verbe a un sujet nominal). Ce sont des exigences sur la cooccurrence des traits dans les catégories, elles ne sont valables que dans les cas où la grammaire n'exprime pas explicitement le contraire.

Une projection d'une règle r qui satisfait FFP, HFC, CAP et LP est admissible si et seulement si :

- a. les FSDs y sont satisfaites, ou bien,
- b. il n'existe aucune projection de r où les FSDs soient respectées, ou bien,
- c. la satisfaction des FSDs entraînerait la modification d'une (ou plusieurs) autre(s) catégorie(s) dans la projection de r .

3.2.2.4.5.4.3. *Tableau comparatif notation traditionnelle / notation GPSG*

K. Baschung propose, page 105, un récapitulatif de la notation GPSG :

S	{<N,->,<V,+>,<BAR,2>,<SUBJ,+>}
XP	{<BAR,2>}
X2	{<BAR,2>}
X1	{<BAR,1>}
V	{<N,->,<V,+>}
VP	{<N,->,<V,+>,<BAR,2>,<SUBJ,->}
V2	{<N,->,<V,+>,<BAR,2>}
N	{<N,+>,<V,->}
NP	{<N,+>,<V,->,<BAR,2>}
N2	{<N,+>,<V,->,<BAR,2>}
N1	{<N,+>,<V,->,<BAR,1>}
P	{<N,->,<V,->}
PP	{<N,->,<V,->,<BAR,2>}
P2	{<N,->,<V,->,<BAR,2>}
P1	{<N,->,<V,->,<BAR,1>}
NP[ACC]	{<N,+>,<V,->,<BAR,2>,<CASE,ACC>}
NP[+PLU]	{<N,+>,<V,->,<BAR,2>,<PLU,+>}
S/NP	{<N,->,<V,+>,<BAR,2>,<SUBJ,+>,<SLASH,<N,+>,<V,->,<BAR,2>>}
VP[AGR[NP]]	{<N,->,<V,+>,<BAR,2>,<SUBJ,->,<AGR,<N,->,<V,->,<BAR,2>>}

TABLEAU 17 : LA NOTATION GPSG SELON K. BASCHUNG

¹ Cf. *op. cit.* [Baschung91], p. 110.

De son côté, Anne Abeillé résume, page 105, l'équivalence entre la notation traditionnelle et les notations GPSG dans le tableau suivant :

Catégories	Abréviation	Traits	Contraintes
Verbes	V	V=+,N=-,Barre=0	Mode
Groupes verbaux	SV ou V ²	V=+,N=-,Barre=2,Suj=-	Mode,~Sous-cat
Phrases	P ou V ²	V=+,N=-,Barre=2,Suj=+-	Mode,Compform,~Sous-cat
Noms	N	V=-,N=+,Barre=0	~Mode
Pronoms Groupes nominaux	SN ou N ²	V=-,N=+,Barre=2	~Sous-cat
Prépositions	Prep	V=-,N=-,Barre=0	Prepform
Adverbes Groupes prép.	SP ou Prep ²	V=-,N=-,Barre=2	Prepform, ~Sous-cat
Adjectifs	A	V=+,N=+,Barre=0	~Mode
Groupes adject.	SA ou A ²	V=+,N=+,Barre=2	~Sous-cat
Complémenteurs	Comp	Compform	~Barre,~Sous-cat,~V,~N
Déterminants	Det	Det=+	~V,~N,~Sous-cat, ~Barre

~ = négation,
 Prepform = valeur lexicale de la préposition concernée
 Compform = valeur lexicale du complément concerné
 Barre = niveau de projection

TABLEAU 18 : NOTATION TRADITIONNELLE VS NOTATION GPSG

On constate que les deux auteurs n'ont pas adopté la même notation (même si la signification est identique) et qu'il existe en plus des traits différents dans les deux tableaux.

3.2.2.4.5.5. Le traitement des prépositions dans GPSG

La sous-catégorisation des groupes prépositionnels fonctionne de la même façon que celle des verbes, si ce n'est qu'il faut spécifier la valeur lexicale de la préposition en utilisant l'attribut Prepform, noté dans les entrées lexicales des prépositions et dans les règles syntagmatiques.

SV → {Sous-cat=3}, SP {Prepform=à}
 à : {V=-,N=-,Barre=0,Prepform=à}
 ressembler : {V=+,N=-,Barre=0,Sous-cat=3}.

En ce qui concerne la représentation des groupes prépositionnels, il est préférable de considérer le groupe nominal (GN) ou le groupe verbal (GV) plutôt que la préposition comme tête des compléments prépositionnels. Le groupe prépositionnel (GP) devient un GN avec un trait Prepform=x, et la préposition devient un simple marqueur.

Ex. : SN {Prepform=x} → Prep {Prepform=x}, T {Prepform=nil}

SV {Prepform=x} → Prep {Prepform=x}, T {Prepform=nil}

Ainsi il est plus simple de faire la différence entre verbes à complément prépositionnel verbal et verbes à complément prépositionnel nominal :

Jean tient / téléphone à Marie
 Jean tient / * téléphone à aller à la plage

1 $SV \rightarrow V, SV \left\{ \begin{array}{l} \text{Prepform} = \text{à} \end{array} \right\}$ pour *tenir* les deux règles fonctionnent, mais pour *téléphoner*
2 $SV \rightarrow V, SN \left\{ \begin{array}{l} \text{Prepform} = \text{à} \end{array} \right\}$ seule la deuxième est possible

Grâce à cette représentation, il est beaucoup plus facile d'exprimer les restrictions de sélection sur les compléments prépositionnels (les traits sémantiques étant des traits de Tête).

3.2.2.4.5.6. Conclusion

En conclusion il faut retenir l'enrichissement des règles syntagmatiques avec :

- * la décomposition des catégories syntaxiques en traits pour éviter la prolifération des catégories (et donc des règles de réécriture),
- * le contrôle de ces traits par des principes généraux de composition et de propagation (basés sur l'unification),
- * le format DI/OL pour exprimer les variations d'ordre des mots dont est susceptible une structure syntaxique donnée,
- * les métarègles pour exprimer les liens entre phrases (et éviter la multiplication de règles indépendantes).

Selon A. Abeillé, il faudrait apporter quelques changements, comme par exemple : augmenter le nombre de traits syntaxiques, augmenter les contraintes de propagation locales, ajouter d'autres contraintes aux métarègles.

3.2.2.4.6. Grammaire syntagmatique guidée par la tête (HPSG, Head Phrase Structure Grammar)

3.2.2.4.6.1. Naissance, motivations de départ, objectifs

Ce formalisme est élaboré au début des années 80 par Carl Pollard et Ivan Sag. Leur motivation de départ est l'intégration plus explicite des différents niveaux d'analyse linguistique (phonétique, syntaxe et sémantique).

Ils se sont inspirés de GPSG, des travaux de Pollard sur la grammaire de tête (*Head Grammar*), de la Théorie du liage et du gouvernement (*Government and Binding Theory*) pour ce qui est de la notion de modularité, et du recours à des principes de liage de contrôle..., de la Grammaire d'unification fonctionnelle pour la représentation uniforme des items lexicaux, des syntagmes et des règles de grammaire sous forme de structures de traits, de la Grammaire lexicale fonctionnelle pour l'enrichissement des cadres de sous-catégorisation et notion de règle lexicale, des grammaires catégorielles pour la saturation progressive des prédicats et recours à une hiérarchie de fonctions grammaticales, et de travaux en logique et en informatique sur le typage et l'héritage.

3.2.2.4.6.2. Organisation du modèle

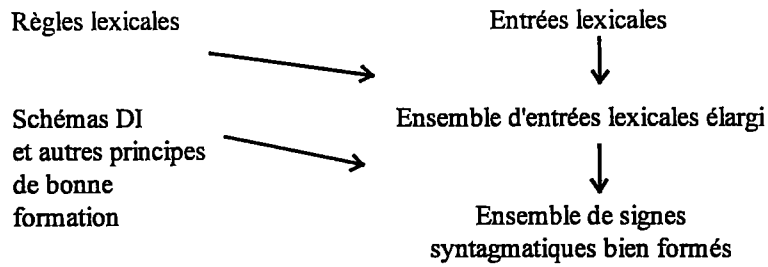


FIGURE 11 : ORGANISATION DU MODELE HPSG ¹

3.2.2.4.6.3. Notations utilisées dans HPSG

- ⇒ : produit d'une règle lexicale
- { } : ensemble de structures de traits
- [] : structure de traits
- ⊕ : concaténation de listes
- <> : liste (ordonnée) de traits ou de structures de traits

3.2.2.4.6.4. Le formalisme HPSG

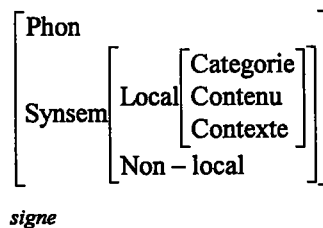
3.2.2.4.6.4.1. Description du modèle

En HPSG, une structure de traits est un signe linguistique, c'est-à-dire un mot, un syntagme ou une règle, comprenant des informations phonétiques, syntaxiques, sémantiques et discursives. L'utilisation systématique de structures de traits typées et à nombre d'attributs fixe permet d'éviter la mise en relation de sous-structures hétérogènes. Les structures de traits autorisent le partage des structures, ce qui évitent les transformations (opérations sur les arbres). Les structures de traits sont donc plus flexibles que les grammaires de transformation. Elles peuvent être hiérarchiquement structurées.

Les propriétés des structures de traits sont les suivantes. Elles sont :

- * réflexives : chaque structure de traits se subsume elle-même,
- * transitives : si A subsume B et B subsume C alors A subsume C,
- * anti-symétriques : si A subsume B et B subsume A alors A = B.

Dans les structures de traits, on a toujours un attribut Phon à valeur phonétique.



attribut Phon ⇒ informations phonétiques

¹ La figure est extraite de *op. cit.* [Abeillé91], p. 149.

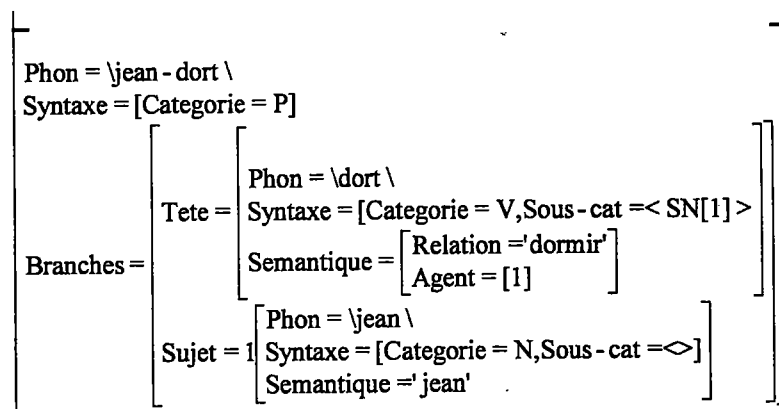
attribut Synsem \Rightarrow informations syntaxiques et sémantiques divisées en :

* traits locaux scindés en :

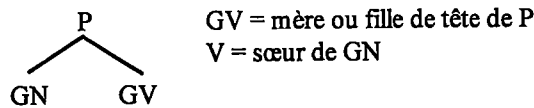
- traits catégoriels (attribut « Catégorie » divisé en : Tête, Sous-cat et Marque),
- traits sémantiques (attributs « Contenu » et « Contexte »),

* traits non-locaux.

Les principaux traits sont donnés dans cette structure mais il en existe d'autres. Pour décrire des structures complexes, on utilise un attribut Branche qui se divise en Tête et Sujet qui prennent pour valeur un signe entier (cf. ci-dessous l'exemple pour une phrase).



L'information de tête est héritée de la mère. Les filles héritent les informations de la mère. L'information de tête est partagée entre la fille et la mère. S'il y a différentes filles, elles ont différentes relations avec leur mère. Le verbe est considéré comme la fille de tête du GV.



Les attributs peuvent prendre différentes valeurs : une valeur atomique (attribut morpho-syntaxique tels que Mode, Personne... ou sémantique comme Relation), un signe (attribut Branche), une valeur non atomique, qui est une structure plus restreinte qu'un signe (attribut Local), une liste ou un ensemble, la valeur est notée entre chevrons <...>, (attribut Sous-cat, attribut Slash) ; la liste vide est notée <>. ¹

Un attribut non local, appelé Slash ², est utilisé pour indiquer qu'un syntagme est « manquant » et propager cette information au niveau des syntagmes supérieurs. Le trait Slash est directement introduit par une règle lexicale qui fait correspondre à un prédicat donné une seconde entrée dans laquelle un élément est supprimé de la liste Sous-cat et apparaît en valeur du trait Slash.

¹ Cf. *op. cit.* [Abeillé91], pp. 153-154.

² *Ibid.*, p. 180.

Une règle lexicale d'« extraction » est de la forme suivante :

$$\begin{aligned} & \Rightarrow \\ \text{Synsem}=\text{Cat}=[\text{Sous-cat}=\langle X, Y[\text{Local}=[1]], \dots \rangle \\ \left[\begin{array}{l} \text{Local} = \text{Cat} = [\text{Sous-cat} = \langle X, \dots \rangle] \\ \text{Non-local} = \text{Hérit} = [\text{Slash} = \{[1]\}] \end{array} \right] \end{aligned}$$

extrac

La coindexation (notée [1]) indique le partage entre les traits locaux de l'élément « manquant » (sous l'attribut Slash) et ceux de l'élément sous-catégorisé (appartenant à la liste Sous-cat du prédicat avant extraction).

Contrairement à GPSG, HPSG n'utilise pas les traits binaires N et V mais un trait Part (parties du discours) qui peut prendre six valeurs : Det, N, V, A, Prep et Marqueur (pour les mots fonctionnels tels que les complémenteurs). Les adverbes sont considérés comme des prépositions sans complément (Part=Prep, Sous-cat= \diamond), les conjonctions de subordination comme des prépositions à complément phrastique.

Ex. SV (ou verbe intransitif) : $\text{Synsem}=\text{Local}=\text{Cat}=\left[\begin{array}{l} \text{Tete} = [\text{Part} = \text{V}] \\ \text{Sous-cat} = \langle \text{SN} \rangle \end{array} \right]$

Bien que basée sur GPSG, HPSG exploite plus le pouvoir de l'unification.

« HPSG revient à la concaténation comme seul mode de combinaison, remplace les métarègles par des règles de redondance lexicale et supprime beaucoup des restrictions qui, dans GPSG, garantissaient un domaine d'information fini. Les appariements entre chaînes et éléments d'information (appelés signes) sont déterminés par un algorithme ascendant d'application des règles ; le formalisme est donc procédural par définition ». ¹

En HPSG, le cadre de sous-catégorisation ordonne le sujet en premier et non en dernier comme dans les autres grammaires.

L'organisation hiérarchique permet d'utiliser un principe d'héritage : chaque sous-type « hérite » de toutes les spécifications du type supérieur (et ainsi de suite jusqu'au sommet de la hiérarchie), c'est-à-dire que les traits ou les attributs qui appartiennent à la description d'un type sont transmis à tous ses sous-types sans qu'on ait besoin de les répéter.

La hiérarchie des types peut être définie à l'aide de la relation de subsomption qui définit un ordre partiel sur les structures de traits. L'organisation hiérarchique des signes linguistiques se présente de la façon suivante :

¹ Cf. *op. cit.* [Miller-Torris90], p. 73.

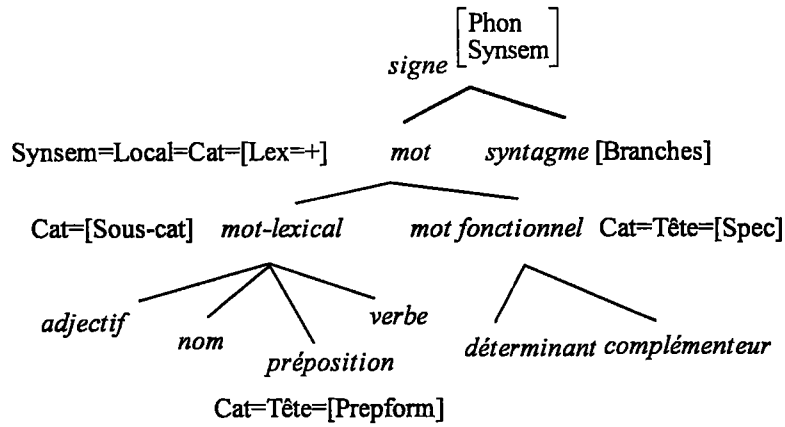


FIGURE 12 : ORGANISATION HIERARCHIQUE DES SIGNES LINGUISTIQUES DANS HPSG ¹

On peut avoir recours à des classifications croisées et à la notion d'héritage multiple, c'est-à-dire qu'un objet doit pouvoir hériter ses spécifications de plusieurs types à la fois.

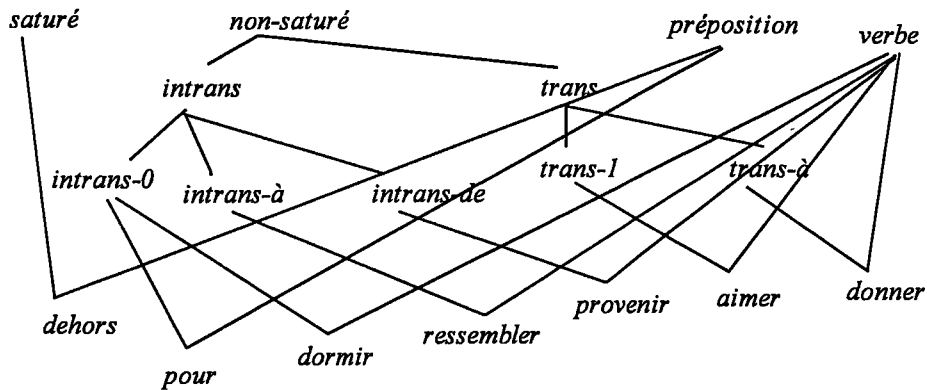


FIGURE 13 : ORGANISATION SIMPLIFIÉE DE QUELQUES SIGNES LEXICAUX (MOTS NON FONCTIONNELS) ²

L'unification s'étend aux structures de traits typées (il faut que les valeurs et les types soient compatibles pour qu'il puisse y avoir unification).

« L'unification de deux structures de traits A et B typées (resp. t et t') a pour résultat, si elle existe, la plus petite structure qui est à la fois une extension de A et une extension de B (comme pour l'unification simple) de type t'', où t'' est le plus grand type qui soit à la fois un sous-type de t et un sous-type de t'. » ³

¹ La figure est extraite de *op. cit.* [Abeillé91], p.159.
² La figure est extraite de *op. cit.* [Abeillé91], p. 160.
³ Cf. *op. cit.* [Abeillé93], p. 161.

3.2.2.4.6.4.2. Principes et théorèmes

On distingue deux types principaux de principes de bonne formation des signes syntagmatiques : les schémas de dominance immédiate et les principes de partage des traits.¹

* schémas de dominance immédiate : il y en a cinq, ils correspondent à différents types de syntagmes : (influence de la théorie X-barre)

- syntagme saturé avec complément(s)

sans le principe de Sous-catégorisation avec le principe de Sous-catégorisation

Synsem=Cat=[Sous-cat=<>]

Synsem=Cat=[Sous-cat=<>]

Branches $\left[\begin{array}{l} \text{B - Tete} = [] \text{ syntagme} \\ \text{B - Comp} = \langle X \rangle \end{array} \right]$

Branches $\left[\begin{array}{l} \text{B - Tete} = \text{Synsem} = \text{Cat} = [\text{Sous - Cat} = \langle X \rangle] \\ \text{B - Comp} = \langle [\text{Synsem} = X] \rangle \end{array} \right]$

- syntagme non saturé avec complément(s)

$\left[\begin{array}{l} \text{Synsem} = \text{Cat} = [\text{Sous - cat} = \langle X \rangle] \\ \text{Branches} \left[\begin{array}{l} \text{B - Tete} = \text{Synsem} = \text{Cat} = [\text{Sous - cat} = \langle X, Y_1 \dots Y_n \rangle] \\ \text{B - Comp} = \langle [\text{Synsem} = Y_1], \dots, [\text{Synsem} = Y_n] \rangle \end{array} \right] \end{array} \right]$

- syntagme avec ajout(s)

$\left[\begin{array}{l} \text{Branches} \left[\begin{array}{l} \text{B - Tete} = [\text{Synsem} = [1]] \\ \text{B - Ajout} = [\text{Synsem} = \text{Cat} = \text{Tete} = [\text{Modif} = [1]]] \end{array} \right] \end{array} \right]$

- syntagme avec marqueur :

$\left[\begin{array}{l} \text{Synsem} = \text{Cat} = [\text{Marque} = [1]] \\ \text{Branches} \left[\begin{array}{l} \text{B - Tete} = \text{Synsem} = [2] \\ \text{B - Marq} = 1[\text{Synsem} = \text{Cat} = [\text{Spec} = [2]]] \end{array} \right] \end{array} \right]$

- syntagme avec élément antéposé

$\left[\begin{array}{l} \text{Synsem} = \text{Local} = \text{Cat} = \text{Tete} = [\text{Part} = V] \\ \text{Branches} \left[\begin{array}{l} \text{B - Tete} = \text{Synsem} = \text{Non - local} = \left[\begin{array}{l} \text{Hé rit} = [\text{Slash} = \{[1]\}] \\ \text{A - lier} = [\text{Slash} = \{[1]\}] \end{array} \right] \\ \text{B - Anté p} = \text{SynsemLocal} = [1] \end{array} \right] \end{array} \right]$

Ce sont des schémas et non des règles ; ils ne codent pas l'ordre des constituants.

* principes de partage des traits : ce sont différents principes qui règlent la propagation des traits à l'intérieur d'un syntagme :²

- principe des traits de Tête : la valeur de l'attribut Tête inclus dans Cat doit être partagée avec celle de l'attribut Tête du signe de la branche B-Tête du syntagme

¹ Ibid., p. 162 sqq.

² Ibid., pp. 163-164.

Synsem = Cat = [Tête = [1]]
 Branches = B-Tête = Synsem = Cat = [Tête = [1]]

- principe de Sous-catégorisation : la liste de l'attribut Sous-cat est mise à jour au fur et à mesure que le syntagme est saturé (dès qu'un complément est réalisé, dans la branche B-Comp, il est éliminé de la liste de Sous-cat du syntagme correspondant). Un syntagme est saturé quand Sous-cat = \diamond .

La valeur de la liste Sous-cat de la branche Tête d'un syntagme doit correspondre à la concaténation de la liste Sous-cat du syntagme (L1) et de la liste L2 des signes appartenant à la branche Compléments (ou plus précisément la liste des traits Synsem de ces signes).

Synsem = Cat = [Sous - cat = L1]
 Branches $\left[\begin{array}{l} \text{B - Tete} = \text{Synsem} = \text{Cat} = [\text{Sous - cat} = \text{L1} \oplus \text{L2}] \\ \text{B - Comp} = \text{L2} \end{array} \right] \oplus = \text{concatenation de listes}$

- principe sémantique : le but est de régler la propagation des traits sémantiques (Contenu et Contexte). La valeur du trait Contenu de la catégorie dominante est identique à celle du trait Contenu de la catégorie qui est la tête sémantique (branche Ajout ou à défaut branche Tête).¹ Ce principe règle la propagation des traits d'accord entre catégorie Tête et catégorie dominante (du N au SN), puisqu'ils font partie du trait Index, lui même inclus dans Contenu. « Pollard et Sag (1994), développent une théorie syntaxico-sémantique des phénomènes d'accord (entendus au sens large, incluant les phénomènes de contrôle). Les traits morpho-syntaxiques, qui appartiennent à l'attribut Catégorie, ne sont pas forcément partagés en cas d'accord (e. g. pas de partage du trait Cas (ou du trait Part) entre le sujet implicite d'une infinitive et le syntagme contrôleur). Leur théorie s'appuie donc sur certains traits syntaxico-sémantiques (tels que Genre et Nombre), regroupés dans l'attribut Index appartenant à Contenu, et sur les traits sémantico-pragmatiques de l'attribut Contexte » (\Rightarrow possibilité d'analyser des phrases du type *vous êtes (pl.) belle (sing.), Jean et sa manie de bavarder m'a toujours déplu...*).

Exemple pour la phrase *Jean dort* :

$$\left[\begin{array}{l} \text{Contenu} = \left[\begin{array}{l} \text{Relation} = \text{'dormir'} \\ \text{agent} = [1] \end{array} \right] \\ \text{Branches} \left[\begin{array}{l} \text{B - Tete} = \text{Contenu} = \left[\begin{array}{l} \text{Relation} = \text{'dormir'} \\ \text{agent} = [1] \end{array} \right] \\ \text{B - Comp} = \text{Contenu} = \left[\begin{array}{l} \text{Index} = i \\ \text{Restrictions} \left\{ \left[\begin{array}{l} \text{Relation} = \text{'nomin'} \\ \text{Nom} = \text{'jean'} \\ \text{Arg} = i \end{array} \right] \right\} \end{array} \right] \end{array} \right] \end{array} \right]$$

¹ Cf. *op. cit.*[Abeillé93], pp. 169-171.

- principe du trait Spec : le trait Spec permet aux marqueurs de contrôler la valeur de la branche Tête du syntagme, il joue le même rôle que le trait Modif pour les ajouts.¹

Synsem = Cat = [Marque = [1]]
Branches [B - Tete = Synsem = [2] B - Marq = 1 [Synsem = Cat = [Spec = [2]]]]

- principe du trait Marque : Ce principe permet de « remonter » au niveau du syntagme la valeur du complémenteur, qui n'est pas la tête du syntagme.

Synsem = Cat = [Marque = [1]]
Branches [B - Tete = Synsem = [2] B - Marq = 1 [Synsem = Cat = [Spec = [2]]]]

Dans un syntagme, le trait Marque (inclus dans Catégorie mais pas dans Tête) a pour valeur le signe de la branche Marqueur, s'il y en a une, sinon il a la valeur non marquée.

- principe d'ordre des mots : ce sont des principes généraux qui règlent l'ordre linéaire d'apparition des constituants propre à chaque langue.²
- principe des traits non locaux : il règle la propagation non-locale, c'est-à-dire la propagation entre compléments, ajouts ou spécifieurs, et catégories mères, et non entre tête et mère, du trait Slash, mais aussi des traits interrogatifs et relatifs.

La valeur du trait Hérit d'un syntagme est l'union des traits Hérit des catégories Branches moins la valeur du trait A- lier de la branche Tête.

3.2.2.4.6.4.3. Tableau comparatif notation traditionnelle / notation HPSG

Catégories traditionnelles	Notation HPSG
Phrase	Part=V, Sous-cat=<>
Nom propre, pronom, SN	Part=N, Sous-cat=<>
Nom commun sans complément	Part=N, Sous-cat=<Det>
Nom prédicatif	Part=N, Sous-cat=<Det, SP, ...>
Adjectif sans complément, SA	Part=A, Sous-Cat=<>
Adjectif avec complément	Part=A, Sous-cat=<SP>
Verbe intransitif, SV	Part=V, Sous-cat=<SN>
Verbe transitif	Part=V, Sous-cat=<SN, SN...>
Préposition	Part=Prep, Sous-cat=<SN>
Conjonction de subordination	Part=Prep, Sous-cat=<SV/P>
Adverbe, Adv, SP	Part=Prep, Sous-cat=<>

TABLEAU 19 : NOTATION TRADITIONNELLE VS NOTATION HPSG³

¹ Cf. *op. cit.* [Abeillé93], pp. 172-174.

² *Ibid.*, p. 182.

³ Ce tableau est extrait de *op. cit.* [Abeillé91], p. 156.

3.2.2.4.6.5. *Le traitement des prépositions dans HPSG*

En français, il faut pouvoir faire la différence entre les SP sous-catégorisés par un verbe comme *tenir* qui admet des SP(à) nominaux ou infinitifs et un verbe comme *téléphoner* qui interdit les SP(à) infinitifs. Comme pour GPSG, on considère que dans ce cas, la préposition n'est pas la tête de ce type de complément, on introduit ces prépositions dans la branche Marqueur.

Exemple pour à (emplois non prédicatifs) :

Cat=[Tête=[Prepform=à], Spec=[SN/SV]]

Sous-cat=<SN,SN[Marque=à]> (-> *tenir à quelque chose, téléphoner à quelqu'un*)

Sous-cat=<SN[1],SV[Mode=inf,Marque=à,Sous-cat=<SN[1]>]> (-> *tenir à faire quelque chose*).

Les cadres de sous-catégorisation sont notés dans les entrées lexicales par un attribut spécifique (Sous-cat) à valeur de liste, ils sont exprimés en termes catégoriels (comme en GPSG).

dormir => Sous-cat=<SN>

aimer => Sous-cat=<SN,SN>

ressembler à SN => Sous-cat=<SN,SP[Prepform=à]>

provenir de SN => Sous-cat=<SN,SP[Prepform=de]>

donner SN à SN => Sous-cat=<SN,SN,SP[Prepform=à]>.

3.2.2.4.6.6. *Conclusion*

HPSG est donc une approche lexicaliste basée sur les contraintes qui vise à modéliser les langages humains comme des systèmes de contraintes sur les structures de traits typés. Les idées principales des travaux actuels sont les suivantes :¹

- * lexicalisme strict : la structure des mots et la structure des phrases sont gouvernées par des principes indépendants,
- * structures concrètes surfacistes : on évite, si possible, d'utiliser des structures abstraites (par ex. catégories vides et projections fonctionnelles) au profit de structures de constituants relativement conservatrices,
- * prédiction géométrique : l'information linguistique est hiérarchiquement organisée de façon à prévoir l'impossibilité de certains types de phénomènes linguistiques,
- * localité de la sélection de tête (une sorte de prédiction géométrique) : selon la théorie HPSG actuelle de la valence, les têtes lexicales sélectionnent seulement les objets synsem (une sorte de complexe syntactico-sémantique) de leurs compléments, sujets, ou spécifieurs. Il s'en suit que la sélection de catégorie, la répartition des rôles, la répartition des cas, l'accord de la tête et la sélection sémantique obéissent à un type particulier de localité déterminée par les traits de sélection de la valence,

¹ Traduction d'un document internet intitulé « Head-Driven Phrase Structure Grammar (HPSG) : Some Leading Ideas ».

- * encodage local des dépendances non bornées (un autre exemple de prédiction géométrique) : Cette théorie basée sur les traits prévoit l'existence de phénomènes grammaticaux sensibles à de telles spécifications, c'est-à-dire les phénomènes qui apparaissent seulement dans le domaine des constructions des dépendances non bornées,
- * classification lexicale croisée : dans HPSG, les mots contiennent beaucoup d'information. L'information lexicale n'est pas simplement listée, en fait; elle est plutôt organisée en termes de hiérarchies à héritage multiple et de règles lexicales qui permettent aux mots complexes d'être dérivés à partir de la logique du dictionnaire. Les recherches actuelles développent des extensions de dictionnaires hiérarchiques qui permettent aux règles lexicales d'être éliminées et aux modèles de liage d'être dérivés d'une façon générale à partir des propriétés sémantiques,
- * classification hiérarchique croisée des constructions grammaticales : les types de phrases sont traités en termes de hiérarchies à héritage multiple,
- * théorie du liage basée sur l'oblicité : les généralisations sur les contraintes de liage d'éléments référentiellement dépendants sont établies en termes d'oblicité relative (o-command), plutôt qu'en termes de supériorité configurationnelle (c-command),
- * théorie de contrôle basée sur la sémantique : l'identification des contrôleurs dans les constructions de contrôle est sémantique de nature,
- * théorie de linéarisation : les travaux actuels en HPSG explorent les modes de sérialisation qui ne sont pas basés sur le modèle de la grammaire de structure de phrase traditionnelle (où les phrases sont des chaînes de mots définies de façon dérivationnelle en termes de structure de phrase). Ceci a des conséquences sur le traitement des constituants discontinus, autorisant même l'introduction de niveaux d'organisation syntaxique linéaire qui sont, dans une certaine mesure, dissociés des relations combinatoires entre les items mis en série.

Les intérêts principaux de HPSG sont : ¹

- * la systématisation de l'utilisation d'une seule structure de données : les structures de traits, pour la représentation du lexique, de la syntaxe, de la sémantique et des phrases de la langue. Elle permet l'enrichissement de la notion d'arbre syntagmatique par un typage différencié des différentes branches, la mise au premier plan de la description conjointe des phénomènes syntaxiques et sémantiques (et de leurs interactions),
- * le contrôle sur l'organisation des structures de traits (organisation hiérarchique des traits et de leur typage) pour éviter la prolifération et l'éclatement des catégories,
- * la systématisation de l'utilisation de règles lexicales,
- * l'utilisation de règles enchâssées qui permet un accès à la plupart des informations nécessaires (cependant, incorporation d'informations non locales

¹ Cf. *op. cit.* [Abeillé91], pp. 195-196.

par l'introduction de traits supplémentaires et de principes de propagation qui leur sont propres : CL pour les pronoms clitiques...).

La quatrième conférence internationale sur HPSG s'est tenue du 18 au 20 juillet 1997 au *LSA Linguistic Institute* à l'université Cornell.

3.2.2.4.7. Grammaire d'arbres adjoints (TAG, Tree Adjoining Grammar)

3.2.2.4.7.1. Naissance, motivations de départ, objectifs

Ce formalisme est élaboré au milieu des années 70 par A. Joshi et al. C'est une extension des grammaires en chaîne. Leurs motivations de départ sont de définir un système à la capacité générative plus contrainte que celle des grammaires transformationnelles, d'effectuer un travail mathématique visant à formaliser la composante récursive de la grammaire (par l'opération d'adjonction).

Les recherches formelles en informatique et en mathématiques sont très actives en TAG. Le développement s'effectue en plusieurs étapes : travaux formels (A. Joshi 1985, K. Vijay-Shanker & A. Joshi 1988, D. Weir 1988, Y. Schabes et al. 1988), syntaxiques (A. Kroch & A. Joshi 1985, A. Abeillé 1988, 1991), sémantiques (S. Schieber & Y. Schabes 1990).

TAG utilise des arbres élémentaires (et non des règles de réécriture) comme unités de base associées aux items lexicaux, combinées par deux opérations : la substitution et l'adjonction.

C'est une grammaire lexicalisée c'est-à-dire que chacun des arbres élémentaires a un nœud-feuille occupé par une catégorie terminale. Le codage lexical est très important, il est établi à partir de traits à valeur atomique uniquement.

3.2.2.4.7.2. Notations utilisées dans TAG

a : arbre initial	b : arbre auxiliaire
γ : arbre dérivé	\downarrow : nœud à substitution
* : nœud-pied	\Rightarrow : produit d'une règle lexicale
t : trait amont	b : trait aval
\Rightarrow : résultat d'une combinaison d'arbres (arbre dérivé)	
[] : ensemble des formes fléchies pour un même lemme	

3.2.2.4.7.3. Le formalisme TAG

3.2.2.4.7.3.1. Description du modèle

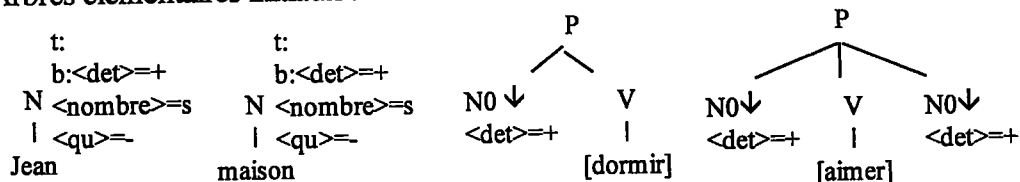
Une grammaire syntagmatique est une grammaire basée sur des règles de réécriture. Une grammaire arborescente est une grammaire basée sur des arbres élémentaires initiaux (notés a) combinés par substitution, et auxiliaires (notés b) combinés par adjonction. Ils ont un nœud-feuille (appelé nœud-pied et noté *) portant un non terminal de même catégorie que le nœud-racine, pour les autres nœuds (terminaux ou non), il y aura substitution (notée \downarrow). Dans une grammaire arborescente, on distingue deux types

d'arbres : un arbre dérivé c'est-à-dire une structure syntagmatique associée à une phrase du langage, il résulte de la combinaison de plusieurs arbres, et un arbre de dérivation qui indique par quelles combinaisons d'arbres élémentaires cette structure a été obtenue, c'est donc une représentation arborescente d'une combinaison d'arbres élémentaires : les nœuds sont étiquetés par un couple (arbre élémentaire, adresse) et les branches ne sont pas ordonnées. Un arbre élémentaire est un arbre de profondeur finie, unité de base d'une grammaire TAG. Un quasi-arbre est un cas particulier d'arbre non standard où l'on peut ne spécifier que partiellement les relations de dominance.

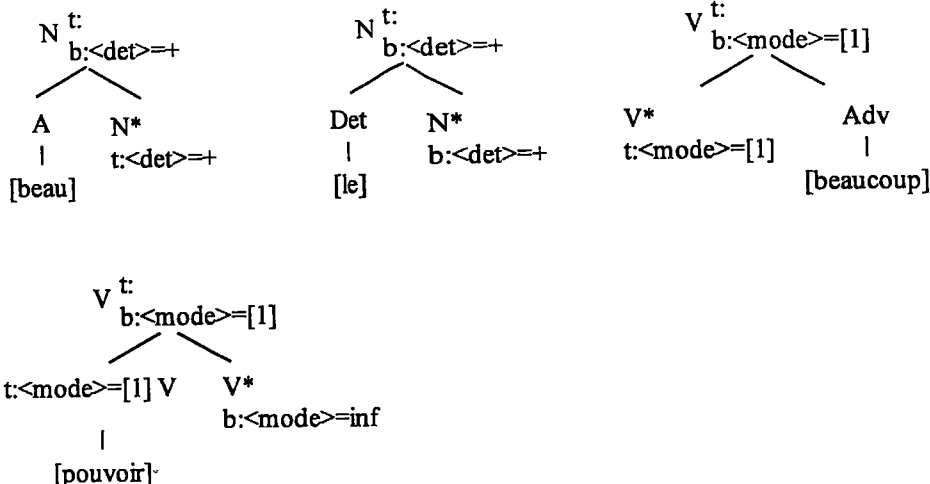
Une grammaire d'arbres adjoints est un ensemble fini d'arbres élémentaires combinés par les opérations d'adjonction ou de substitution. Ce type de grammaire a permis l'implémentation d'un prototype d'analyse de large couverture syntaxique et lexicale pour le français (Abeillé, 1991).

Un arbre élémentaire a au moins un de ses nœuds-feuilles occupé par un item lexical qui lui sert de tête, et peut être de deux types : initial (noté a) ou auxiliaire (noté b), selon qu'il se combine avec d'autres arbres par substitution ou par adjonction. Les arbres auxiliaires ont un nœud-feuille (appelé pied) portant un non terminal de même catégorie que le nœud-racine. Les nœuds-feuilles portent des terminaux ou une marque pour la substitution. Les arbres initiaux sont utilisés pour les noms, les verbes à arguments nominaux ou prépositionnels, et les arbres auxiliaires pour les modifieurs (relatives, adjectifs, adverbes) et plus généralement les ajouts au sens harrissien du terme (verbes auxiliaires ou opérateurs, verbes à complétives, déterminants...), ils représentent les structures linguistiques minimales qui ne contiennent aucune récursivité. On peut les caractériser de la façon suivante : 1) tous les nœuds internes sont étiquetés par des non terminaux, 2) tous les nœuds-feuilles sont étiquetés par des terminaux ou par des nœuds non terminaux marqués pour la substitution. Les arbres élémentaires doivent correspondre à des structures argumentales complètes. On peut les caractériser ainsi : 1) tous les nœuds internes sont étiquetés par des non terminaux, 2) tous les nœuds-feuilles sont étiquetés par des terminaux ou par des nœuds non terminaux marqués pour la substitution, à l'exception d'un seul nœud non terminal, appelé le nœud-pied, qui ne peut être utilisé que pour adjoindre à un arbre à un autre nœud, le nœud-pied a la même étiquette que le nœud-racine de l'arbre.

Arbres élémentaires initiaux :



Arbres élémentaires auxiliaires :



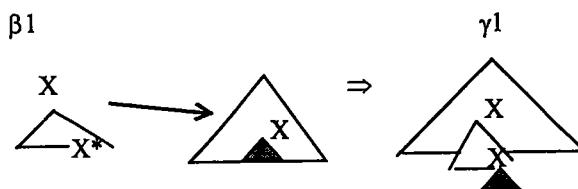
Le signe \downarrow indique les nœuds à substitution, le signe * les nœuds pieds, les crochets [] une des formes fléchies (ou élidées) du lemme correspondant.

Les traits amont (notés t: pour *top*) indiquent les relations du nœud avec les nœuds qui le dominent et les traits aval (notés b: pour *bottom*) indiquent les relations du nœud avec les nœuds qu'il domine.

Le langage TAG est produit par la combinaison des arbres élémentaires : ¹

- * l'adjonction est une opération qui insère un arbre auxiliaire dans un arbre élémentaire ou dérivé : le nœud X où a lieu l'adjonction est remplacé par l'arbre auxiliaire (opération facultative et réitérable). Le nœud-racine et le nœud-pied de l'arbre auxiliaire doivent s'apparier avec le nœud auquel s'adjoint l'arbre auxiliaire. Le nœud adjoint se divise et son trait amont s'unifie avec le trait amont du nœud adjoint de la racine, alors que son trait aval s'unifie avec le trait aval du nœud adjoint du pied.

Adjonction (d'un arbre auxiliaire) :

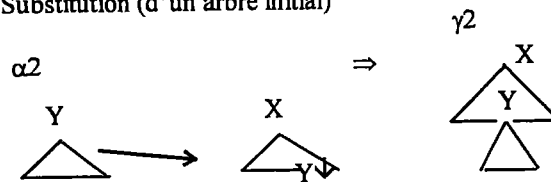


- * la substitution, similaire à l'opération utilisée dans les grammaires de réécriture, et moins puissante que l'adjonction, insère un arbre initial (ou dérivé d'un arbre initial) de racine X à un nœud-feuille de même catégorie dans un arbre élémentaire ou dérivé (opération obligatoire et non réitérable à un nœud donné). Le nœud-racine et le nœud de substitution doivent avoir le même nom. La structure de traits d'un nouveau nœud créé par substitution hérite de l'union des traits des nœud originaux. Le trait amont du nouveau nœud correspond à l'union des traits amont des deux nœuds originaux, alors que le trait aval du

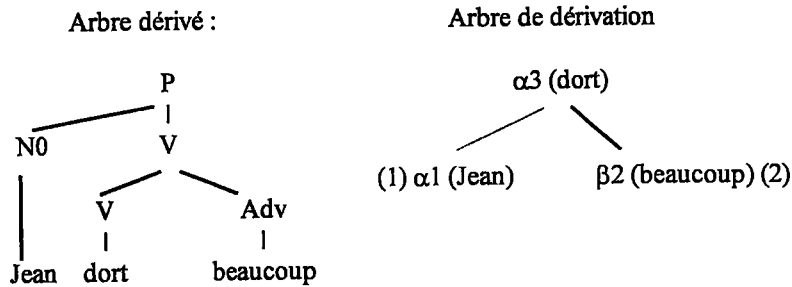
¹ Cf. *op. cit.* [Abeillé91], pp. 148-149.

nouveau nœud représente simplement le trait aval du nœud amont de l'arbre de substitution (puisque le nœud de substitution n'a pas de trait aval).

Substitution (d'un arbre initial)



La combinaison d'arbres élémentaires donne un « arbre dérivé » associé à un « arbre de dérivation. »



Dans l'arbre de dérivation, on utilise un trait en pointillé pour la substitution, un trait plein pour l'adjonction. Un arbre dérivé n'indique pas forcément de façon univoque à partir de quels arbres élémentaires il a été obtenu (contrairement aux grammaires basées sur des règles de réécriture).

En cas de substitution, les traits amont du nœud de l'arbre substitué doivent s'unifier avec les traits amont du nœud où a lieu la substitution. En cas d'adjonction, il doit y avoir unification des traits amont du nœud racine de l'arbre auxiliaire avec les traits amont du nœud qui reçoit l'adjonction, et des traits aval du nœud pied de l'arbre auxiliaire avec les traits aval du nœud recevant l'adjonction. A la fin d'une dérivation, traits amont et traits aval doivent s'unifier à chaque nœud dans l'arbre dérivé.¹

Chaque arbre élémentaire correspondant à une unité lexicale, l'arbre de dérivation représente la structure de la phrase sous forme de dépendances lexicales : les informations strictement syntaxiques (symboles auxiliaires, niveaux syntagmatiques intermédiaires, ordre des mots) n'y figurent pas en tant que telles.

Cette double représentation permet de rendre compte de nombreux cas de discordance entre constituance (découpage des syntagmes) et relations de dépendance (plus sémantiques).

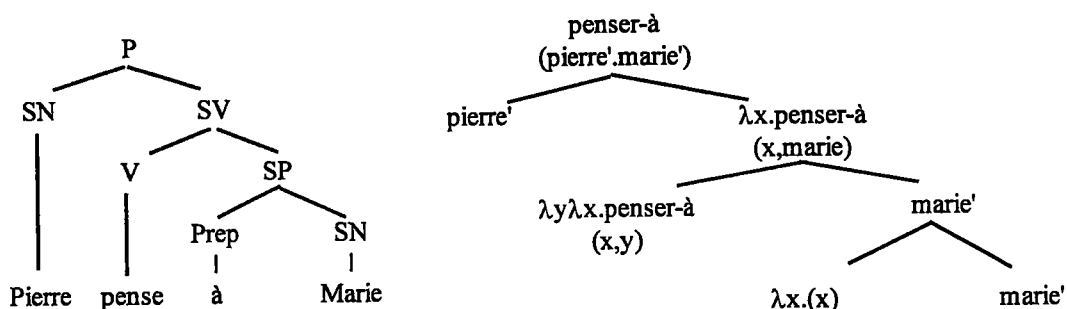
En TAG, les dépendances sont représentées de la façon suivante :²

- * dépendances locales : à l'intérieur des arbres élémentaires, par leur topologie ou par les traits syntaxiques associés à leurs nœuds. Les arbres des prédicats définissent les nœuds associés à leurs arguments, et les arbres des modificateurs ceux des catégories qu'ils modifient.

¹ Cf. *op. cit.* [Abeillé91], p. 160.

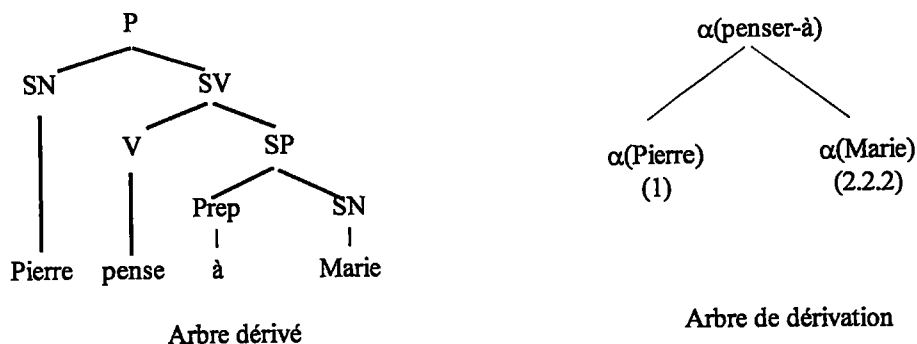
² *Ibid.*, p. 150.

- * dépendances moyennes : à l'intérieur des structures élémentaires de la grammaire, entendues cette fois comme « quasi »-arbres, ou descriptions arborescentes (Vijay-Shanker, 1992), les différents sous-arbres d'une même structure devant se combiner à différents nœuds d'un seul et même arbre élémentaire.
- * dépendances à longue distance : à l'intérieur des arbres élémentaires, mais de profondeur étendue (supérieure à 2), avec un (ou plusieurs) nœud(s) à l'intérieur permettant d'adjoindre à volonté des constituants intermédiaires.



En TAG, une représentation sémantique doit être associée à chaque nœud de l'arbre de dérivation, c'est-à-dire à chaque arbre élémentaire, on peut ignorer certains nœuds syntaxiques dans la représentation sémantique.

Sémantique compositionnelle en TAG

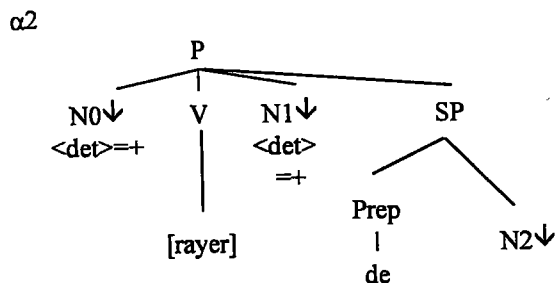
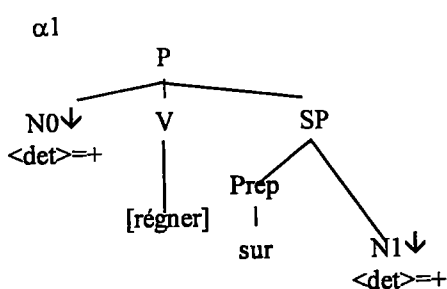


Les considérations sémantiques qui interviennent dans la définition des unités syntaxiques sont les suivantes :¹

- * l'ancrage sémantique : tout arbre syntaxique élémentaire a un correspondant sémantique non vide, ceci exclut la plupart des éléments fonctionnels (prépositions « vides », complémenteurs, certains pronoms relatifs) en tant qu'entité autonome de la syntaxe. « Les éléments fonctionnels sans autonomie sémantique sont co-têtes de l'arbre du verbe, du nom ou de l'adjectif qui les gouverne ».

¹ Cf. *op. cit.* [Abeillé91], pp. 151-152.

Verbes avec co-têtes fonctionnelles



- * la non-compositionnalité : un arbre élémentaire correspond à une seule unité sémantique, une limite est imposée au nombre d'items lexicaux (têtes) ainsi réalisés dans un arbre élémentaire donné ; ces éléments doivent former une seule unité sémantique.

On peut définir la notion de tête comme suit : ¹

- * il s'agit forcément d'une tête lexicale,
- * il ne s'agit pas forcément d'un seul élément lexical, puisqu'on peut avoir des co-têtes ou têtes multiples,
- * une tête vaut pour un arbre élémentaire quelle que soit sa profondeur et n'est pas limitée à un seul niveau syntagmatique,
- * il n'y a pas forcément de partage des traits morpho-syntaxiques entre la tête lexicale et la racine de l'arbre élémentaire.

Il existe trois types de traits morpho-syntaxiques (comme en GPSG) : ²

- * les traits « d'accord » doivent s'unifier en parties amont des nœuds concernés,
- * les traits « de tête » de la partie amont du premier nœud qui domine la tête lexicale de l'arbre élémentaire (prédicatif) doivent s'unifier avec les traits « de tête » de la partie aval du nœud dominant immédiatement (et ainsi de suite pour chaque couple de nœuds dans une relation de dominance immédiate jusqu'au nœud-racine). Dans le cas d'arbre auxiliaire correspondant à un modifieur, l'unification doit se faire entre traits de tête (amont) du nœud-pied et traits de tête (aval) du nœud-racine,
- * les traits « de pied » en partie aval de la catégorie dominante doivent s'unifier avec les traits « de pied » en partie amont de la (ou des) catégorie(s) dominée(s) pertinente(s).

Au niveau des traits sémantiques, on peut formaliser certaines opérations, notamment la distinction entre traits inhérents et traits afférents :

- * les traits inhérents sont associés aux nœuds lexicaux en partie aval,
 - le trait est activé s'il est recopié au nœud syntagmatique supérieur, c'est-à-dire s'il est prévu un partage de valeur entre nœud-pied et nœud-racine dans l'arbre auxiliaire qui s'est adjoint,

¹ *Ibid.*, p. 214.

² Cf. *op. cit.* [Abeillé91], p. 161 sqq.

- le trait est inhibé (ou virtualisé) s'il n'est pas transmis au nœud syntagmatique supérieur (ce qui se fait là aussi dans la définition de l'arbre auxiliaire),

* les traits afférents sont associés aux nœuds-racines des modificateurs.

En ce qui concerne la sous-catégorisation, on ne fait pas appel à des traits spécifiques tels que Pred (comme dans LFG) ou Sous-cat (comme dans HPSG et GPSG), elle est représentée par la structure même des arbres élémentaires associés au prédicat considéré. Le sujet est sous-catégorisé comme les compléments (comme dans HPSG et LFG).

Au modèle de base ont été ajoutées différentes extensions :

- * le format DL/OL (Dominance linéaire/Ordre linéaire) pour les variations d'ordre des mots (A. Joshi, 1987)
- * les TAG « ensemblistes » pour certaines dépendances à distance (D. Weir, 1988)
- * la notion de descriptions arborescentes pour réinterpréter les structures élémentaires comme des « quasi-arbres » (K. Vijay-Shanker, 1992)
- * les TAG « Synchrones » pour associer une interprétation sémantique aux dérivations (S. Schieber & Y. Schabes, 1990).

Le format DL/OL est inspiré du format ID/LP du modèle GPSG. Les arbres élémentaires ne représentent que des relations de dominance, les relations d'ordre sont notées séparément.

Ex. *N0 parler de N1 à N2* (un seul arbre DL auquel on associe les relations d'ordre suivantes :

$N0 < V$ $V < SP1$ $V < SP2$
 $Prep1 < N1$ $Prep2 < N2$

si les SP ne sont pas ordonnées, on aura deux réalisations : *N0 parler de N1 à N2*, *N0 parler à N2 de N1*

Les avantages du format DL/OL par rapport à ID/LP sont les suivants :

- * il est possible d'exprimer des relations entre éléments appartenant à des niveaux syntagmatiques différents,
- * il est possible d'aboutir à des structures arborescentes non standard (avec des branches qui se croisent) utiles pour des langues à ordre des mots plus libre que le français.

Mais il y a des contraintes :

- * on ne peut exprimer directement de relations entre des nœuds appartenant à des arbres élémentaires différents,
- * les relations d'ordre exprimées entre deux nœuds sont forcément conservées au niveau de leurs descendants.

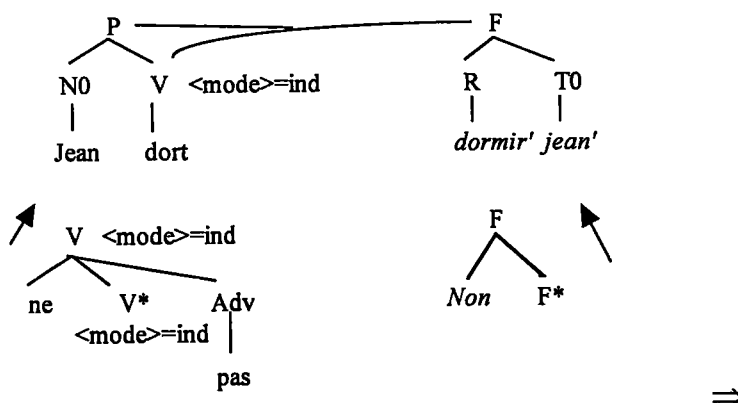
Dans les TAG « ensemblistes »¹ on définit comme structures élémentaires des ensembles (finis) d'arbres élémentaires (D. Weir, 1988). Les arbres d'un même ensemble doivent être utilisés en même temps, lors de la combinaison avec une autre structure ; ils peuvent être de types différents (arbre initial et arbre auxiliaire par exemple). « Il permet de regrouper dans une même structure élémentaire, un arbre initial de racine N pour un nom prédicatif (par ex. *exemple*) et un arbre auxiliaire de racine P pour son complément prépositionnel (par ex. *de N1*) ». Si on combine les arbres d'un même ensemble avec un même arbre élémentaire, les dérivations sont toujours représentables, et il y a équivalence (faible) avec le modèle TAG standard, si on combine les arbres d'un même ensemble avec des arbres différents, on obtient un modèle plus puissant.

Pour coder les structures élémentaires d'une grammaire TAG, Vijay-Shanker propose d'utiliser un langage de description spécifique qui décompose les relations de dominance, les relations de précedence et l'étiquetage des nœuds en autant d'équations. Les structures élémentaires sont réinterprétées comme des « quasi-arbres »², c'est-à-dire des structures arborescentes où les relations de dominance ne sont que partiellement spécifiées. On peut noter deux avantages : le regroupement dans un même modèle des TAG « classiques » et des TAG « ensemblistes » et le rapprochement des modèles précédents davantage basés sur l'unification.

S. Shieber & Y. Schabes 1990 ont défini le formalisme des TAG « synchrones » qui permet le calcul d'une (ou de plusieurs) représentation(s) sémantique(s) associée(s) à une dérivation syntaxique. (...) Les dérivations syntaxiques et sémantiques doivent être « synchrones », c'est-à-dire se développer comme suit à partir d'une paire d'arbres élémentaires E1 et E2, l'un syntaxique, l'autre sémantique :

- choisir un nœud A de E1 et un nœud B de E2 reliés par un lien,
- choisir une autre paire d'arbres E3 et E4 de racines respectives A et B,
- combiner E3 en A et E4 en B et supprimer le lien,
- faire de même avec d'autres nœuds et d'autres paires d'arbres, jusqu'à dérivation complète de la phrase considérée

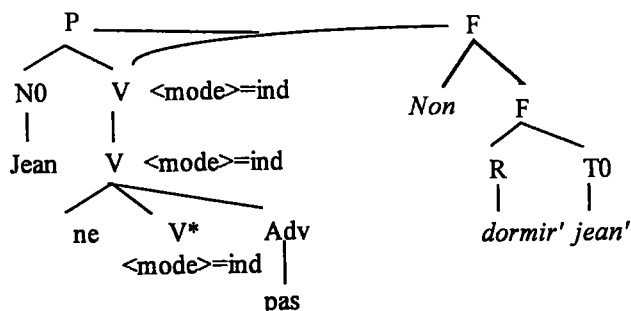
Dérivation sémantique synchrone de *Jean ne dort pas* :³



¹ Cf. *op. cit.* [Abeillé91], p. 238.

² *Ibid.*, p. 239.

³ Les graphiques sont extraits de *op. cit.* [Abeillé91], p. 241.



3.2.2.4.7.3.2. Principes et théorèmes

Les principes de bonne formation pour les arbres auxiliaires sont les suivants : ¹

- * principe d'ancrage lexical (ou lexicalisation) : tout arbre élémentaire a au moins une tête lexicale non vide. La chaîne vide (sauf cas d'ellipse) ne peut être la seule tête d'un arbre élémentaire.
- * principe de cooccurrence prédicat-arguments (ou principe de sous-catégorisation) : tout prédicat contient dans sa structure élémentaire au moins un nœud pour chacun des arguments qu'il sous-catégorise (=> sous-catégorisation comme dans GPSG, HPSG).
- * principe de consistance sémantique : tout arbre syntaxique élémentaire a un correspondant sémantique non vide ; la plupart des éléments fonctionnels (prépositions « vides », complémenteurs, certains pronoms relatifs) sont exclus en tant qu'entité autonome de la syntaxe.
- * principe de non-compositionnalité : un arbre élémentaire correspond à une seule unité sémantique ; il n'y a pas de limite au nombre d'items lexicaux (têtes) réalisés dans un arbre élémentaire donné, mais contrainte : ces éléments doivent former une seule unité sémantique.

3.2.2.4.7.3.3. Tableau comparatif notation traditionnelle / notation TAG

Catégories traditionnelles	Représentation en TAG
Nom, pronom, SN	Arbre initial de racine N
Déterminant	Arbre auxiliaire de racine N
Adjectif, SA (attribut)	Arbre initial de racine A
Adjectif, SA (épithète)	Arbre auxiliaire de racine N
Verbe intransitif, verbe à complément nominal ou prépositionnel	Arbre initial de racine P
Verbe à complétive	Arbre auxiliaire de racine P
Verbe auxiliaire ou modal	Arbre auxiliaire de racine V
Préposition (modifieur de nom ou circonstant)	Arbre auxiliaire de racine N, V ou P
Conjonction de subordination	Arbre auxiliaire de racine P ou V
Adverbe, Sadv	Arbre auxiliaire de racine V, P ou A

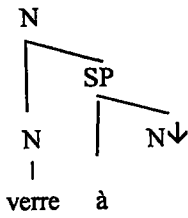
TABLEAU 20 : NOTATION TRADITIONNELLE VS NOTATION TAG ¹

¹ Cf. *op. cit.* [Abeillé91], pp. 212-213.

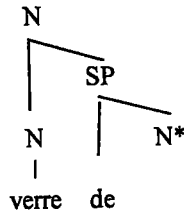
3.2.2.4.7.4. Le traitement des prépositions dans TAG

Constructions en N prép. N :²

(a) verre-à



(b) verre-de



(a) vin



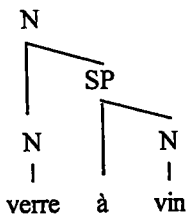
Tête syntaxique : *verre* en (a, arbre initial) et (b, arbre auxiliaire)

Tête sémantique : *verre* en (a), *N* en (b)

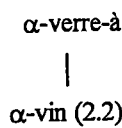
L'arbre associé à *verre de N* s'adjoint au nœud racine de l'arbre de *vin*, tandis que l'arbre de *verre à N* reçoit, lui, la substitution (à son nœud feuille N) de l'arbre de *vin*. La structure syntagmatique associée aux deux séquences *verre à vin* et *verre de vin* est la même.

En revanche, les relations de dépendance sont inversées : « la séquence *verre à vin* est dérivée à partir de l'arbre initial associé à *verre à N*, tandis que la séquence *verre de vin* est dérivée à partir de l'arbre associé à *vin* :

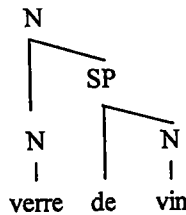
Arbre dérivé



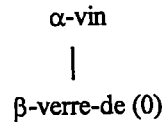
Dérivation



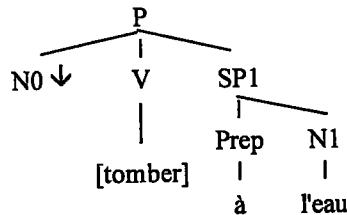
Arbre dérivé



Dérivation



Expressions verbales semi-figées :



Pour les verbes à complément prépositionnel, on n'utilise pas le trait Prepform (ou Pcas) pour sélectionner la bonne préposition. Grâce au principe de consistance sémantique, les prépositions marqueurs d'argument, sémantiquement « vides », sont

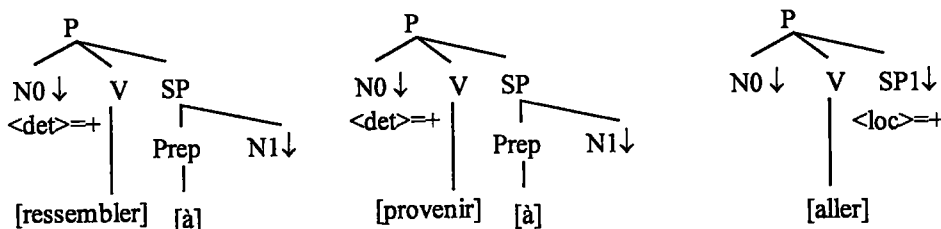
¹ Le tableau est extrait de *op. cit.* [Abeillé91], p. 215.

² Cf. *op. cit.* [Abeillé91], pp. 153-154.

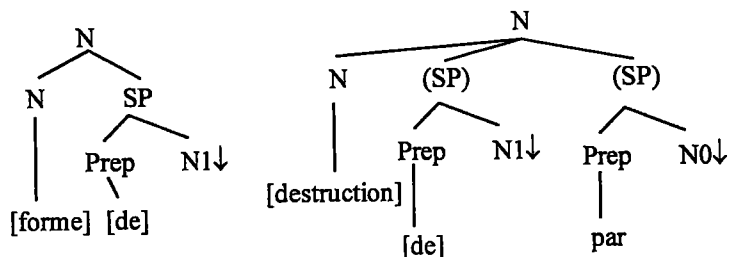
représentées comme co-têtes fonctionnelles, elles font partie de la définition de l'arbre élémentaire du verbe.¹

En ce qui concerne les compléments locatifs, les prépositions ne sont pas vides (*aller à Paris* est différent de *aller vers Paris*); elles sont considérées comme unités autonomes, têtes d'un arbre initial de racine SP, qui a sa propre sémantique et qui est substitué dans l'arbre du verbe à complément locatif.²

Arbres élémentaires avec complément prépositionnel :



Quand un verbe comme *rêver de* peut être suivi d'un GN ou d'une infinitive, il a deux types d'arbres élémentaires : un arbre initial avec un nœud N1 (à substitution) correspondant à sa construction nominale, et un arbre auxiliaire avec un nœud-pied P1 pour sa construction infinitive. Il en est de même pour les compléments de noms, d'adjectifs, d'adverbes et de prépositions. Ils auront dans leurs arbres élémentaires un nœud (ou un sous-arbre) pour chacun de leurs compléments :³



Si les arbres fournissent un point de départ pour une analyse sémantique en terme de dépendances lexicales, ils ne sont pas suffisants et doivent être complétés par des informations plus explicites. On pourra les relier :

- * à une représentation sous forme de graphes conceptuels : graphes orientés bipartites, dont les sommets sont soit des concepts (notés entre crochets), soit des relations (notées par des ovaux) (cf. Sowa, 1984).

Pour passer d'un arbre de dérivation à un graphe conceptuel, il faut l'enrichir en :

- remplaçant les adresses de combinaison des arbres élémentaires par des étiquettes correspondant à des relations sémantiques (agent, objet, lieu...)
- ajoutant des arcs pour les unités remplissant simultanément plusieurs rôles sémantiques

¹ Cf. *op. cit.* [Abeillé91], p. 228.

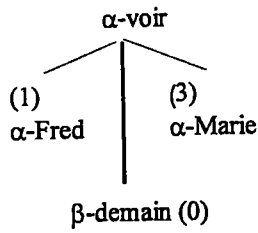
² *Ibid.*

³ *Ibid.*, p. 229.

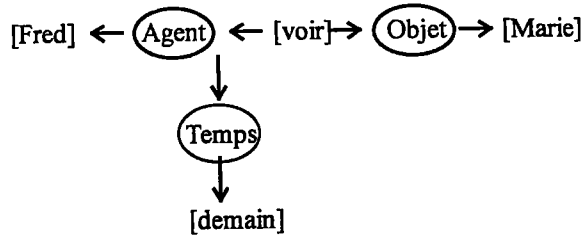
- éventuellement, remplaçant les nœuds correspondant à des unités lexicales sémantiquement complexes en un sous-graphe correspondant.

Ex. : *Fred verra Marie demain*

Arbre de dérivation

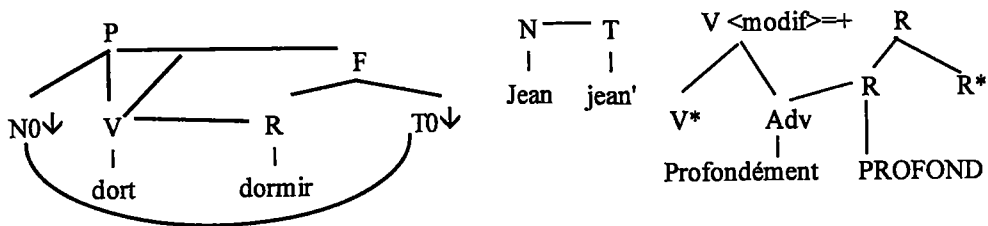


Graphe conceptuel



- * ou à une représentation plus sophistiquée à base d'arbres sémantiques : on associe à chaque arbre élémentaire au moins un arbre sémantique, et on définit des liens entre leurs nœuds qui contraignent les dérivations ultérieures (cf. Schieber & Schabes, 1990). Les arbres syntaxiques et sémantiques reliés ne sont pas forcément du même type (l'un peut être un arbre initial, l'autre un arbre auxiliaire). Le nom propre est couplé à un terme (T), le nœud V de l'arbre phrastique aux nœuds sémantiques R (pour relation) et F (pour formule), ce qui correspond à la portée des modifieurs qui peuvent s'adjoindre à V :

Couples d'arbres élémentaires syntaxiques et sémantiques



3.2.2.5. CONCLUSION GENERALE SUR LES GRAMMAIRES D'UNIFICATION

À la fin de cette présentation sur les grammaires d'unification, on constate qu'elles ont des points communs :

- * elles sont nées d'une critique de la grammaire générative transformationnelle,
- * leur développement se caractérise par une utilisation importante d'outils formels, il a été accéléré en raison de la progression du Traitement Automatique des Langues (TAL); elles enrichissent leurs modèles mathématiques sous-jacents par des principes de bonne formation proprement linguistiques,
- * elles cherchent à articuler plus explicitement descriptions lexicales, syntaxiques et sémantiques, cette articulation est indispensable pour toute application informatique ; la compatibilité entre ces trois types de descriptions se fait grâce à l'unification,

- * l'accent est mis sur les descriptions linguistiques, elles ont recours à un style d'analyse syntaxique plus « concret » ce qui restreint le nombre d'étapes intermédiaires dans la production d'une phrase

Mais il existe également des différences :

- * les motivations de départ différentes : LFG voulait intégrer des travaux en psycholinguistique en abandonnant l'appareil transformationnel ; GPSG voulait construire un modèle équivalent à une grammaire hors contexte ; HPSG visait l'intégration plus explicite de la syntaxe et de la sémantique ; TAG avait pour but de formaliser la composante récursive de la grammaire et avait le souci plus linguistique d'une « lexicalisation » intégrale de la grammaire,
- * les structures de traits ne sont pas utilisées de la même façon dans les différentes grammaires d'unification : dans HPSG, elles remplacent les représentations arborescentes ; dans LFG, les traits complexes sont utilisés pour formaliser la notion de fonction grammaticale ; dans GPSG, les traits sont un moyen de vérifier les contraintes et d'exprimer des dépendances entre des éléments situés à des niveaux syntagmatiques différents ; dans TAG, la notion de description arborescente est une représentation équivalente des structures de traits complexes,
- * en HPSG, le cadre de sous-catégorisation ordonne le sujet en premier et non en dernier comme dans les autres grammaires.
- * malgré quelques points communs (grammaires à traits...), GPSG et UCG présentent des différences : GPSG est une grammaire syntagmatique indépendante du contexte mais UCG est une grammaire catégorielle, GPSG est basée sur la sémantique de Montague alors que UCG est basée sur une extension de la *Discourse Representation Theory* de Kamp, GPSG code la sous-catégorisation dans les règles et UCG la code dans le lexique.

Nous présentons, à présent, les avantages des grammaires d'unification :

- * elles présentent à la fois des aspects théoriques et des propriétés computationnelles,
- * elles développent la notion de structures de traits qui jouent un rôle clé dans la représentation de l'information linguistique, les structures de traits sont très proches des objets habituellement manipulés par les linguistes. La décomposition en traits permet de regrouper les objets linguistiques sur la base de telle ou telle propriété commune et d'obtenir si besoin est des classifications croisées,
- * elles sont souvent utilisées pour définir la relation entre les chaînes de caractères et la représentation du sens, elle peuvent donc être utilisées de façon bidirectionnelle et la même grammaire peut servir à la fois pour le parsing et la génération,
- * elles intègrent le lexique, la syntaxe et la sémantique, réhabilitent les descriptions de « surface », définissent des principes généraux de bonne formation des énoncés,
- * leur conception « intégratrice » est un atout pour le TAL ; elles sont faciles à implémenter, elles se basent sur des modèles logiques ou mathématiques

étudiés par ailleurs et pour lesquels ont été définies des méthodes de programmation,

- * dans l'interprétation des équations, on identifie les valeurs des traits concernés et on les remplace partout par cette valeur commune (on ne dit pas que le nom s'accorde avec le déterminant ni que le déterminant s'accorde avec le nom, il y a simplement « partage » de valeurs),
- * en ce qui concerne les grammaires d'unification les plus « syntagmatiques », elles réinterprètent les règles de réécriture comme des descriptions de sous-arbres permettant de filtrer, parmi les différents sous-arbres possibles, ceux qui sont compatibles avec la grammaire considérée,

Il n'y a pas vraiment d'inconvénients communs à l'ensemble des grammaires d'unification, si ce n'est peut-être l'utilisation en trop grand nombre d'entrées lexicales qui peut entraîner une perte d'efficacité. En revanche, on peut citer les inconvénients que chacune présente :

- * dans des applications de taille réelle utilisant LFG, on risque de perdre en efficacité en raison de la multiplication des entrées lexicales liées à l'utilisation généralisée des règles lexicales,
- * dans GPSG, les auteurs ont tendance à proposer des règles très compactes, où ne figure explicitement que peu d'information, donc les principes de bonne formation sont plus complexes ce qui peut annuler (pour de grosses applications) l'avantage informatique lié à la réduction de la capacité générative de la grammaire. Les réalisations pratiques se sont traduites par un éclatement difficile à maîtriser. D'un point de vue descriptif, on se heurte vite à des blocages,
- * en HPSG, il n'est pas clair que la construction d'un lexique en vraie grandeur pourra éviter l'éclatement et la prolifération des types et sous-types. Comme l'utilisation de règles lexicales est systématisée, l'organisation d'un système complet risque de faire apparaître un certain éclatement.

	LFG	GPSG	HPSG	TAG
<i>Naissance</i>	Bresnan & Kaplan, fin des années 70	Gazdar, Klein, Sag, Pullum début des années 80	Pollard, Sag début des années 80	Joshi et al. milieu des années 70
<i>Catégories</i>	Atomes	Ensemble de traits	Structure de traits typées	Arbres élémentaires avec « tête lexicale »
<i>Traits syntaxiques enchâssés</i>	Structures fonctionnelles	Quelques traits à valeur catégorielle	Structure de traits typées	non utilisés
<i>Principes de combinaison</i>	Règles hors contexte (avec annotations fonctionnelles) et principes généraux	Règles hors contexte (avec catégorie « tête »)	Structures de traits (schémas DI) et principes généraux	Adjonction ou substitution et unification
<i>Résultat d'une analyse</i>	Structure en constituants et structure fonctionnelle	Arbre syntagmatique	Structure de traits typée	Arbre syntagmatique et arbre de dépendance
<i>Représentations sémantiques</i>	Structure argumentale et structure sémantique	Logique intensionnelle (Montague)	Structure de traits typée (sémantique situationnelle)	Arbres sémantiques (synchrone)
<i>Généralisations syntaxiques</i>	Règles « lexicales »	Métarègles	Règles « lexicales »	Règles « lexicales » (famille d'arbres élémentaires)
<i>Equivalent mathématique</i>	Grammaires contextuelles	Grammaires hors contexte	Grammaires non contraintes	Grammaires « légèrement contextuelles »
<i>Principes</i>	d'unicité de cohérence de complétude	des traits de pied des traits de tête d'accord contrôle de restrictions sur la cooccurrence des traits de spécifications de traits par défaut	schémas de dominance immédiate principe de partage des traits	d'ancrage lexical de cooccurrence prédicat-arguments de consistance sémantique de non-compositionnalité

TABLEAU 21 : TABLEAU COMPARATIF DES MODELES DE GRAMMAIRES BASEES SUR L'UNIFICATION

Ce tableau est extrait de *op. cit.* [Abeillé93]. Il ne concerne que les grammaires d'unification de type « théorie ».

4. LE TRAITEMENT AUTOMATIQUE

4.1. GENERALITES

Premièrement, l'analyse linguistique est un préalable au traitement informatique. Elle permet une préanalyse du corpus, en fonction de la tâche. Ce point de vue que nous adoptons [F. Rastier] n'est pas partagé par la communauté de l'informatique linguistique. Deuxièmement, elle guide l'analyse informatique, dans la stratégie d'usage des logiciels. Enfin, elle permet l'interprétation des résultats du traitement. Dans le meilleur des cas, elle intervient alors aussi bien avant, pendant et après la mise en œuvre informatique.¹ La formalisation du contenu est indispensable à toute implémentation informatique.

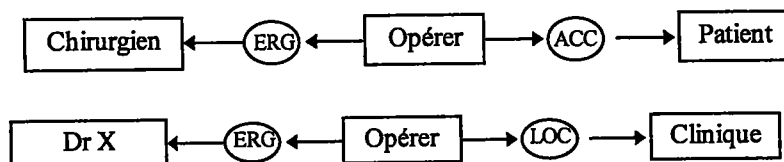
L'informatique linguistique s'est constituée autour d'une linguistique restreinte pour l'essentiel à la morphosyntaxe, et a produit toute une gamme d'analyseurs morphologiques et syntaxiques. (cf. Sabah, 1989). La linguistique informatique utilise l'informatique pour les besoins de la linguistique théorique d'une part, et appliquée d'autre part.

Il faut faire une distinction entre la forme et le contenu ; cette séparation n'est pas sans soulever certains problèmes, Wilks (1992) examine ces rapports et défend la primauté du contenu ; il ne faut pas baser les représentations du contenu sur des symboles arbitraires, elles doivent être interprétables en terme de langage naturel ; Jayez (1988) s'intéresse au problème illustré par la distinction entre formalisation qui repose sur l'idée d'une certaine continuité entre les langues et les langages formels qui sont censés en représenter le contenu et représentation qui est une conception minimaliste, qui partant d'un contenu donné, propose de l'inclure dans un formalisme dont le but est de faciliter certains traitements du contenu.

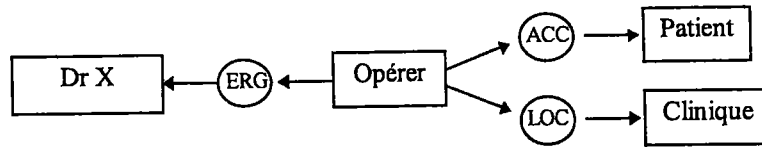
Il y a une équivalence entre le formalisme des graphes conceptuels et la logique des prédicats (Sowa, 1984) :

Graphe : [chirurgien] - (ERG) - [opérer]
Prédicat : ERG (/chirurgien/,/opérer/)

Si les graphes conceptuels sont utilisés pour représenter des contenus différentiels, on limite les relations à un petit nombre de primitives casuelles. Si les graphes conceptuels sont utilisés en tant que représentation informatique, on ajoute la notation des procédures formelles de manipulation de graphes : la projection qui consiste à restreindre un premier graphe à un fragment d'un second graphe, la jointure qui consiste à fusionner les parties identiques de deux graphes tout en ne conservant que les éléments les plus spécifiques (Landau, 1990) :



¹ F. RASTIER, C. CAVAZZA, A. ABEILLE : *Sémantique pour l'analyse, De la linguistique à l'informatique*, Masson, Paris / Milan / Barcelone, 1994, p. 2.



Assemblage de deux graphes conceptuels

Les graphes ont une notation permettant la présentation du contenu lexical et peuvent faciliter une description logique ultérieure. Il se pose toutefois un problème, en effet, il est difficile de décrire les réarrangements du contenu sémantique lors de l'interprétation en termes des procédures de traitement des graphes conceptuels.¹

Saint-Dizier (1992) a introduit la propagation des contraintes dans le TALN en intégrant à l'analyse syntaxique l'utilisation de traits sémantiques contenus dans le lexique. Il retient trois avantages de la programmation logique avec contraintes : « la plus grande efficacité par rapport à la programmation logique traditionnelle, la généralité et la réutilisabilité du code ainsi développé et l'expressivité plus grande du formalisme à base de contraintes ». « La propagation de contraintes ouvre également des perspectives pour décrire les transformations du contenu, puisqu'elle propose une solution à la recombinaison de structures de traits qui ne repose pas sur une approche purement logique ». La contrainte exprime la restriction de sélection. Une règle avec contraintes s'écrit de la façon suivante : Tête $\square \rightarrow$ Corps, {Contrainte}.

4.1.1. Lexique

La construction de structures où sont répertoriées les unités de la langue avec un certain nombre de propriétés nécessaires au traitement automatique est un des objectifs des dictionnaires électroniques.² L'homogénéité du codage est très importante. La base de données d'un dictionnaire électronique doit être formelle avec des entrées explicites et suffisantes, et exhaustive pour permettre la reconnaissance des mots déjà répertoriés, des mots délaissés dans les ouvrages « dictionnaires », des mots nouveaux non relevés mais prédictibles par construction.

Dans le cadre de la traduction automatique (TA), le dictionnaire joue un rôle très important, notamment en ce qui concerne les informations sémantiques. Les informations lexicales ont pour but de diriger la simulation informatique du processus de schématisation sous-jacent à la sélection des prépositions. Le plus difficile est de n'indiquer que les informations pertinentes. Zelinsky-Wibbelt indique les types de connaissances qui doivent apparaître dans le codage lexical : la disposition spatiale d'une entité, sa fonction typique, ses propriétés inhérentes / intrinsèques, les contraintes sur l'idéalisation.³ L'absence d'un trait dans un sémème doit toujours être explicitement exprimée, par exemple par sa négation.

¹ Cf. *op. cit.* [Rastier94], p. 99 sqq.

² A. DUGAS : « La création lexicale et les dictionnaires électroniques », in *Dictionnaires électroniques du français*, Langue Française 87, pp. 23-28, Larousse, Paris, 1990, p. 11.

³ W. RAMM : *Conceptual Information and the Interpretation of Spatial Prepositions : A comparison Between Approaches to Systemic-Functional Text Generation and Cognitive Based Machine Translation*, IAI Working Paper N° 25, Saarbrücken, 1992, p. 27.

Chaque discipline établit son lexique selon des modes différents : ¹

- * en psychologie cognitive, le lexique mental est important, il y a une tentative de reconstituer l'organisation du lexique en mémoire,
- * en intelligence artificielle, on utilise des réseaux sémantiques pour représenter les taxinomies^o lexicales, dans des domaines limités,
- * en linguistique informatique, on construit des dictionnaires électroniques pour les traitements automatiques.

Depuis dix ans, on assiste à une multiplication des grammaires lexicales qui stockent dans le lexique les traits morphologiques, sémantiques et syntaxiques et au développement des dictionnaires électroniques.

Pour poser correctement les problèmes de la description lexicale, il faut :

- * récuser les critères traditionnels selon lesquels un nom représente un objet, un verbe, une action... ; un verbe peut désigner une substance (neiger), un nom une action (course),
- * admettre que les descriptions et représentations puissent être différentes selon les disciplines,
- * admettre que les modes de structuration du lexique puissent varier selon les domaines et champs décrits.

Il existe deux types de codage :

- * méthode sémasiologique^o (traditionnelle) : elle code les unités lexicales par leur signifiant graphique auquel sont associés tous les signifiés qu'il peut véhiculer, puis structure leur inventaire. Cette méthode est utilisée par la lexicographie
- * méthode onomasiologique^o : elle part des classes de signifiés pour les structurer. Cette méthode est utilisée par la lexicologie.

Le but de l'assemblage du contenu sémantique est essentiellement de reconstituer les acteurs, qui sont représentés au moyen de molécules sémiques. Elles sont les seules structures interprétatives construites et servent de base à la représentation sémantique. Pour les besoins de l'implémentation, le contenu lexical doit être formalisé pour permettre la construction de ces structures interprétatives par réorganisation du contenu.

Le choix du formalisme structurant une description en traits différentiels est libre, mais il doit suivre les contraintes imposées par la tâche pratique.

Il faut faire attention aux ambiguïtés lexicales : *vers* [Prép.] et *vers* [N] par exemple.

Dans le cadre du traitement automatique, il est important, dans un dictionnaire, de préciser les éléments qui font partie du schéma de construction du terme et de les distinguer des éléments qui apparaissent dans son environnement sans dépendre de ses propriétés structurelles. C'est notamment le cas des verbes pour lesquels il faut préciser les éléments qui appartiennent à leur structure argumentale et qui sont nécessaires. Les autres éléments (compléments circonstanciels) seront traités dans le cadre de la grammaire mais pas dans celui du lexique. Attention, il arrive cependant que la limite

¹ Cf. *op. cit.* [Rastier94], p. 44 sqq.

entre ces deux types de compléments soit floue, surtout pour les compléments de temps, de lieu ou d'instrument « qui tantôt dépendent étroitement du verbe en raison des liens sémantiques très forts que celui-ci impose, tantôt au contraire disposent d'une certaine autonomie sémantique par rapport au verbe et s'appliquent à la phrase dans son ensemble ». ¹

Dans les formalismes informatiques, deux aspects pour les rapports forme-contenu : l'aspect statique qui correspond à la description du contenu dans le formalisme et l'aspect dynamique qui équivaut aux procédures de traitement.

Les formalismes lexicaux sont bien développés en syntaxe mais très peu développés en sémantique. Pustejovsky (1991) travaille sur le lexique génératif. Son objectif est de définir un formalisme général permettant une décomposition lexicale, ces structures servant de base à la compositionnalité sémantique. Il utilise la notion de qualia ² inspirée d'Aristote.

« Dans la théorie du lexique génératif, une entrée lexicale est considérée comme un prédicat, décrit par la structure de qualia. Elle se compose de quatre facettes : le rôle constitutif, qui correspond à la relation entre le prédicat et ses constituants, le rôle formel, ensemble des caractéristiques qui distinguent le prédicat à l'intérieur d'un domaine, le telic ou fonction du référent et le rôle agentif, qui regroupe différents facteurs ». ³

Ex. porte (x,y)

rôle constitutif : ouverture (y)
rôle formel : objet-physique (x)
telic : passer-par (x,y)
rôle agentif : meuble (x)

attention le contexte n'est pas pris en compte.

La première étape consiste à étudier le corpus en identifiant les phénomènes sémantiques avant la formalisation dans le cadre différentiel. Dans une deuxième étape, on décrit, dans le formalisme différentiel, à partir du corpus, les traits inhérents (ainsi que certains traits afférents stabilisés) de chaque entrée lexicale.

Les champs sémantiques sont les classes de généralité supérieure au taxème, les domaines sont en nombre restreint dans un texte relevant d'une spécialité technique, les dimensions sont décrites en dernier lieu.

Dans la description du contenu, on n'introduira pas nécessairement de distinction entre sèmes génériques et sèmes spécifiques. La distinction entre ces deux types de sèmes a été établie par B. Pottier.

4.1.2. Syntaxe

En traitement automatique, l'analyse syntaxique consiste à associer à la chaîne découpée en unités, une représentation des groupements structurels entre ces unités ainsi

¹ A. BORILLO : « A propos de la localisation spatiale », in *Sur les compléments circonstanciels*, Langue Française 86, pp.75-84, Larousse, Paris, 1990, p. 75.

² Une structure de qualia regroupe les attributs essentiels de l'objet ou de l'action désignés par l'entrée lexicale, dans une perspective qui est donc nettement référentielle

³ Cf. *op. cit.* [Rastier91], p. 96.

que des relations fonctionnelles qui unissent les groupes d'unité. L'analyse syntaxique est un intermédiaire entre l'analyse morphologique et l'analyse sémantique.

L'analyse syntaxique nécessite une représentation plus complexe. Il n'est pas obligatoire de reconnaître des groupements de constituants du niveau supérieur pour analyser une phrase. On peut, en se déplaçant, sur la phrase reconnaître une catégorie grammaticale pour un item lexical et prévoir le suivant avec un certain degré de certitude. On obtient une suite de catégories grammaticales que l'on peut regrouper en unités du niveau supérieur.

La représentation en arbres de structures syntaxiques est basée sur l'identification des unités ; la structure est hiérarchisée, l'arbre possède des nœuds étiquetés.

4.1.3. Sémantique

« L'utilisation de la théorie sémantique pour des applications informatiques consiste en premier lieu à transposer une lecture descriptive en lecture réductive, c'est-à-dire à sélectionner les unités sémantiques pertinentes pour la tâche ». ¹

Pour le traitement automatique, il faut que la description du contenu soit en continuité avec le texte.

La sémantique cognitive subordonne le langage aux concepts, et le contenu linguistique n'y est jamais que l'expression dans la langue des concepts qui sont soumis aux lois propres d'une sphère conceptuelle. ²

La sémantique différentielle, qui dispose du texte en tant qu'objet empirique, est en mesure de décrire la diversité des contenus. C'est une théorie complète de l'interprétation reposant notamment sur une description du contenu lexical.

4.2. LE TRAITEMENT ET LA TRADUCTION DES PREPOSITIONS

A.-M. Berthonneau et Pierre Cadiot ne sont pas certains que l'on puisse implémenter les prépositions.

« Ce n'est pas dire qu'à terme on ne puisse exploiter et valoriser ces descriptions dans le sens d'une implémentation, ou au moins d'une opérationnalisation. Les prépositions sont notamment un bon objet pour les modélisations de type connexionniste. » ³

Au niveau de la description syntaxique, il faut identifier les unités hypotaxiques. On va de l'analyse de l'interprétation de la chaîne de surface à l'identification des groupes structurels.

Si une préposition a une valeur casuelle, la préposition peut être inscrite comme trait particulier du lexème régissant les étiquettes des entrées lexicales du dictionnaire d'analyse.

¹ Cf. *op. cit.* [Rastier94], p. 16.

² Nous citons la note 1 p. 84 de *op. cit.* [Rastier94] : *Selon les théories, cette sphère conceptuelle peut avoir un fondement géométrique ou topologique (Pottier 92) inspirée par le monde physique, ou au contraire, demeurer purement symbolique et former une sorte de langage interne ou « langage de la pensée » (Fodor 75).*

³ Cf. *op. cit.* [Berthonneau93], p. 12.

La forme graphématique de la préposition est donc un signal pour la construction d'une unité syntagmatique dans laquelle le groupe prépositionnel représente la partie subordonnée du syntagme. C'est la rection du lexème supérieur indiquée par la préposition introduisante. Dans l'exemple *Nous glissons sur ce problème*, la rection du verbe indiquée par la préposition *sur* permet de construire une unité syntagmatique du type « vb-GP ».

L'analyse ne peut pas se baser que sur les formes graphématiques ; il faut pouvoir faire la différence entre un objet prépositionnel et un adverbe. Il faut voir si la distribution de la préposition est rattachée à la découverte des valences syntaxiques. On commence par l'analyse de ces dernières en s'appuyant sur l'unité lexicale supérieure de la structure hypotaxique (c'est-à-dire le verbe et sa rection), puis on continue par l'étude de la préposition et de sa distribution immédiate (c'est-à-dire le groupe nominal qui la suit). Dans le groupe prépositionnel à valeur adverbiale, le substantif ne se rattache pas à la préposition sans opérer une actualisation de certains traits et certaines restrictions sémantiques concernant la combinaison entre la préposition et le nom.

On peut faire la différence entre objet prépositionnel et adverbe grâce à une étude des traits sémantiques des noms qui suivent la préposition, c'est-à-dire les indices inhérents tels que indication de temps, de qualité, de lieu, d'état, d'instrument, d'événement... Les indices sont donnés dans les entrées lexicales, traités par ordinateur et ainsi on peut établir des règles concernant, par exemple, l'identification de valence casuelle telle que : s'il y a dans une phrase un verbe avec une rection caractéristique en forme de groupe nominal introduit par une préposition *à*, et s'il y a dans la même phrase plusieurs candidats à cette rection, le nom de l'objet prépositionnel ne doit en aucun cas être muni de traits sémantiques tels que « temps ».

La firme a fusionné avec une autre [- qualité → objet prépositionnel]
La firme a fusionné avec prudence [+ qualité → adverbe]

Une stratégie d'analyse exploitant les traits sémantiques inhérents permet dans la plupart des cas de reconnaître les fonctions et de lever les ambiguïtés à ce niveau de distinction des valeurs fonctionnelles en cas d'ambiguïtés « objet prépositionnel - adverbe ».

Il faut ensuite établir les restrictions sémantiques entre la préposition et sa cooccurrence, et séparer les éléments prépositionnels casuels à valeur morphématique et les éléments sémantiquement autonomes.

Il est nécessaire d'éviter une liaison entre le verbe et l'adverbe appartenant au cadre verbal car la préposition introduisant le groupe adverbial ne serait plus soumise à l'analyse sémantique bien qu'elle ait sa propre valeur sémantique, il est donc impératif, dans le dictionnaire d'analyse, d'entrer des indications de connexion pour le verbe et la préposition qui effectivement se rattachent casuellement, c'est-à-dire qui sont en général équivalents aux cas de surface caractérisés par des morphèmes flexionnels (*permettre à, répondre à...*).

On peut analyser les prépositions quant à leur appartenance à une structure profonde : si la préposition introduit un syntagme adverbial, elle garde sa qualité d'élément autonome dans la structure profonde. Si elle est seulement un indice de cas superficiel,

elle est interprétée comme élément de surface et non comme élément autonome dans la structure profonde.

On a alors deux types de prépositions : les prépositions vides traduites au niveau morphématique et les prépositions pleines qui seront analysées plus en détail afin d'obtenir une traduction correcte.

On se heurte à des problèmes de traduction quand une préposition peut avoir plusieurs traductions : par exemple *auf* signifie *sur* dans *Der Bleistift liegt auf dem Tisch*, mais *auf* ne sera pas traduit par *sur*, et *sur* par *auf* dans les énoncés suivants : *J'ai de l'argent sur moi*, *Antworte auf meine Frage*.¹ Un des problèmes principaux de l'analyse automatique des prépositions est de savoir si la préposition identifiable en structure de surface, l'est aussi en structure profonde.

Les prépositions sont très polysémiques. Leur traduction dépend des modèles d'usage spécifiques fournis par la langue source et la langue cible. Il est difficile d'effectuer une traduction basée sur un appariement direct entre les deux langues.

Que les deux langues possèdent ou non des prépositions ayant une étymologie commune, il sera impossible de toujours traduire, et ce quel que soit son emploi, une préposition de l'une des langues par la même préposition dans l'autre langue. La correspondance entre des prépositions de deux langues ne peut être que celle d'éléments dont le schème représentatif est identique ou du moins assez proche. La correspondance ne devient effective qu'à partir du moment où l'on recherche les contours de l'identité et de la différence (les différences ne doivent pas l'emporter sur les ressemblances).

La correspondance entre deux éléments X et Y appartenant à deux langues différentes, c'est le fait que, dans la traduction, c'est par Y que l'on rend le plus souvent X et par X que l'on rend le plus souvent Y.

L'allemand dispose d'une deuxième possibilité de codification des informations sémantiques pertinentes précises grâce à deux caractéristiques spécifiques de la langue :

- * l'existence de lexèmes P transitifs qui peuvent attribuer différents cas
- * la possibilité de deux sérialisations différentes à l'intérieur du complexe P1 et par là même une structure du groupe prépositionnel différente, dans laquelle la tête P0 de la phrase peut apparaître soit à droite soit à gauche.²

Comme les prépositions peuvent présenter des ambiguïtés monolingues, il est d'autant plus difficile de faire un appariement entre les prépositions de la langue source et celles de la langue cible. Pour essayer de lever ces ambiguïtés, il convient de faire une analyse plus poussée en exploitant des informations complémentaires concernant notamment le contexte syntaxique et sémantique de la préposition.³

¹ Cf. *op. cit.* [Marcq72], p. 7.

² N. FRIES : *Präpositionen und Präpositionalphrasen im Deutschen und im Neugriechischen*, Max Niemeyer Verlag, Tübingen, 1988, p. 143.

³ W. RAMM : *Conceptual Information and the Interpretation of Spatial Prepositions : A Comparison Between Approaches to Systemic-Functional Text Generation and Cognitive Grammar Based Machine Translation*, IAI Working Paper N° 25, Saarbrücken, 1992, p. 3.

Trois paramètres entrent en compte pour la génération des prépositions dans la grammaire de Penman :

- * le type d'élément « circonstanciel » (le type de groupe prépositionnel) à générer, c'est-à-dire le type de relation que le groupe prépositionnel doit exprimer,
- * le type du concept « objet » à réaliser comme complément de la préposition, c'est-à-dire quelle entité peut être reliée à telle autre grâce à une préposition,
- * le type « procès » de toute la phrase.¹

Si on ne fait pas intervenir de contraintes sémantiques, on risque de générer des constellations « circonstance / objet » sémantiquement incompatibles.

Ce qui manque dans l'implémentation de nombreux systèmes de TALN à l'heure actuelle, c'est la prise en compte systématique des informations contextuelles afin de limiter les interprétations sémantiquement possibles, de contrôler le choix d'une perspective sous laquelle on doit présenter un objet ou un événement.

Il faut en premier lieu lever les ambiguïtés sémantiques, puis faire une description, c'est-à-dire établir une relation en partant d'une valeur lexématique désambiguïsée pour élucider les circonstances de leurs rôles sémantiques dans la phrase. Si on a une signification lexématique on a aussi une information sur le rôle relationnel de la préposition au niveau sémantique de la phrase.

Une situation spécifique de l'énoncé entraîne l'emploi et le choix de telle ou telle préposition, il faut faire une analyse sur l'exploitation de l'environnement de la préposition dans la chaîne lexématique.

On analyse les différents membres d'un groupe prépositionnel : le contenu sémantique d'une préposition est souvent déterminé par le contenu des éléments nominaux situés à droite de la préposition. L'exploitation des traits morphosyntaxiques dans le groupe prépositionnel ne donne pas de clarification satisfaisante du contenu de la préposition, c'est pourquoi on s'appuie plus sur une information sémantique du groupe nominal suivant la préposition. Toutefois le rôle des classes nominales en combinaison avec l'information morphosyntaxique est limité car le français est une langue non flexionnelle.

Le livre est sur la table	}	si on analyse le nom situé après la préposition, on a une indication de lieu mais on n'a pas d'indication de mouvement ou de direction, il faut donc élargir le contexte.
Il pose le livre sur la table	}	

La traduction de prépositions ou de groupes prépositionnels est réalisée grâce à la traduction d'un type de schématisation d'une langue en un type de schématisation dans une autre langue. Les différentes étapes sont (Zelinsky-Wibbelt, en préparation, p. 17)² :

1. la schématisation de la relation spécifique à une langue exprimée par la préposition dans la langue source,

¹ Cf. *op. cit.* [Ramm92], p. 12.

² *Ibid.*, pp.27-28.

2. l'extraction (à partir de cette relation spécifique à une langue) d'une relation qui est supposée constante entre les langues,
3. le transfert des résultats de l'analyse dans la langue cible qui implique :
 - la relation extraite du sens prépositionnel de la langue source,
 - la situation de discours inférante,
 - les unités lexicales sémantiquement contraintes par la préposition, i.e., fonctionnent comme trajector et landmark,
4. la spécification de la relation transférée en une relation spécifique dans la langue cible en l'unifiant avec les informations spécifiques à une langue fournit les autres unités lexicales transférées, c'est-à-dire, de déterminer le type de schématisation sous-jacent à une expression équivalente dans la langue cible,
5. la génération de l'expression dans la langue cible à partir de la relation obtenue dans la dernière étape.

Selon Johannes Ritzke :

« un des problèmes principaux pour l'analyse automatique des prépositions, consiste à savoir si une préposition, en tant qu'élément concret identifiable dans la structure superficielle, le restera encore dans la structure profonde, base commune des langues qui sont à traduire ».

On peut reconnaître assez facilement les éléments prépositionnels dans une phrase grâce à leur forme graphématique.¹

Quelques difficultés se présentent pour l'analyse automatique :

- * les éléments prépositionnels sont chargés de fonctions différentes en ce qui concerne leur syntaxe et leur sémantique,
- * une forme graphématique peut exprimer plusieurs fonctions.²

Il existe deux types de groupes prépositionnels :

- * l'objet prépositionnel à valeur casuelle (cas de surface) et lexématiquement figée (*l'édifice s'impose à notre admiration ; le savant penche plutôt pour l'autre hypothèse.*)
- * l'adverbe, avec une préposition à valeur sémantique autonome (*les voitures glissent sur le verglas ; elle a été prise à Milan, au siège de la maison mère.*)

Les prépositions sont identifiées grâce à leur forme graphématique, donc, dans le cadre du traitement automatique, la grammaire d'analyse doit résoudre les ambiguïtés fonctionnelles pour faire la distinction entre ces deux types, grâce par exemple à une étude sur les traits sémantiques des noms qui suivent la préposition. Les traits sémantiques représentent des indices inhérents qui sont assignés aux entrées lexicales des noms et traités par les programmes informatiques, de sorte qu'il est possible de formuler des règles de grammaire. On obtient une distinction des valeurs fonctionnelles mais pas une détermination détaillée du sens de la préposition.

¹ C. SCHWARZE : *Analyse des prépositions*. IIIème Colloque franco-allemand de linguistique théorique du 2 au 4 février 1981 à Constance, Max Niemeyer Verlag, Tübingen, 1981, p. 141.

² *Ibid.*, p. 142 sqq.

Dans le dictionnaire d'analyse, il faut affecter d'indications de connexion uniquement les verbes et les prépositions qui se rattachent casuellement, c'est-à-dire qui sont en général équivalentes aux cas de surface caractérisés par des morphèmes flexionnels (*permettre à, répondre à...*).

Une préposition, en structure profonde, garde sa qualité d'élément autonome, si elle introduit un syntagme adverbial.

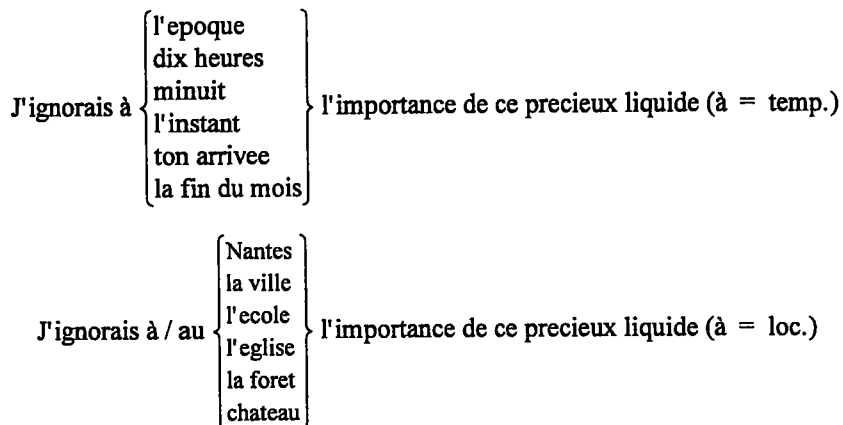
La traduction des prépositions vides, à valeur casuelle (cas de surface) est réalisée au niveau morphématique. En revanche, l'analyse des prépositions pleines a un rôle très important dans le système pour obtenir une traduction correcte.

Deux étapes complémentaires dans l'analyse sémantique :

- * éclaircir la sémantique lexicale de la préposition en résolvant les ambiguïtés lexématiques,
- * élucider les circonstances de son rôle sémantique dans la phrase.

Il y a un ordre à respecter dans le processus analytique : « L'analyse sémantique des prépositions commence par la définition et l'inventaire des variantes sémantiques de chaque préposition comme entrée lexicale ». Il faut tenir compte aussi bien d'informations morphosyntaxiques que d'informations sur des classes sémantiques de l'environnement nominal.

En français, on a pu remarquer que « l'exploitation des traits morphosyntaxiques à l'intérieur d'un groupe prépositionnel ne conduit pas à une clarification satisfaisante du contenu de la préposition, excepté certains cas simples où la catégorie lexématique de l'élément suivant est suffisante (*Dans peu, il y aura du soleil*) ». On accorde donc plus d'importance à la mise en valeur d'informations sémantiques du groupe nominal qui suit la préposition. Les exemples ci-dessous ont été présentés dans [Schwarze81], page 150.



Une analyse qui ne porte que sur les éléments qui suivent la préposition est insuffisante, il faut aussi tenir compte des éléments qui la précèdent. Il faut utiliser un contexte élargi.

Dans cette deuxième étape, on tiendra essentiellement compte des verbes, qui sont presque toujours des facteurs déterminatifs dans une relation adverbiale.

Les verbes jouent alors un grand rôle : les classes verbales précisent et déterminent une valeur sémantique exacte :

- * verbes de mouvement orienté : *tomber, monter...*
 - impliquant un point de départ : *venir, s'enfuir de + GN,*
 - impliquant une arrivée : *aller, arriver à / dans + GN,*
- * verbes de mouvement non orienté : *marcher, se promener, flâner pendant / sous + GN*
- * verbes duratifs : *sommeiller, fumer, dormir, se jouer...*
- * verbes inchoatifs : *rougir, pâlir, se mettre en route, se liquéfier...*
- * verbes perfectifs : *terminer, finir, trouver...*

Plus la classification est raffinée, meilleure est l'analyse ; mais il faut cependant faire attention à ne pas mettre trop d'informations (seules les informations pertinentes sont utiles) sinon le temps de calcul sera trop important.

En conclusion, le procédé de l'exploitation automatique est le suivant : « en partant d'une interprétation de faits grammaticaux existant dans la chaîne graphématique, nous parvenons à une description grammaticale qui nous procure à la fin la résolution des ambiguïtés ». ¹

4.3. LES PROJETS D'IMPLEMENTATION UTILISANT LES GRAMMAIRES D'UNIFICATION

4.3.1. Les projets d'implémentation utilisant UCG

SYSLID : SYntaktisch-Semantische Linguistikkomponente für sprachverstehende Dialogsysteme

Partenaires du projet : Dr.-Ing. Ute Ehrlich, Daimler-Benz AG Forschung und Technik Sprachverstehende Systeme (F3S)

Dans le projet SYSLID, on développe une composante pour l'analyse linguistique du sens dans des systèmes de dialogues en langue naturelle. Le but de tels systèmes de dialogue est de permettre aux utilisateurs un accès facile aux bases de données informatiques, en simulant la compréhension linguistique et les capacités de communication d'un agent de renseignement humain.

Pour comprendre la langue orale dans les systèmes de dialogue, il faut résoudre deux tâches importantes.

- * La reconnaissance acoustique des mots doit, à partir du signal linguistique digitalisé, retrouver les mots prononcés. Pour effectuer cette tâche, on compare le signal des expressions de l'utilisateur à analyser avec le modèle linguistique qui est généré dans une phase d'entraînement. Le travail est d'autant plus difficile que, dans le signal acoustique d'une phrase prononcée sans intonation

¹ Cf. *op. cit.* [Schwarze81], p. 156.

particulière, les intervalles entre les mots sont difficilement perceptibles. De plus, la prononciation d'un mot peut varier d'un homme à un autre de même qu'elle peut être modifiée dans situations différentes. La composante de reconnaissance des mots ne peut donc livrer que des hypothèses sur ce qui peut effectivement avoir été dit.

- * L'analyse linguistique du sens doit, à partir des hypothèses fournies par la reconnaissance des mots, rechercher la signification de ce qui a été dit. Pour ce travail on utilise des formalismes de représentation et des procédés d'analyse de la linguistique informatique et de l'intelligence artificielle. En raison de l'incertitude de la reconnaissance acoustique des mots, il est très important d'utiliser des procédés d'analyse robustes. Dans le contexte de la compréhension linguistique informatique, on entend, par robustesse, la capacité d'un système à tolérer les violations de règles grammaticales jusqu'à un certain point. Ceci est nécessaire car, d'une part la langue parlée spontanément contient souvent des irrégularités grammaticales et, d'autre part, en raison des erreurs qui résultent de la reconnaissance des mots, on peut laisser passer des hypothèses agrammaticales pour la composante de compréhension de la langue.

La deuxième tâche, c'est-à-dire le développement d'une composante linguistique robuste, est traitée dans SYSLID. Ce travail comprend deux sous-domaines :

- * le développement d'une grammaire formelle, qui définit les combinaisons entre les mots (syntaxe) et la génération d'une représentation du sens (sémantique) lors d'une telle combinaison. Le développement grammatical en SYSLID s'effectue grâce au formalisme UCG (*Unification Categorial Grammar*). Dans cette optique, le dictionnaire a le rôle central dans la représentation du savoir linguistique, car l'information concernant les combinaisons possibles est codée en premier lieu dans les entrées des mots simples du dictionnaire.
- * le développement d'un parser, grâce auquel on assigne aux hypothèses de la composante de reconnaissance des mots une représentation de la signification. Cette composante fait passer au parser une quantité d'hypothèses, évaluées au niveau acoustique, sous forme d'un graphe de mots. Il a pour tâche de chercher le meilleur chemin grammaticalement admissible parmi les graphes. La grammaire formelle sert de base de savoir linguistique au parser. Les travaux menés dans SYSLID sur le parser recherchent avant tout, la façon d'augmenter l'efficacité et la robustesse de l'analyse.

Un système automatique de renseignements concernant les trains, auquel on pouvait demander les liaisons intercity entre deux villes allemandes, a été choisi comme application prototypique dans la première phase du projet de 1992 à 1994. La composante linguistique développée dans SYSLID, qui se compose d'une grammaire pour le domaine d'application « renseignements ferroviaires » et d'un parser robuste de graphes de mots, a été intégrée dans le système de démonstration de Daimler-Benz, partenaire du projet.

Dans la phase actuelle du projet, des grammaires sont développées pour deux applications dans le domaine des systèmes d'organisation des transports.

- * Dans la première, il s'agit d'un projet de demande de modifications du réseau de transport dans la base de données du système STORM (*Stuttgart Transport Operation by Regional Management*). La demande se fait grâce à un dialogue en langue naturelle entre l'homme et la machine.
- * Dans la seconde, les recherches portent sur un scénario en langue naturelle de demande radio en voiture. Ainsi cela doit permettre au conducteur, dans un dialogue ciblé avec l'ordinateur de bord, de filtrer, parmi la quantité d'informations concernant la circulation, celle qui lui convient.

Les deux projets sont réalisés par Daimler-Benz à Ulm, la partie développée dans le projet SYSLID est intégrée comme composante linguistique d'analyse.

4.3.2. Les projets d'implémentations utilisant LFG

4.3.2.1. LE PROJET LFG PARGRAM

Le projet LFG PARGRAM représente un effort de collaboration entre les chercheurs de Xerox PARC à Paolo Alto, du Centre de Recherche Rank Xerox à Grenoble (France), et de l'Université de Stuttgart (Allemagne). L'objectif de ce projet est la production de grammaires importantes pour l'anglais, le français et l'allemand, écrites en collaboration, et basées sur un ensemble commun de principes linguistiques avec un ensemble défini en commun de traits grammaticaux. La grammaire anglaise est écrite à Xerox PARC par Tracy King, Maria-Eugenia Nino, et Mary Dalrymple, la grammaire française au centre de Recherche Rank Xerox par Frédérique Segond et Caroline Brun, et la grammaire allemande à l'Université de Stuttgart par Christian Rohrer, Miriam Butt, et Christian Fortmann.

Le projet PARGRAM est actuellement mené dans deux environnements de développement grammatical différents. Au début du projet, les trois sites utilisaient le *Xerox LFG Grammar Writer's Workbench*, un environnement de développement grammatical écrit en Medley Lisp, une variante de Interlisp (Kaplan et Maxwell 1993). Le *Grammar Writer's Workbench* est une implémentation d'analyse complète du formalisme syntaxique de LFG, incluant divers traits intégrés depuis l'introduction de la théorie LFG par Kaplan et Bresnan (1982), tels que l'incertitude fonctionnelle, la précedence fonctionnelle, la généralisation pour la coordination, et des projections multiples. Il inclut une notation de règles en structure c très riche, plus différentes sortes d'outils abrégatifs (dont des formes-types^o paramétrées, des macros). Plus récemment, les sites Xerox ont été transférés vers une nouvelle plate-forme, l'environnement linguistique Xerox, qui est conçu dans le même esprit que le *Grammar Writer's Workbench* mais tourne sous Unix et Tcl/Tk au lieu de l'environnement de programmation Medley. Le noyau du projet est un parser efficace basé sur l'unification (Maxwell et Kaplan 1991). L'environnement linguistique Xerox est toujours en développement ; actuellement (mars 1996) il implémente les mêmes traits que le *Grammar Writer's Workbench* à l'exception de l'incertitude fonctionnelle et de la

La quasi-totalité des projets d'implémentation présentés dans les différentes grammaires a été trouvée sur Internet et traduite pour plus de clarté.

généralisation pour la coordination. Un générateur basé sur l'unification est également en développement.

La grammaire se compose de règles syntaxiques et de macros abrégatives ; LFG autorise les expressions régulières du côté droit des règles syntaxiques (y compris la notation en étoile de Kleene) et les combinaisons booléennes arbitraires de prédicats réguliers, ainsi, dans la grammaire, les règles réduisent effectivement un grand nombre de règles écrites dans un cadre plus conventionnel (Kaplan et Maxwell 1993). Les dictionnaires utilisés se composent d'entrées pour les stemmes, de définitions de formes-types, et de règles lexicales. L'environnement linguistique Xerox permet une interface avec un analyseur morphologique d'états finis externe, et ainsi les sites utilisant cet environnement de développement complètent leurs dictionnaires avec des entrées pour l'information concernant la flexion morphologique fournie par le parser. Actuellement, le *Medley Grammar Writer's Workbench* n'autorise pas l'utilisation de parser morphologique ; en revanche, des règles de découpage des suffixes sont introduites, ce qui permet ainsi de reconnaître les terminaisons flexionnelles de mots dans des langues comme l'anglais. Les sites utilisant cet environnement de développement utilisent des tableaux morphologiques pour encoder ces terminaisons suffixales et les informations qu'elles renferment.

Bon nombre de constructions principales sont couvertes par les grammaires anglaise, française et allemande. Toutes les grammaires couvrent les déclaratives et les impératives, y compris la négation, les grammaires française et allemande les impératives à la deuxième personne de même que les constructions infinitives impératives, la grammaire allemande les interrogatives standards, la grammaire anglaise les interrogatives standards de même que les interrogatives principales et enchâssées, les constructions déclaratives en incluant les inversions existentielles et négatives.

Les trois grammaires couvrent une série de modifications nominales : les déterminants, les adjectifs, la composition des noms, les propositions relatives, et les modificateurs nominaux des groupes prépositionnels. Les grammaires anglaise et allemande couvrent également les modifications de noms par les groupes verbaux et les adjectifs déverbaux. L'interaction compliquée entre le cas et l'accord nominal (flexion de l'adjectif et du déterminant), de même que l'occurrence des GNs sans tête (par ex. : « der blaue ») se retrouvent également dans la grammaire allemande. Toutes les grammaires couvrent différentes constructions de coordination.

Les trois grammaires incluent des formes-types parallèles pour un grand nombre de sous-catégorisations verbales et de diathèses, y compris le passif, les verbes à particules, divers types de contrôle, les constructions prédicatives, les groupes prépositionnels sous-catégorisés, et les séquences et sélections auxiliaires. La grammaire allemande permet également les constructions casuelles bizarres (par ex. : « helfen » - aider) et leur interaction avec la passivation.

Les trois grammaires couvrent une série de modifications verbales, y compris les adverbes et leur place, et les conjonctions de subordination. Les grammaires française et allemande fournissent différentes sortes de résolution temporelle (temps composés), de constructions réflexives, et de pronoms clitiques pour le français et explétifs pour

l'allemand. La grammaire anglaise couvre également différentes constructions spécialisées (par ex. : l'expression du temps).

Enfin, la grammaire allemande couvre les différentes possibilités d'ordre des mots et de topicalisation que l'on trouve en allemand. La distribution de la topicalisation des GV est notamment entièrement implémentée.

4.3.2.2. LE XEROX LFG GRAMMAR WRITER'S WORKBENCH

Contact : Ron Kaplan

Le *Xerox LFG Grammar Writer's Workbench* est une implémentation d'analyse complète du formalisme syntaxique LFG, comprenant également les différents traits introduits depuis le papier d'origine KB82 (incertitude fonctionnelle, précedence fonctionnelle, généralisation pour la coordination, projections multiples, etc.). Développé par Ron Kaplan et John Maxwell, il comprend une notation très riche des règles en structure c plus différentes sortes de outils abrégatifs (dont des formes-types paramétrées, des macros). Le système possède une interface pilotée par souris élaborée pour afficher les différents structures et sous-structures grammaticales - l'idée est d'aider le linguiste à comprendre et à éliminer les erreurs d'une grammaire sans avoir à comprendre les détails des algorithmes spécifiques du traitement.

Le *Workbench* tourne sur la plupart des systèmes Unix (Sun, DEC, HP...) et sous DOS sur PC, bien que la plupart des expérimentations aient eu lieu sur Sun. Il n'a pas d'interface télétype - il fonctionne seulement comme un programme graphique (et l'interface n'est pas à l'heure actuelle celle qui est nécessaire sur de petits écrans standards PC). Il requiert au moins 16 MB de ram, plus un disque de 40MB ou plus.

Le *Workbench* est écrit en Medley Lisp (un descendant de Interlisp), on peut obtenir une licence gratuite chez Xerox. Xerox fournit également le *Workbench* gratuitement pour une utilisation en recherche et en enseignement.

4.3.2.3. LE KONSTANZ LFG WORKBENCH

Contact : Bruce Mayo

À l'Université de Constance (Allemagne), il existe un *LFG-Workbench* simple qui est utilisé pour un cours d'introduction à LFG. Il accepte les règles syntaxiques dans une notation conventionnelle avec des équations fonctionnelles simples (pas d'opérateurs booléens) et des équations de contraintes (=c), et permet à l'utilisateur de projeter une structure f sur un ensemble d'implications sémantiques via une sémantique lexicale écrite selon la notation Prolog. La version de B. Mayo tourne sur Macintosh avec le Prolog II+ de Marseilles, qui est disponible pour bon nombre d'autres ordinateurs, y compris les machines fonctionnant avec MS-DOS et Unix. Ses entrées et sorties linguistiques sont en ligne, mais les dialogues utilisateurs utilisent essentiellement les menus et les fenêtres du système Prolog. B. Mayo n'a pas essayé de le tester ou de l'adapter sur d'autres environnements.

4.3.2.4. LE SYSTEME LFG DE AVERY ANDREW

Avery Andrew possède un petit système LFG qui tourne sur PC (sur XT, en fait), qui, à la base, est élaboré pour produire de petits fragments pour illustrer les aspects de l'analyse grammaticale en LFG. Il utilise quelques notations non standards (en général dans un souci de brièveté), il lui manque cependant quelques éléments qui devraient pourtant être présents (comme par exemple un traitement des dépendances à grande distance), mais il a un composant morphologique primitif, quelque chose que Avery Andrew trouve essentiel pour une utilisation.

4.3.2.5. UN PARSER LFG POUR LE TURC

Auteur : Zelal Gungordu

Il intègre un analyseur morphologique à deux niveaux, des règles LFG utilisant le formalisme du parser CMU-CMT. Il couvre 80 % de la langue turque sans pré-édition, et même plus avec un minimum de pré-édition. Il fonctionne relativement vite. Il ne traite pas certaines classes de phrases telles que les conditionnelles et les causatives, et possède peu de données concernant la ponctuation. En sortie, on obtient une structure de traits LFG bien que les notions LFG de sujets et d'objets ne soient pas toujours bien observées.

4.3.3. Les projets d'implémentation utilisant GPSG

4.3.3.1. GPSG-TOOLS

Auteurs : Carla Umbach, Guido Dunker

Type : parser

Tâche : Traduction Automatique

Ce sont des outils utilisés pour transformer des fichiers de définition externes en format interne (règles Prolog). Chaque outil comprend :

- * un parser
- * un analyseur de contexte
- * un générateur qui produit des règles Prolog utilisées dans le système de traduction automatique de Berlin (Projet Kit-Fast).

4.3.3.2. GPSG PARSER

Auteur : Wilhelm Weisweber

Type : parser

Tâche : Analyse syntaxique et développement d'une version constructive de GPSG.

Le parser graphique ascendant est intégré dans un système de traduction automatique expérimental. Il interprète directement le format ID/LP et les métarègles.

Il se compose d'un analyseur morphologique (sur la base de SUTRA) et d'un parser. Les données sont des règles de grammaire allemandes et un lexique.

4.3.3.3. KIT-FAST

Le projet Fast, qui fait partie du groupe de projet Kit, a développé et implémenté un système de traduction automatique (TA) expérimental. Les représentations syntaxiques et sémantiques des phrases sont des structures générées respectivement par des *Generalized Phrase Structure Grammars* (GPSG) et des *Functor-Agreement-Structures* (FAS). Le niveau conceptuel de représentation est réalisé grâce à BACK, un système de représentation des connaissances basé sur KL-ONE, développé par le projet Kit-Back et intégré dans le système de TA. Le contenu du texte se trouve dans la Abox (connaissances assertives) du système BACK, et les connaissances linguistiques dans la Tbox (connaissances terminologiques).

Le système Kit-Fast est basé sur le transfert. La traduction d'une phrase se compose d'une analyse morphologique, syntaxique, sémantique et conceptuelle, d'un transfert, d'une génération et d'une synthèse morphologique. Les algorithmes pour l'analyse et la synthèse sont basés sur le système SUTRA. L'analyse syntaxique est réalisée par un parser GPSG qui interprète directement les règles ID, les états LP et les métarègles. Les analyses sémantique et conceptuelle, le transfert ainsi que la génération sont réalisés par un algorithme sur la base d'une réécriture de termes. Après l'analyse sémantique, l'expression FAS (*Funktor-Argument-Struktur*, structure foncteur-argument) résultante est analysée conceptuellement, c'est-à-dire qu'elle est appariée avec une expression du ABox-Tell-Language (ATL), grâce auquel le contenu de la phrase est ajouté à la représentation du contenu du texte dans la Abox du système BACK.

La première étape vers la traduction de textes a été d'interpréter les relations anaphoriques dans la langue source. C'est pour cette raison qu'un algorithme pour les relations anaphoriques a été développé et implémenté.

Le système de TA utilise deux représentations textuelles : une pour l'information structurelle du texte, et une pour le contenu du texte. Au cours de la traduction, les représentations sont construites de façon incrémentielle à partir des représentations de phrases. En principe, une représentation textuelle est nécessaire à chaque niveau, mais cela conduirait à des représentations redondantes aux niveaux syntaxique et sémantique. C'est pour cette raison, que les auteurs ont décidé de prendre un niveau sémantique plus général (FAS) pour la représentation des aspects structurels du texte. Le niveau FAS¹ représente le texte comme une série de représentations syntaxiques FAS et reflète ainsi l'ordre linéaire de même que les caractéristiques structurelles spécifiques du texte.

Les composantes du système de TA sont les suivantes :

- * un analyseur morphologique basé sur le système SUTRA,
- * un parser GPSG pour l'interprétation des règles ID, des états LP et des métarègles,
- * un interpréteur de règles de réécriture pour l'analyse sémantique et conceptuelle, le transfert et la génération,
- * un synthétiseur morphologique basé sur le système SUTRA,
- * un module pour l'évaluation des relations anaphoriques,
- * le système de représentation des connaissances BACK,

¹ W. RAMM : *Grundlagen für eine wissensbasierte Maschinelle Übersetzung*, IAI Working Paper N° 22, Saarbrücken, 1991, p. 22.

- * des outils pour le développement de lexiques, de grammaires et des systèmes de réécriture.

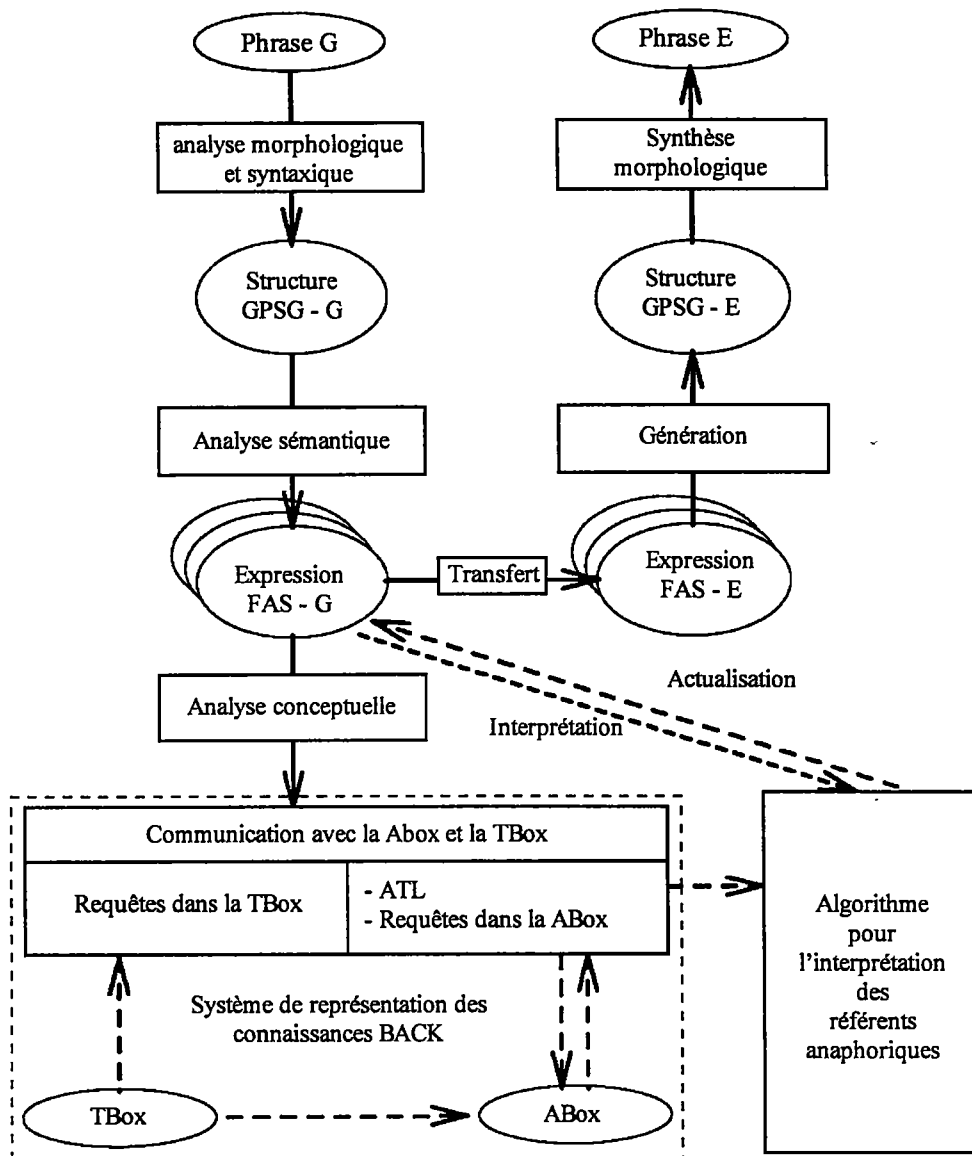


FIGURE 11 : ARCHITECTURE DU SYSTEME DE TA DANS LE PROJET KIT-FAST ¹

Les données linguistiques ont été développées pour traduire un texte allemand. Environ 1 000 phrases ont été testées avec succès grâce au système de TA.

Les données linguistiques comprennent :

- une grammaire allemande
 - * 22 catégories principales, 34 traits
 - * 22 alias
 - * 76 règles ID
 - * 23 états LP
 - * 5 métrarègles

¹ Le schéma proposé est extrait de W. WEISWEBER : *Installations- und Benutzungshandbuch des experimentellen MÜ-Systems des Projekts Kit-Fast*, Kit-Report 116, Berlin, 1994. Nous l'avons traduit pour plus de clarté. G = allemand, E = anglais.

- * 23 FCRs
- * 265 entrées lexicales (stemmes)
- 134 règles de réécriture pour l'analyse sémantique (allemand)
- 37 règles de réécriture pour l'analyse conceptuelle (allemand)
- 248 règles de réécriture pour le transfert (allemand→anglais)
- 182 règles de réécriture pour la génération (allemand)
- 8 facteurs pour l'évaluation des relations anaphoriques en allemand :
 1. accord
 2. liage
 4. proximité
 4. préférence pour l'objet sémantique
 5. préférence thématique
 6. identité des rôles
 7. préférence négative pour les adjoints libres
 8. consistance conceptuelle

4.3.4. Les projets d'implémentation utilisant HPSG

4.3.4.1. ALE : ATTRIBUTE LOGIC ENGINE

Contacts : Bob Carpenter

Type : Parser

Plate-forme : SICStus ou Quintus-Prolog ou autres compilateurs compatibles
SICStus 2.1 #8 ou plus pour pouvoir traiter les structures cycliques

ALE 2.0 est un système freeware écrit en Prolog et est développé à CMU par Bob Carpenter et Gerald Penn. Il intègre un parser syntaxique et une programmation logique des contraintes et utilise des structures de traits typées comme termes. On peut ajouter aux types des contraintes arbitraires, que l'on peut déclarer comme ayant des conditions d'identité structurelle extensionnelles. Alors que ALE a été développé pour traiter des grammaires HPSG, il peut aussi exécuter des grammaires PATR-II, des grammaires DCG, des programmes Prolog, Prolog-II, et des programmes LOGIN, etc. Avec un codage approprié, il peut exécuter quelques aspects des grammaires LFG.

Les termes utilisés dans les grammaires ALE et les programmes logiques sont spécifiés en utilisant une version typée de la logique attribut-valeur de Rounds-Kasper, qui inclut des variables, une disjonction complète, et des inéquations. Les programmes sont alors compilés en instructions Prolog du niveau inférieur qui correspondent aux opérations de base de la logique typée de Rounds-Kasper. On impose, au niveau des descriptions, une grande discipline sur les traits, ce qui permet de détecter un grand nombre d'erreurs au moment de la compilation.

Les systèmes de programmation logique et d'analyse peuvent être utilisés séparément ou simultanément. Les traits du système de programmation logique incluent la négation, la disjonction et les coupures.

Le système de structure syntaxique utilise un parser ascendant. Les composantes de la structure syntaxique et de la programmation logique du système autorisent la définition de macros paramétriques et leur libre emploi dans les descriptions.

4.3.4.2. ALEP : ADVANCED LINGUISTIC ENGINEERING PLATFORM

Contacts : Andrew Bredenkamp

Type : Grammaire, plate-forme de développement grammatical

Une grammaire HPSG pour l'anglais est actuellement en développement à l'Université d'Essex. La grammaire est implémentée en utilisant ALEP ; le système ne permet pas, jusqu'à présent, un héritage multiple⁰. Les règles lexicales HPSG sont implémentées comme des règles qui s'appliquent pendant le temps de travail, de nouvelles entrées lexicales sont alors créées au fur et à mesure si nécessaire.

ALEP est une plate-forme pour les développeurs de logiciels linguistiques. ALEP fournit un environnement multitâches, basé sur Motif comprenant des outils pour développer des grammaires et des dictionnaires. Le système s'appuie sur un formalisme basé sur l'unification, et sur une machine linguistique performante dotée de composantes d'analyse, de génération et de transfert. Une forme de morphologie à deux niveaux est conservée, tout comme l'est le texte d'entrée et de sortie écrit selon la norme SGML.¹ Les règles de grammaire sont compilées selon le codage de base de la machine sur laquelle tourne le système. Les règles lexicales sont compilées selon des clauses stockées sur le disque dans un système de gestion de base de données. Le système comporte également un logiciel linguistique de démonstration : une grammaire d'analyse pour l'allemand, une grammaire de synthèse pour l'anglais, une grammaire de transfert de l'allemand vers l'anglais et des dictionnaires allemand et anglais.

ALEP 1.0 tourne sur des stations SUN SPARK. ALEP 2.0 tourne également sur des IBM RS 600 et des Intel 486 sous Solaris.

ALEP est un système ouvert : les utilisateurs peuvent personnaliser le système et les développeurs peuvent y intégrer de nouveaux outils, y compris de nouveaux formalismes et machines linguistiques. Pour promouvoir sa fiabilité, ALEP a un langage de configuration basé sur la logique attribut-valeur. On peut programmer les extensions du système grâce à ce langage que l'on peut également utiliser pour construire une interface afin de séparer les applications ; ainsi le langage de l'utilisateur joue le même rôle en ALEP que LISP en EMACS.

Toutes les tentatives ont été menées pour se conformer aux normes et pour réutiliser les ressources existantes. Le logiciel EMACS a été utilisé de façon extensive.

ALEP a été implémenté par le Groupe de Développement des Applications chez IBM en Belgique. La partie graphique (FEGRAMED) est fournie par DFKI à Sarrebruck et par P-E International à Luxembourg. Le logiciel de démonstration ALEP est fourni par l'IAI à Sarrebruck. Le formalisme d'unification et la machine linguistique ont été conçus par SRI à Cambridge. La composante de traitement de texte a été conçue par Sema Group en Belgique. L'architecture du système est le résultat d'une étude menée par l'IAI à Sarrebruck, CAP Gemini SCS à Hambourg et SNI à Munich.

¹ SMGL (Standard Generalized Mark-up Language) : Norme de description de documents, notamment en vue de leur échange par exemple pour une impression.

4.3.4.3. BABEL

Contacts : Stefan Müller

Type : Grammaire, composante morphologique, parser

Plate-forme : Plateforme indépendante (sources Prolog en C)

Babel est une implémentation Prolog d'un fragment HPSG de l'allemand. La grammaire est écrite dans un format ID/LP et des constituants discontinus sont adoptés afin de décrire l'ordre des mots relativement libre en allemand. Un dictionnaire de stemmes est développé en temps différé pour produire un dictionnaire pleines formes qui compte, jusqu'à présent, 3 077 entrées. La grammaire s'accroît tous les jours. Stefan Müller utilise un parser graphique ascendant.

4.3.4.4. JPSG : JAPANESE PHRASE STRUCTURE GRAMMAR

Contacts : Hiroshi Tsuda, Koiti Hasida, Hidetosi Sirai

Type : Prolog/CLP/parser

Plate-forme : UNIX, Macintosh, MS-DOS

Hiroshi Tsuda a développé un langage de programmation logique contraint appelé cu-Prolog afin d'implémenter des formalismes grammaticaux basés sur les contraintes, en particulier JPSG (*Japanese Phrase Structure Grammar* ; une grammaire du style HPSG pour le japonais). cu-Prolog autorise les prédicats Prolog définis par l'utilisateur comme contraintes, et les PSTs (*partially-specified term*, termes partiellement spécifiés) pour décrire les structures de traits. cu-Prolog peut aussi être utilisé comme un interpréteur Prolog DEC10. Une version UNIX de cu-Prolog (avec des exemples des grammaires JPSG et HPSG) est enregistrée sous le nom ICOT Free Software (IFS). cu-Prolog a également été prévu pour MS-DOS et Macintosh.

4.3.4.5. HDRUG

Contacts : Gertjan van Noord

Type : Grammaire, parser

HDRUG est un environnement utilisé pour développer les grammaires logiques, les analyseurs, les générateurs pour les langues naturelles. Le progiciel est écrit en Sicstus Prolog et utilise ProTcl comme interface entre Sicstus Prolog et Tcl/Tk. Tcl/Tk est un langage script puissant pour développer des applications pour l'environnement X-windows. Le progiciel est fourni avec un certain nombre de grammaires en exemple, dont une grammaire catégorielle, une grammaire d'arbre adjoint, une grammaire d'unification dans l'esprit de HPSG, une grammaire d'extraposition et une grammaire à clause finie. Chaque grammaire est fournie avec un ensemble de parsers, tels que Earley, des parsers gauche-droite et des parsers guidés par la tête. Certaines grammaires sont livrées avec des variantes du générateur guidé par la tête.

HDRUG tient compte de la comparaison de différents parsers/générateurs ; il a des possibilités étendues pour compiler des équations de traits en terme Prolog ; il peut produire des sorties Prolog graphiques (Tk), et ordinaires d'arbres, des structures de traits, des termes Prolog (et leurs combinaisons), des graphes d'information statistiques, et des tableaux d'information statistiques, etc.

En utilisant uniquement les menus et les boutons, il est possible d'analyser des phrases, de les générer à partir de représentations de forme logique, de visualiser les arbres d'analyse produits par le parser ou le générateur, de changer une version particulière du parser au fur et à mesure, de comparer les résultats d'analyse d'une ou plusieurs même(s) phrase(s) par un ensemble de parsers différents, etc.

4.3.4.6. HPSG-PL

Contacts : Fred Popowich

Type : Système de développement grammatical

Plate-forme : Quintus and Sicstus Prolog (TreeTool tourne dans les environnements SunView et X-Windows)

HPSG-PL est une implémentation Prolog du formalisme HPSG, développé par Fred Popowich, Sandi Kodric et Carl Vogel (1991), Kodric, Popowich & Vogel (1992). Il a été développé pour être un outil de travail permettant d'élaborer et de tester des grammaires écrites dans le cadre de HPSG.

Le système se compose d'un compilateur lexical, d'un processeur contraint, d'un parser graphique et d'un module pour relier le parser à une interface graphique. En utilisant ce système, un utilisateur peut examiner les propriétés du formalisme HPSG, et peut étudier les caractéristiques de grammaires spécifiques qui utilisent ce formalisme.

Le système peut aussi être utilisé en conjonction avec l'interface graphique TreeTool développée par Baker et al (1990). TreeTool est un programme C qui utilise l'environnement Suntools pour afficher les arbres graphiques correspondant aux représentations du style Prolog des termes des arbres.

4.3.4.7. PLEUK

Contacts : Jo Calder

Type : shell ¹ de développement grammatical

Plate-forme : Sun SPARC SunOS {4,5}, HP-UX; SICStus Prolog 2.1#9 ou suivantes

Pleuk (Calder, 1993) est un shell de développement grammatical écrit par Jo Calder, Kevin Humphreys et Mike Reape. De nombreux formalismes grammaticaux différents peuvent être intégrés dans Pleuk, comme par exemple :

HPSG-PL : Un système de développement de grammaires HPSG produit à l'Université Simon Fraser (Canada) par Fred Popowich, Sandi Kodric et Carl Vogel.

Mike : Un simple système d'unification basé sur des graphes, amélioré par des opérations supplémentaires pour le traitement de l'ordre libre des mots proposé par Mike Reape dans diverses publications.

SLE : Un formalisme basé sur des graphes amélioré par des relations arbitraires à la façon de Johnson et Rosner (EACL, 1989) ainsi que Doerre et Eisele. Ce système a été utilisé pour le développement de plusieurs grammaires HPSG.

¹ ou interpréteur de commandes.

Term : Un système de grammaires d'unification basé sur des termes, développé à l'origine les grammaires d'unification catégorielle (Zeevat, Klein et Calder).

DCG : Grammaire à clause finie

Pleuk représente un unique «contrôleur de dérivation», un système graphique qui permet à l'utilisateur d'«augmenter les dérivations» grâce à des actions incluant la sélection de matériel lexical (ou autre) et son insertion dans des structures plus grandes. Ce qui offre la possibilité d'utiliser des parsers et des générateurs pour étudier les conséquences des descriptions grammaticales.

4.3.4.8. SOAS HPSG ELLIPSIS

Contacts : Shalom Lappin, Howard Gregory.

Le projet HPSG Ellipsis a débuté en septembre 1995 à SOAS (Université de Londres), il est placé sous la direction de Shalom Lappin (Département de Linguistique). En juin 1996, Howard Gregory a rejoint le projet, succédant à Hsue-Hueh Shih.

Le projet a pour but d'implémenter l'algorithme de résolution des ellipses proposé par Lappin et Shih en utilisant une grammaire de traits typés basée sur HPSG et écrite en ProFIT (Prolog avec Features (traits), Inheritance (héritage) et Templates (formes-types)).

4.3.4.9. TDL : TYPE DESCRIPTION LANGUAGE

Contacts : Hans-Ulrich Krieger

Type : Langage de description de types et système d'inférence

Plate-forme : SUN Sparc, HP 9 000, IBM RS 6 000, Apple Macintosh, LINUX PCs, Franz Allegro CL, Macintosh CL, Lucid CL, CLISP.

TDL est un langage typé de représentation basé sur les traits et un système d'inférence, spécialement conçu pour soutenir des théories grammaticales très lexicalisées telles que HPSG. La définition de types en TDL se compose de contraintes de types et de traits sur les connecteurs booléens. Le travail avec des types partiellement ou totalement étendus est possible. Le contrôle des connaissances est spécifié à un niveau séparé. Un raisonnement efficace en TDL est réalisé grâce à plusieurs modules spécialisés. TDL est une partie d'un système plus grand qui fournit des composantes supplémentaires : un parser, un composant d'apprentissage basé sur des explications, un éditeur de traits, une importante grammaire HPSG pour l'allemand (approx. 1 500 définitions de types), etc.

4.3.4.10. TFS : TYPED FEATURE STRUCTURES SYSTEM 6.1

Contacts : Martin Emele

Type : Formalisme de représentation basé sur les contraintes

Plate-forme : Common Lisp

Le formalisme de représentation de structures de traits typées (*TFS, Typed Feature Structure*) est élaboré pour essayer de fournir une synthèse de quelques uns des concepts

clés des formalismes grammaticaux basés sur l'unification (structures de traits), des langages de représentation des connaissances (héritage) de la programmation logique (variables logiques et déclarativité). L'architecture de contraintes basée sur l'héritage incluse dans le système TFS intègre deux paradigmes computationnels : l'approche par objets (*object-oriented*) propose des objets complexes, récursifs, éventuellement reliés en réseaux représentés comme des structures de traits typées avec des restrictions attribut-valeur, des contraintes d'(in)égalité, et un héritage multiple ; l'approche *relational programming* propose la déclarativité, des variables logiques, le non-déterminisme avec un « backtracking », et l'évaluation existentielle des questions. L'interpréteur du formalisme est décrit comme un système de réécriture des termes où l'unification de structures de traits typées est utilisée pour détecter les contradictions entre une question et les contraintes imposées par le système de types de traits.

Les grammaires d'unification sont organisées comme des réseaux d'héritage des structures de traits typées. Les structures linguistiques complexes sont décrites au moyen de contraintes de type récursives qui correspondent aux définitions de classes dans les formalismes orientés vers l'objet. L'utilisation d'une méthodologie orientée vers l'objet avec héritage est très intéressante pour le traitement du langage naturel et offre un certain nombre d'avantages tels que l'abstraction et la généralisation, le partage de l'information, la modularité et la réutilisabilité des descriptions. A travers le développement d'un grand nombre d'applications différentes, il a été démontré que le formalisme est assez flexible et adéquat pour représenter l'approche des principes et des paramètres des théories linguistiques computationnelles modernes. Il a été utilisé avec succès pour encoder des fragments d'une grammaire HPSG et un système de traduction automatique bilingue basé sur l'approche des contenus partagés.

4.3.4.11. UBS : UNIFIKATIONSBASIERTE SPRACHE

Auteur : Frieder Stolzenburg

Tâche : Analyse linguistique, test de théorie linguistique (HPSG, Pollard & Sag, 1987)

Le formalisme HPSG est complexe et comprend plus de structures de données et d'opérations que les formalismes grammaticaux précédents. Aux structures de traits régulières s'ajoutent la disjonction, la négation, les ensembles, les fonctions générales et les types. Ces nouveaux aspects ne peuvent pas être directement implémentés en Prolog. UBS a donc été développée pour traiter ces aspects. Le système suit la tradition de GULP (Covington, 1989), une extension de Prolog qui facilite la spécification des structures de traits dans les programmes Prolog en autorisant l'indépendance de l'ordre et de l'arité. GULP a malgré tout des limites puisqu'il ne peut pas traiter la négation, la disjonction, les ensembles ou les fonctions comme des valeurs de traits. UBS est maintenant capable de traiter les structures de traits typées, l'unification, la disjonction, la négation (restreinte), les valeurs dépendantes fonctionnelles (relations) et les ensembles (unification d'ensembles).

4.3.4.12. PROJET DE GRAMMAIRE SYNTAGMATIQUE DU FRANÇAIS

Responsable : D. Godard (CNRS, Paris 7) et I. Sag (Stanford)

Collaboration NSF-CNRS : URA 1 028, TALANA et Stanford

Participants : A. Abeillé, F. Namer, E. Guimier, M.-P. Jouannaud, B. Robichaud,
G. Recourcé, P. Miller

Durée : octobre 1994 à octobre 1997

Il s'agit d'écrire une grammaire syntagmatique du français basée sur les structures de traits, sur le modèle de la grammaire HPSG de C. Pollard et I. Sag pour l'anglais. Il s'agit de reprendre certaines des analyses proposées en grammaire générative, en les complétant, et de s'attaquer à des phénomènes peu décrits (comme la coordination). Les travaux déjà réalisés concernent les auxiliaires de temps, le passif, les verbes de perception, les constructions causatives, la négation et le placement des adverbes. Un travail linguistique contrastif français-anglais est également réalisé.

4.3.4.13. AUTRES TRAVAUX

Philippe Blache travaille sur :

- * l'interprétation des HPSG en programmation par contraintes : interprétation directe de HPSG reposant sur l'utilisation de contraintes actives, développement d'un parser HPSG du français en LIFE
- * l'intégration Prosodie/Syntaxe en HPSG : développement d'un parser syntaxique HPSG guidé par la prosodie. Une application de ce système est la désambiguïsation par l'analyse de l'intonation.

4.3.5. Les projets d'implémentation utilisant TAG

4.3.5.1. LE SYSTEME XTAG

Auteur : Doran et al.

4.3.5.1.1. Présentation générale du système

La phrase d'entrée est soumise à l'analyseur morphologique et au système de balisage. Cet analyseur extrait l'information morphologique pour chaque mot individuel de la phrase à partir d'une base de données morphologique. La phrase augmentée, dont chaque mot est repéré par des étiquettes de parties du discours et des informations morphologiques, constitue l'entrée de l'analyseur qui consulte alors la base de données syntaxique et la base de données d'arbres pour ressortir les structures d'arbres adéquates pour chaque mot de la phrase. L'analyseur compose alors les structures pour obtenir l'analyse de la phrase.

4.3.5.1.2. Analyseur morphologique

Les données morphologiques sont extraites, au départ, du *Collins English Dictionary* ([Hanks, 1979]) et du *Oxford Advanced Learner's Dictionary* ([Hornby, 1974]), puis nettoyées et complétées à la main ([Karp et al., 1992]). La base de données se compose d'environ 317 000 items fléchis, avec leurs racines et les informations flexionnelles (telles que cas, nombre, temps). On distingue treize parties du discours : Nom, Nom propre, Pronom, Verbe, Particule verbale, Adverbe, Adjectif, Préposition, Complémenteur, Déterminant, Conjonction, Interjection, et Contraction nominale/verbale. Les noms et les verbes représentent les catégories les plus importantes avec approximativement 213 000 et 46 500 formes fléchies.

4.3.5.1.3. Système de balisage des parties du discours

Un système de balisage des parties du discours ([Church, 1988]), établi à partir du corpus du *Wall Street Journal*, est intégré à XTAG. XTAG utilise ces informations pour réduire le nombre d'analyses spécifiques en filtrant les parties du discours possibles fournies par l'analyseur morphologique pour chaque mot.

4.3.5.1.4. Parser

XTAG utilise un parser du type Earley qui a été étendu pour traiter les structures de traits associées aux arbres ([Schabes, 1990]).

Le parser utilise l'entrée de la base de données syntaxique pour chaque item lexical de la phrase pour sélectionner un ensemble de structures élémentaires à partir de la base de données d'arbres.

L'arbre d'analyse représente la structure de constituants de surface, alors que l'arbre de dérivation représente le déroulement dérivationnel de l'analyse. L'opération de composition est indiquée par la nature des arcs : un trait pointillé pour la substitution, un trait plein pour l'adjonction. Le nombre situé à côté du nom de chaque arbre correspond à l'adresse du nœud où a lieu l'opération. L'arbre de dérivation peut aussi être interprété comme un graphe de dépendance avec des arcs non étiquetés entre les mots de la phrase.

Des méthodes complémentaires, qui tirent avantage des FB-LTAG, ont été implémentées pour prouver la performance du parser. Par exemple, la taille de l'arbre et la position de l'ancre dans celui-ci sont utilisées pour éliminer les arbres incorrects lors du premier passage du parser. Des informations statistiques sur les fréquences d'utilisation des arbres ont été obtenues par l'analyse de corpus. Ces informations ont été compilées dans une base de données statistique qui est utilisée par le parser.

4.3.5.1.5. Base de données syntaxique

La base de données syntaxique associe des items lexicaux à des arbres et à des familles d'arbres adéquats basés sur diverses informations de sélection. Les entrées (plus de 37 000) de cette base de données ont été extraites, à l'origine, du *Oxford Advanced Learner's Dictionary* ([Hornby, 1974]) et du *Oxford Dictionary for*

Contemporary Idiomatic English ([Cowie et Mackin, 1975]) puis modifiées et complétées à la main ([Egedi et Martin, 1994]).

Chaque entrée syntaxique constitue un champ INDEX, la forme infléchié sous laquelle l'entrée est compilée dans la base de données un champ ENTRY, qui contient tous les items lexicaux qui va ancrer le(s) arbre(s) associé(s), un champ POS, qui indique la partie du discours pour le ou les items lexical (aux) dans le champ ENTRY, et ensuite soit un champ TREES, soit un champ FAM mais pas les deux en même temps. Le champ TREES donne une liste d'arbres individuels à associer à l'entrée, alors que le champ FAM donne une liste de familles d'arbres. Une entrée syntaxique peut aussi contenir une liste de formes-types de traits (FS) qui se développe en équations de traits qui doivent être placées dans le ou les arbre(s) spécifié(s).

La base de données syntaxique subit actuellement quelques modifications destinées à faciliter son utilisation et sa mise à jour. De plus, le nombre d'entrées sera augmenté afin d'étendre sa couverture, et les défauts utilisés par le système XTAG sera accessible à partir de la base de données elle-même.

4.3.5.1.6. Assignment de cas

La théorie GB (*Government and Binding theory*, théorie du gouvernement et du liage) propose le «filtre casuel» suivant comme exigence sur la structure s.

FILTRE CASUEL

Un cas abstrait doit être assigné à chaque GN de surface. ([Haegeman, 1991])

Le cas abstrait est considéré comme universel. On suppose que les langues qui ont un marquage casuel morphologique riche, tel que le latin, et les langues qui ont un marquage casuel morphologique très limité, comme l'anglais, possèdent toutes des systèmes complets de cas abstrait qui se différencient uniquement par l'étendue de la réalisation morphologique.

En GB, le cas abstrait est assigné aux GNs par divers « case assigners »¹, à savoir les verbes, les prépositions, et INFL. On dit que les verbes et les prépositions assignent le cas accusatif aux GNs qu'ils gouvernent, et INFL assigne le nominatif aux GNs qu'il gouverne. La notion de cas abstrait et de filtre casuel sont utiles pour rendre compte d'un certain nombre de phénomènes comprenant la distribution du nominatif et de l'accusatif et la distribution de GNs de surface et de catégories vides (telle que PRO).

MINIMALISME ET CAS

Une différence conceptuelle importante entre la théorie GB et le Minimalisme est que dans le Minimalisme, les items lexicaux transportent leurs traits avec eux plutôt que de se voir assigner les traits basés sur les nœuds sur lesquels ils se trouvent. Pour les noms, cela signifie qu'ils transportent les cas avec eux, et que le cas est contrôlé quand ils sont en position SPEC de AGR ou en position AGR, qui, par conséquence, disparaît ([Chomsky, 1992]).

LES CAS EN XTAG

¹ Nous conservons le terme anglais car nous n'avons pas d'équivalent avec un nom en français, il faut utiliser une périphrase du type « les éléments qui assignent les cas ».

La grammaire anglaise XTAG adopte la notion de cas et de filtre casuel pour de nombreuses raisons identiques à celles formulées dans GB. Toutefois, à certains égards, l'implémentation des cas de la grammaire anglaise XTAG ressemble plus au traitement dans le cadre de Minimalisme de Chomsky ([Chomsky, 1992]) que le système de la littérature GB ([Chomsky, 1986]). Comme dans le Minimalisme, les noms dans la grammaire XTAG transportent avec eux le cas qui est éventuellement «contrôlé». Toutefois, dans la grammaire XTAG, les cas des noms sont contrôlés par rapport aux valeurs des cas assignées par le verbe au cours de l'unification des structures de traits. Contrairement au Minimalisme de Chomsky, il n'y a pas de nœud AGR séparé ; le contrôle du cas vient directement du verbe. L'assignation du cas par le verbe ressemble plus à l'approche de GB qu'aux spécifications de la relation de tête SPEC dans le Minimalisme.

La plupart des noms anglais n'ont pas de formes séparées pour le nominatif et l'accusatif et il peut y avoir une ambiguïté entre les deux. Les pronoms, évidemment, sont morphologiquement marqués pour le cas, et chacun d'entre eux transporte le cas approprié dans ses traits.

Dans la grammaire anglaise XTAG le cas est assigné par deux composants : les verbes et les prépositions. Les prépositions assignent l'accusatif (acc) grâce à leur trait *assign-case*, qui est directement lié au trait casuel de leurs objets.

Dans la grammaire XTAG, les verbes représentent l'autre partie du discours qui peut assigner un cas. Comme XTAG ne distingue pas les nœuds INFL et les nœuds GV, les verbes doivent fournir une assignation de cas sur la position sujet en plus du cas assigné à leurs GNs compléments.

L'assignation de cas aux GNs compléments est traitée par l'assignation de valeurs casuelles des compléments directement dans l'arbre que le « case assigner » (le verbe) ancre. Il faut noter que les assignations de cas pour les GN compléments se trouvent déjà dans les arbres, même s'il n'y a pas encore d'item lexical qui ancre l'arbre. Puisque chaque verbe qui sélectionne ces arbres (et d'autres arbres dans chaque cadre de sous-catégorisation) assigne le même cas aux compléments, l'élaboration de traits casuels dans l'arbre a exactement le même résultat que la mise en valeur du trait casuel dans chaque entrée lexicale du verbe.

Le cas assigné à la position sujet varie en fonction de la forme verbale. Comme la grammaire XTAG traite le verbe infléchi comme une seule unité plutôt que le diviser en INFL et nœuds V, le cas, tout comme le temps et l'accord, est indiqué dans les traits des verbes, et doit être transmis de façon appropriée. La forme morphologique de verbe détermine la valeur du trait d'assignation de cas.

La plupart des formes d'un verbe assignent un cas nominatif, bien que certaines formes, tel que le participe passé, n'assignent aucun cas quel qu'il soit. La distinction entre un cas zéro et l'absence de cas est une indication de la divergence de la théorie standard GB.

Dans la théorie GB, l'absence de cas sur un GN signifie que seul PRO peut occuper la place du GN. Avec l'unification des traits telle qu'elle est utilisée dans la grammaire

FB-LTAG, l'absence de cas sur un GN signifie que n'importe quel GN peut occuper cette place, quelque soit son cas.

Ceci est dû au mécanisme de l'unification, dans lequel, si quelque chose n'est pas spécifié, il peut s'unifier avec n'importe quoi. Ainsi on a un cas spécifique zéro pour traiter les formes verbales qui, dans la théorie GB, n'assignent pas de cas. PRO est le seul GN avec un cas zéro. Les formes verbales qui n'assignent pas de cas, comme le participe passé mentionné ci-dessus, peuvent le faire parce qu'elles ne peuvent pas apparaître sans un auxiliaire qui tient compte de l'assignation de cas.

4.3.5.1.7. Les prépositions

Il existe trois types de groupes prépositionnels, et trois places auxquelles ils peuvent s'adjoindre. Les trois types de groupes prépositionnels sont : les prépositions avec GN complément, les prépositions avec complément propositionnel, et les prépositions exhaustives. Les trois places se situent à droite d'un GN, à droite d'un GV, et à gauche d'une P. Chacun des trois types de GP peut s'adjoindre à ces trois places, pour un total de neuf arbres.

Les prépositions qui prennent des GN compléments leur assignent l'accusatif. La plupart des prépositions prennent des GN compléments. Il n'y a que peu de prépositions qui prennent des compléments propositionnels.

4.3.6. Les autres projets d'implémentation utilisant les grammaires d'unification

En règle générale, les projets présentés ci-dessous utilisent plusieurs modèles de grammaire.

4.3.6.1. FORMALISMES GRAMMATICaux ET ANALYSE DU FRANÇAIS

Responsable : Anne Abeillé

Participants : A. Husson, B. Robichaud, B. Daille, F. Namer, G. Recourcé, P. Samvelian

Les membres de ce pôle de recherche, d'une part, mettent au point des spécifications linguistiques écrites dans les formalismes TAG, LFG et HPSG, d'autre part, implémentent des maquettes d'analyse du français. Leur objectif est de réaliser un parser syntactico-sémantique dont la sortie soit suffisamment abstraite pour être utilisée dans des applications diverses, notamment la traduction automatique et l'interrogation de bases de données. Il suppose la définition, à base de corpus représentatifs, d'un cahier des charges pour une grammaire du français « réutilisable » et couvrant le « noyau dur » des phénomènes syntaxiques.

L'équipe TALANA dispose déjà des réalisations suivantes :

- * Une maquette réalisée dans le cadre du projet européen Eurotra, pour lequel l'équipe TALANA était responsable de l'analyse et de la génération du français, ainsi que des modules de transfert vers le français. S'appuyant sur une approche statificationnelle - avec à chaque niveau des règles de réécriture

augmentées de structures de traits complexes combinées par unification - cette maquette couvre de nombreux phénomènes syntaxiques pour lesquels une représentation sémantique a été également définie, notamment les constructions comparatives et les phrases à verbe support.

- * Une maquette réalisée à l'aide du logiciel XTAG pour les grammaires d'arbres adjoints qui dispose d'une large couverture syntaxique, avec en particulier un traitement efficace et élégant des dépendances à distance et des expressions semi-figées grâce au domaine de localité étendu des arbres élémentaires lexicalisés et à la puissance de l'opération contextuelle qu'est l'adjonction. Ce parser a été implémenté pour les 3 000 lemmes les plus fréquents en français. Il est en cours de test sur un corpus de textes courts, pour enfants (d'environ 100 K), dans le cadre d'une collaboration AUPELF-UREF : Projet Linguistique théorique et traduction. Une évaluation sur le corpus de phrases réalisées par l'ISSCO dans le cadre du projet européen Test-suite NLP est également prévue.
- * Une maquette réalisée à l'aide du logiciel européen ALEP, qui permet l'implémentation (en Prolog) de grammaires à base de règles de réécriture hors contexte et de structures de traits typées inspirées de HPSG. Il s'agit d'implémenter les analyses linguistiques proposées dans le cadre du modèle HPSG (auxiliaires, clitiques, causatives, relatives, négation, noms prédicatifs et groupes nominaux complexes).

Ce travail suppose d'une part la mise au point de stratégies « robustes » pour proposer des analyses partielles en cas de phrases « déviantes » (ou comportant des non-attendus), et d'autre part le développement d'outils d'aide à l'extension des lexiques et des grammaires pour les futurs utilisateurs extérieurs. Il doit également déboucher sur la mise en place de critères d'évaluation plus généraux (par ex. quand peut-on dire qu'une phrase d'un texte a été « bien » analysée par un programme ?).

4.3.6.2. PROJET DE RECHERCHE : ENVIRONNEMENT DE TEST GRAMMATICAL - GTU (GRAMMATIK-TEST-UMGEBUNG)

Auteurs : M. Volk, H. Ridder, J. Hubrich, D. Richarz, M. Jung, A. Wagner, A. Fitschen, S. Pieper, M. Wirt, B. Kullbach

GTU est disponible sous UNIX et DOS. C'est un outil simple à utiliser pour l'enseignement de l'analyse syntaxique. GTU décrit un environnement de développement et de test pour les grammaires dans le cadre du paradigme d'unification.

GTU se compose d'un analyseur morphologique, d'un parser, d'éditeurs grammaticaux, d'une série de tests de 350 phrases, d'un module de « debuggage », de contrôleurs grammaticaux statiques, d'un module d'aide hypertexte, d'un module de sortie pour une présentation des résultats d'analyse conviviale pour l'utilisateur. Les trois derniers composants ne sont disponibles que dans la version UNIX.

4.3.6.2.1. Évolutions de GTU - Extension pour les grammaires catégorielles, en particulier UCG

En tant qu'instrument de recherche, KoGiTum est très intéressant dans le fait que différentes grammaires peuvent être développées dans un cadre similaire. La prise en compte des grammaires catégorielles est importante car ce type de grammaires sera linguistiquement significatif à l'avenir.

L'implémentation doit contenir :

- * un parser pour les grammaires UCG,
- * une connexion aux composants du lexique et une gestion de l'interface,
- * un ajustement des résultats aux formats d'édition KoGiTum,
- * des jeux d'essais qui comprennent une quantité représentative de phrases-tests.

La réalisation de ce projet s'effectue en collaboration avec l'*Institute for Information Technology*, institut de recherche de Daimler-Benz à Ulm (Groupe de projet Mecklenburg/Ehrlich).

4.3.6.2.2. Développement d'un parser GTU-HPSG pour KoGiTum

Un parser HPSG efficace a besoin d'algorithmes pour, au moins, les problèmes suivants :

- * Utilisation des ébauches existantes : Emele, Zajac, Stolzenburg
- * Données typées et hiérarchies d'héritage ; défauts
- * Héritage multiple dans le lexique
- * Partage des traits dans la partie grammaticale ; unification avec disjonction et négation
- * Interfaces avec le lexique pour les structures complexes
- * Intégration d'informations sémantiques : sémantique situationnelle
- * Intégration d'informations sur le discours : DRS
- * Complément des formats d'édition déjà existants.

4.3.6.2.3. Développement d'une interprétation sémantique GTU dans les grammaires d'unification

L'interprétation sémantique, dans le concept orienté vers la syntaxe de KoGiTum, est nouvelle, mais correspond à l'aspiration générale d'élaborer des systèmes de traitement linguistique intégrés, comme cela se manifeste notamment dans la philosophie HPSG. Comme ce travail est récent, on peut seulement établir des postulats généraux.

L'interprétation sémantique ne sera pas développée isolément mais sera rattachée à un modèle syntaxique spécifique.

Ceci nécessite une référence exacte aux résultats du travail syntaxique et permet l'utilisation du parser syntaxique et des archives des phrases-test.

Comme cela est d'usage dans le développement syntaxique, on avance en sémantique et on développe des solutions partielles : négation, opérations avec des individus et des groupes (Wagner), présuppositions (Schauer). Chacune de ces «solutions partielles» est, à elle-même, une procédure étendue.

Pour l'interprétation sémantique il sera nécessaire de définir un nouveau niveau de description, (comme peut-être les structures *s* à côté des structures *c* et *f* en LFG).

Comme les descriptions partielles de la sémantique se déroulent dans des « micro-mondes », il se pose le problème de l'intégration des résultats dans un modèle global dont le développement risque d'être plus difficile que le déroulement incrémentiel en syntaxe. L'intégration des descriptions sémantiques partielles est donc un travail particulier.

4.3.6.2.4. Perfectionnement de GTU

En ce qui concerne KoGiTum, le travail d'enrichissement le plus urgent est de compléter le parser ID/LP afin d'obtenir un parser GPSG sans défaut. On peut utiliser des solutions partielles du modèle ID/LP existant, mais il faut également programmer des composants manquants, et notamment :

- * le traitement des données avec HFCs, FFCs, FCRs, et FSDs ¹
- * les traits SLASH et AGR
- * les métarègles : spécifications de formats et contraintes
- * les possibilités d'affichage des arbres partiels (résultats provisoires)

4.3.6.3. LKP : LINGUISTIC KERNEL PROCESSOR

Auteurs : H.-U. Block, M. Gehrke, R. Hunze, S. Schachtl, L.-A. Schmid, C. Zünkler
Tâche : Parsing et génération de l'allemand

Le système utilise la grammaire de trace et d'unification (*TUG, Trace and Unification Grammar*) et autorise la disjonction non contrainte d'équations de traits. TUG est un formalisme grammatical basé sur l'unification qui comprend des types de règles complémentaires pour les règles de mouvement. Il se compose d'un parser, d'un générateur et d'outils lexicaux.

4.3.6.4. GULP : GRAPH UNIFICATION LOGIC PROGRAMMING

Auteur : M.-A. Covington
Tâche : Outil d'implémentation de grammaires

Il résout un problème maudit avec Prolog, à savoir, le manque d'un moyen efficace pour représenter les structures de traits dans lesquelles les traits sont identifiés par leur nom plutôt que par leur position.

Il se compose d'une simple extension de Prolog fournissant une représentation adéquate des structures de traits (structures attribut-valeur), qui peut être utilisée par un parser Prolog descendant ou par n'importe quel parser écrit en Prolog.

¹ pour les abréviations voir index en fin de thèse.

4.3.6.5. CUF

Auteurs : M. Dorna, J. Dörre

Tâche : CUF fournit un formalisme pour une description déclarative de phénomènes linguistiques indépendants du domaine linguistique et un système pour le traitement des descriptions CUF.

Description : description déclarative de phénomènes linguistiques (syntaxe, phonologie, morphologie et sémantique) dans le formalisme CUF et traitement basé sur les contraintes de connaissances linguistiques dans le système CUF.

4.3.6.6. AUTRES TRAVAUX

Philippe Blache travaille sur le rôle et l'utilisation des théories linguistiques pour la traduction automatique. Les théories observées sont GB, GPSG, HPSG et TAG. L'objectif de ce projet est de montrer l'intérêt des théories linguistiques pour la traduction automatique. Ses partenaires sont J.-Y. Morin de l'Université de Montréal, E. Wehrli de l'Université de Genève et A. Abeillé du groupe TALANA de l'Université de Paris VII.

4.4. EUROTRA

Lancé par la Communauté Économique Européenne en novembre 1982, il s'agit du projet de Traduction Automatique le plus ambitieux et le plus important quant au personnel et au potentiel d'extension. L'objectif de ce programme de recherche et de développement est de construire le prototype préindustriel d'un système de T.A. de conception avancée, qui couvre les 9 langues (7 initialement) de la Communauté : allemand, anglais, danois, espagnol, français, grec, italien, néerlandais et portugais.

Des équipes universitaires des pays membres (242 personnes, dont 80 % de linguistes, traducteurs, informaticiens et mathématiciens, réparties sur 21 sites) travaillent pour leur langue respective, sur les modules d'analyse et de synthèse. Elles ont à charge également la réalisation du module de transfert vers leur langue. Le projet se déroule en trois phases. Après la détermination du modèle théorique et la conception du formalisme (1982-1984), les études linguistiques (1985-1988) ont préparé l'élaboration d'une maquette de prototype qui devrait clore la phase 3 (1988-1990) et fonctionner avec un vocabulaire de 20 000 mots (dont 15 000 termes techniques du domaine des télécommunications).

Il s'agit d'un système multilingue qui traite les langues de la Communauté dans leur ensemble, avec une même approche linguistique. On utilise un modèle unique pour passer d'une langue à une des huit autres et on s'appuie sur une même description pour toutes les étapes de la traduction, que la langue soit source ou cible.

L'hypothèse de base est que le processus de traduction est une succession de plusieurs transformations au cours desquelles un texte doit être retranscrit, de sa forme originale (texte source) jusqu'à la sortie correspondante (texte cible) sous des représentations grammaticales successives correspondant à des impératifs linguistiques (niveaux de la morphologie, de la syntaxe de surface, de la syntaxe profonde, de la

sémantique). Chaque niveau de représentation est défini par une grammaire (« générateur ») composée d'un ensemble de règles (« constructeurs »). L'application de ces règles produit des arbres décorés. Les objets d'un niveau de représentation sont transcrits dans le niveau suivant par un autre ensemble de règles (« traducteurs »).

Les niveaux dans EUROTRA :

- * niveau ECS (EUROTRA Configurational Structure) = niveau de la structure des constituants avec l'ordre des mots d'origine,
- * niveau ERS (EUROTRA Relational Structure) = niveau de la structure des valences,
- * niveau IS (Interface Structure) = niveau de la structure prédicat-argument.

La syntaxe à plusieurs niveaux du système EUROTRA permet de prendre en considération plusieurs points de vue pertinents pour le traitement des termes polysémiques.

En résumé, nous dirons que quel que soit le niveau de représentation, on utilise un même formalisme pour les grammaires, les dictionnaires et les « transferts », un formalisme déclaratif basé sur les grammaires d'unification. On soulignera que les étapes d'analyse et de génération sont poussées au maximum pour réduire le transfert au minimum, c'est-à-dire à un transfert lexical. L'interface obtenue à la fin de l'étape d'analyse conserve la structure sémantique et les indications lexicales, elle laisse les informations morphosyntaxiques devenues inutiles. La phase de génération s'appuie sur cette interface pour chercher les équivalents dans la langue cible, sans recours donc à la morphosyntaxe de la langue source. Il est malheureusement des situations dans lesquelles on ne peut pas être suffisamment abstrait au niveau de l'interface et où il faut alors recourir à des « stratégies de compensation ».

5. IMPLEMENTATION

5.1. PRESENTATION DE CAT2

CAT2 est un système de traduction automatique (TA) multilingue basé sur l'unification et le transfert. Il comprend un formalisme bien défini pour exprimer les connaissances linguistiques, et un ensemble d'outils pour les développer et les tester. A l'origine, il a été développé pour la TA, mais il peut tout aussi bien être utilisé pour étudier les aspects de l'analyse ou de la génération du langage. Il a été développé et est amélioré par l'IAI (*Institut der Gesellschaft zur Förderung der Angewandten Informationsforschung e.V. an der Universität des Saarlandes*).

Le cadre de CAT2 est l'héritier direct du cadre de $\langle C,A \rangle, T$, utilisé autrefois dans le projet de TA Eurotra. Le cadre $\langle C,A \rangle, T$ (Arnolds et al., 1985, 1986 ; Arnolds & des Tombe, 1987) a introduit les concepts de constructeurs, d'atomes et de translateurs : les constructeurs décrivaient partiellement des arbres, les atomes des feuilles et les translateurs les relations entre les arbres. Les constructeurs et les atomes qui constituaient un niveau de représentation étaient appelés un générateur ; ainsi les translateurs étaient en fait des relations entre les générateurs.

Cependant, la structure initiale avait un sérieux défaut : les règles constituant un translateur reliaient en fait des représentations avec des descriptions, c'est-à-dire qu'une moitié de la règle de translation décrivait partiellement un arbre et l'autre partie désignait une règle dans le générateur-cible. Ainsi, il fallait une règle de translation séparée pour chaque entrée lexicale étant donné que chaque entrée était nominative.

Le système CAT2 (Sharp, 1988) a changé cela grâce à des translateurs qui relient des représentations avec d'autres représentations, assouplissant ainsi l'étroite connexion entre translateurs et générateurs. D'autres changements sont intervenus pour simplifier la structure originale et pour rendre les différents aspects du formalisme plus précis et plus homogènes. L'efficacité de l'implémentation a également été améliorée.

La version CAT2 2.0 marque un changement puisque l'on passe d'un formalisme basé sur l'unification à un formalisme basé sur des contraintes. La syntaxe est apparue, inspirée des récents travaux sur la logique des traits menés par Johnson (1988) et Kasper (1987a, 1987b). Il est possible maintenant d'exprimer de véritables négations logiques, alors qu'auparavant, l'implémentation s'apparentait à un simple test d'instanciation de traits caractéristiques qui dépendait grandement du temps.

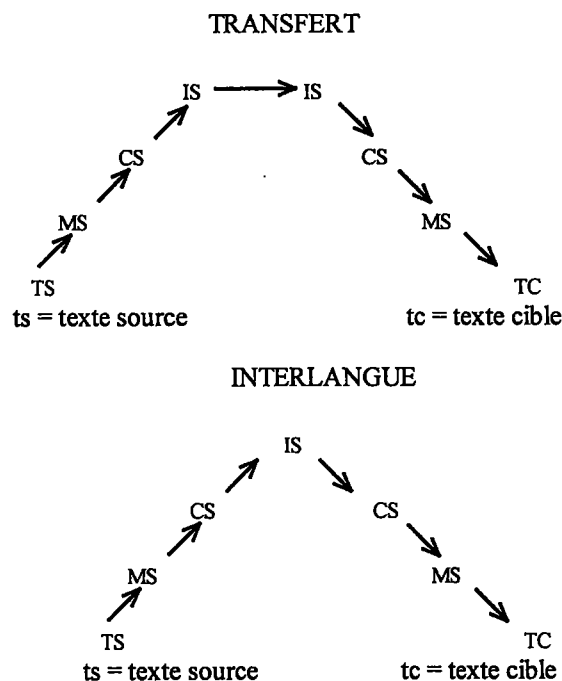
La disjonction de traits a également subi un changement en passant d'une évaluation de la forme disjonctive normale à une évaluation « paresseuse » dans laquelle la disjonction n'est évaluée que lorsque c'est nécessaire, c'est-à-dire quand des informations supplémentaires sont disponibles.

Qu'est-ce qui distingue CAT2 des autres formalismes en langue naturelle ? L'une des principales caractéristiques de CAT2 réside dans le fait que les constructions linguistiques en CAT2 sont des arbres plutôt que de simples structures de traits comme pour ZFK (Kaplan & Bresnan 1982), FUG (Kay 1984) et HPSG (Pollard & Sag 1987). Les arbres sont ordonnés par inhérence selon des structures hiérarchiques alors que les

structures de traits ne le sont pas. Cela rend CAT2 plus approprié pour modéliser des théories linguistiques basées sur une structure de phrase, comme par exemple GM (Chomsky 1986) ou GPSG (Gazdar et al. 1985). Cependant, les traits jouent également un rôle important dans ces théories ; ils forment le contenu des nœuds dans les arbres, et c'est de cette façon qu'ils sont exprimés dans le formalisme CAT2.

L'importance du rôle des niveaux de représentation est une autre caractéristique de la structure de CAT2. Il est reconnu en TA que la structure de constituants d'une langue n'est pas une représentation appropriée pour générer sa traduction dans une autre langue. Pour cette raison, elle est transformée en une forme plus adaptée. Cette forme appelée structure d'interface (IS, *interface structure*) est une structure arborescente normalisée dans laquelle les idiosyncrasies spécifiques au langage sont éliminées.

Dans le meilleur des cas, cette IS est totalement indépendante du langage ; c'est la base du paradigme interlingue de TA. Le paradigme basé sur le transfert dans lequel l'IS d'une langue est liée à l'IS d'une autre langue via un système de règles de transfert est moins ambitieux. En CAT2 chaque niveau de représentation est défini par un générateur et les relations entre chaque niveau sont définies par un translateur. Ceci distingue le formalisme CAT2 de LFG par exemple, dans lequel les structures *c* et les structures *s* sont des formes de représentation totalement différentes et l'équivalent des translateurs est un système d'équation fonctionnelle qui fait office d'annotations aux règles de structure de phrase.



5.1.1. Niveaux de représentation

Une grammaire est un ensemble d'assertions formelles qui définit l'ensemble des séquences bien formées d'une langue et assigne une représentation structurale à cette séquence. Une grammaire CAT2 est constituée d'un ensemble de règles regroupées en générateurs et translateurs. Un générateur est un ensemble de règles qui définit les structures bien formées à un certain niveau de représentation, et un translateur un

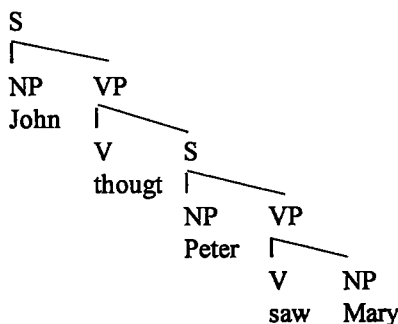
ensemble de règles qui apparie les structures présentes à un niveau de représentation à celles d'un niveau adjacent.

CAT2, comme beaucoup d'approches et de théories linguistiques, scinde le monde linguistique en niveaux de représentation. Le premier niveau s'intéresse à la structure des mots ou morphologie. Le générateur de ce niveau a pour but de déterminer les structures morphologiques bien formées ; on appelle ce niveau MS. Le second niveau s'intéresse à la structure des constituants, c'est-à-dire à la structure de la phrase. Le générateur de ce niveau, appelé CS, définit les structures syntaxiques de la langue. Ces deux niveaux définissent les mots et phrases bien formés, et représentent la grammaire de la langue.

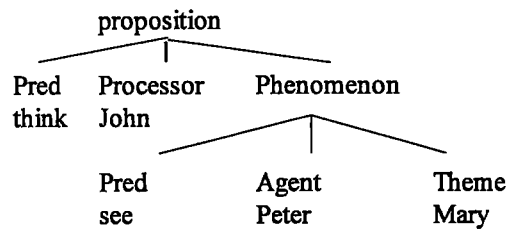
Bien que CAT2 puisse analyser des phrases, son but premier est de traduire des phrases d'une langue dans une autre. Par conséquent, il ne faut pas s'intéresser uniquement aux structures bien formées d'une langue mais il faut aussi voir, au niveau de la traduction, comment les structures d'une langue sont transformées en structures équivalentes dans une autre langue. A cette fin, on a établi un niveau de représentation supplémentaire qui agit comme une interface entre deux langues ; ce niveau est appelé IS (interface structure).

Alors que le niveau CS représente la structure syntaxique de la phrase, le niveau IS se rapporte davantage à la structure sémantique de la phrase, ou plus précisément, à sa structure prédicat-argument.

Représentation CS

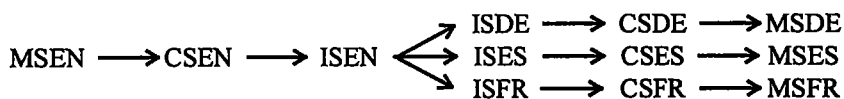


Représentation IS



L'IS fait abstraction des structures syntaxiques spécifiques à une langue et représente la phrase sous une forme plus adéquate pour la traduction. Soit la phrase équivalente en allemand *Hans dachte, daß Peter Maria sah*, la CS correspondante est différente de la CS anglaise, mais les représentations IS des deux langues sont isomorphiques. C'est cette propriété d'isomorphisme qui simplifie le processus de traduction de CAT2 : la traduction de l'IS anglaise en IS allemande est la traduction systématique de chaque constituant.

Diagramme de la stratégie de traduction :



Un texte a la structure morphologique MSEN, qui est analysée comme ayant une structure de constituants CSEN, et ceci est apparié au niveau de la structure d'interface ISEN. A partir de ISEN on peut traduire en ISDE allemande, ISES espagnole, ISFR française, chacune ayant ses représentations CSXX et MSXX correspondantes.

En résumé, le formalisme CAT2 définit deux choses : les niveaux de représentation et l'appariement des structures entre les niveaux. Ceux-ci sont définis par les générateurs et les règles d'appariement par les translateurs.

5.1.2. Règles vs objets

Les règles qui forment un générateur sont utilisées pour construire les structures appelées objets. Un objet est la représentation d'un mot ou d'une phrase, et prend la forme d'un arbre, comme on peut le voir dans les exemples de représentation CS et IS de la section précédente. Un arbre a une racine ou nœud-mère qui se ramifie en nœuds-filles, qui, à leur tour, se ramifient pour se terminer en nœuds-feuilles qui n'ont pas de filles. Chaque nœud de l'arbre contient un ensemble de traits appelé paquet de traits, c'est-à-dire une collection non ordonnée de paires attribut-valeur.

En CAT2, les règles décrivent des morceaux d'arbres, en utilisant des systèmes notationnels pour décrire les branches optionnelles ou alternatives. On se sert aussi de ces systèmes pour décrire les paquets de traits à l'intérieur des nœuds, appelés descriptions de paquets de traits, en utilisant les contraintes positives, négatives, disjonctives... sur les traits et leurs valeurs. Comme les règles et les objets sont décrits sous forme d'arbres et de traits, il est important de faire une distinction : les règles sont des descriptions de représentations et les objets sont les représentations elles-mêmes.

Au cours du processus d'application des règles, CAT2 utilise l'opération d'unification, pour, à la fois, construire les arbres et combiner les traits. L'unification des arbres implique l'unification, de gauche à droite, des nœuds-filles d'une règle avec des sous-objets existants, créant un objet plus grand. L'unification des traits est définie en termes de relation standard de subsomption que l'on trouve dans tous les formalismes de grammaire basés sur l'unification, c'est-à-dire que les traits d'une règle sont combinés avec les traits d'un objet de telle manière que le paquet de traits de l'objet qui en résulte subsume les traits dans la règle et dans l'objet.

Une notion importante liée à l'unification est la satisfaction des contraintes. Les descriptions de traits dans une règle expriment les contraintes sur les traits dans un objet. L'unification ne peut pas résoudre immédiatement les contraintes, c'est-à-dire que les contraintes négatives et disjonctives sont retenues avec l'objet tant qu'il n'y a pas plus d'informations disponibles.

Les générateurs de CAT2 utilisent uniquement deux types de règles : les règles *b* pour construire les structures et les règles *f* pour contrôler le contenu des traits des structures. De même, les translateurs de CAT2 utilisent deux types de règles : les règles *t* pour transformer les structures de niveau à niveau et les règles *tf* pour transformer les traits de structure à structure.

5.2. LE FORMALISME CAT2

CAT2 est programmé en PROLOG et utilise ses conventions syntaxiques pour écrire les états. Le langage de l'utilisateur en CAT2 a ses propres fonctions et ses propres commandes.

Les états de CAT2, c'est-à-dire les règles et les directives de compilation sont des termes PROLOG. Tous les termes doivent se terminer par un point, et être suivis par au moins un espace, ou une tabulation, une fin de ligne ou une fin de fichier. S'ils ne sont pas fixés, les blancs peuvent être utilisés librement dans les termes.

Les termes en CAT2 utilisent les crochets, les accolades et les parenthèses.

On peut placer les commentaires n'importe où dans un fichier CAT2. Est considéré comme commentaire tout ce qui apparaît derrière le symbole de pourcentage (%) et jusqu'à la fin de la ligne, ou tout ce qui est situé entre /* et */.

5.2.1. Constantes et variables

Un terme CAT2 se compose de constantes et de variables.

Une constante est un symbole atomique, c'est-à-dire toute séquence de caractères alphanumériques, incluant le souligné _, et commençant par une minuscule, ou toute chaîne de caractères délimitée par des apostrophes. Toute constante contenant des caractères non alphanumériques doit être mise entre apostrophes. Si le caractère apostrophe est inclus dans la chaîne, il doit être écrit deux fois (par ex. 'don''t'). Une chaîne vide est indiquée par deux apostrophes (i. e. ""). Une constante peut aussi être un nombre (par ex. 3 ou 98.6)? Il faut noter que 3 et '3' sont deux constantes différentes : la première est un nombre, la seconde une chaîne.

Une variable est une séquence de caractères alphanumériques (y compris le souligné) commençant par une majuscule ou un souligné (X, Case, AGR, _, _nil). Les variables représentent des valeurs non spécifiées. Si deux variables, dans une règle ont le même nom, elles représentent la même valeur ; c'est-à-dire que si une variable est instanciée avec une certaine valeur, la seconde est instanciée de façon identique. Une variable qui apparaît une seule fois dans une règle doit être écrite comme la variable anonyme, indiquée par un souligné comme premier caractère. Ce nombre de variables anonymes dans une règle n'est pas fixe ; elles représentent toutes des variables distinctes.

Les noms des variables sont internes à la règle contenant la variable en question. L'instanciation d'une variable X dans une règle n'a pas d'effet sur une variable X dans une autre règle. En d'autres termes, la portée de la variable est la règle.

5.2.2. Un système de traits

5.2.2.1. LES TRAITS

Une description de traits comprend un attribut, un opérateur et une valeur (per=3). La description d'un paquet de traits correspond à un ensemble de descriptions de traits, et se compose de zéro, une ou plusieurs descriptions de traits séparées par des virgules et

entourées d'accolades. Comme c'est un ensemble, les traits ne sont pas ordonnés. La description d'un paquet de traits vide est notée {}. Un attribut est un symbole de constante. Les opérateurs sont = (égal), \neq (non égal), \exists (opérateur existentiel). Une valeur peut être une constante, une variable ou la description d'un paquet de traits, dans ce dernier cas, on parle de trait complexe ; si la valeur est une constante, on parle de trait atomique.

5.2.2.2. L'UNIFICATION

L'unification fusionne les traits contenus dans une règle de description d'un paquet de traits avec un objet d'un paquet de traits de façon à ce que, après chaque unification, l'objet a ses traits d'origine plus ceux de la règle et ne contient pas deux fois le même trait. On note l'unification d'un paquet de traits (FBD pour *Feature Bundle Description*) avec le paquet de traits d'un objet (FB pour *Feature Bundle*) : $FBD \cup FB = FB'$

5.2.2.3. LES CONTRAINTES

Il existe différents types de contraintes :

- contrainte positive : $ATTR=VAL$: le paquet de traits de l'objet doit inclure (via l'unification) le trait $ATTR=VAL$
- contrainte négative : $ATTR\neq VAL$: le paquet de traits de l'objet n'est pas compatible avec le trait $ATTR=VAL$. La contrainte négative n'assigne pas une valeur à un trait, elle contrôle seulement ce que cette valeur ne doit pas être. Si $ATTR$ apparaît dans le paquet de traits de l'objet, alors sa valeur ne doit pas être compatible avec VAL .
- contrainte disjonctive : $ATTR=(VAL1;VAL2;...;VALn)$: la valeur de l'attribut $ATTR$ s'unifie avec $VAL1$, $VAL2$, ..., ou $VALn$. Les valeurs de $VAL1$, $VAL2$... sont des éléments disjoints et doivent être séparés par des ; et mis entre parenthèses.
- contrainte existentielle : $ATTR\exists VAL$: 1) l'attribut $ATTR$ existe dans le paquet de traits de l'objet et 2) sa valeur s'unifie avec VAL .
- contrainte concaténative : $ATTR=SEG1+SEG2+...+SEGN$ (où chaque $SEGi$ est une constante ou une variable) : elle décompose une valeur atomique en segments ou inversement compose une valeur atomique à partir de segments.
- contrainte conditionnelle : $CONDITION \Rightarrow CONSEQUENCE$ (où $CONDITION$ et $CONSEQUENCE$ sont une simple description d'un paquet de traits ou une disjonction de descriptions d'un paquet de traits) : si on peut satisfaire les contraintes dans la description de la $CONDITION$, alors on doit aussi les satisfaire dans la $CONSEQUENCE$ mais si on ne peut satisfaire aucune contrainte dans la $CONDITION$ alors la règle contenant la contrainte conditionnelle ne peut être appliquée.

5.2.2.4. LES ATTRIBUTS FIXES

Il existe trois attributs fixes dans CAT2 :

- lex : le trait *lex* est obligatoire dans toutes les règles b atomiques et sa valeur doit être une constante ou une disjonction de constantes. Sa valeur est généralement le stemme ou la racine du mot, mais on peut choisir n'importe quelle valeur tant qu'elle est atomique.
- string : les règles b atomiques contiennent le trait *string* avec une constante ou une disjonction de constantes comme valeur, indiquant la forme de surface du mot, c'est-à-dire la forme orthographique du mot dans le texte. Ce trait est connu comme CLE EXTERNE, il est utilisé pour un accès rapide aux entrées lexicales durant l'analyse morphologique. Le trait *string* peut apparaître du côté CONSEQUENCE d'une règle b écrite comme contrainte conditionnelle. Ceci est particulièrement utile si plusieurs formes différentes d'un mot ont certains traits en commun. Les traits en commun seront définis du côté CONDITION et les traits variables du côté CONSEQUENCE :

```
{lex=decide} >> ( {string=decision, cat=n,...};
                   {string=decisions, cat=n,...};
                   {string=decide, cat=v,...}).
```

- word# : dans une phrase contenant n éléments, on assigne à chacun d'eux le trait 'word#=i', où i est le numéro de l'élément de 1 à n. Dans *The council adopted the decision*, le premier élément aura comme paquet de traits : {'word#=1; string='the'} (voir la différence de notation de string#). On utilise les apostrophes en raison du dièse (#).

5.2.2.5. LES VALEURS FIXES

Les valeurs fixes sont également au nombre de trois :

- @num : c'est la valeur attribuée à *lex*, si le lexème est numérique (entier, décimal...)
- \$index : c'est une fonction qui génère un entier unique à chaque fois que la fonction est rencontrée. On l'utilise pour générer des indices afin de faire des co-références, pour assigner des entiers uniques à des constituants lexicaux ou syntaxiques. La première fois que '\$index' est rencontré dans une règle, l'entier 1 est généré. Toutes les applications suivantes de \$index augmentent l'indice de 1.
- @unknown : cette valeur est automatiquement assignée à l'attribut *lex*, pendant l'entrée du texte, si l'élément analysé n'est pas trouvé au niveau morphologique.

5.2.3. Les niveaux

On trouve trois types de niveaux : morphologique, syntaxique et relationnel. Ces niveaux sont décrits par des générateurs. Un objet créé par un générateur à un niveau est transformé en un objet à un autre niveau grâce aux translateurs.

5.2.3.1. LES GENERATEURS

@level (LEVEL/TYPE/LANGUAGE)

LEVEL représente la constante atomique indiquant le nom du niveau : (CSFR)

TYPE peut prendre les valeurs morphologique, syntaxique ou relationnel : (morph)

LANGUAGE est une constante atomique indiquant le nom de la langue décrite par le niveau : (french)

Le nom du niveau doit être unique ; aucun autre niveau ne peut avoir le même nom, même s'il a un autre type ou une autre langue.

Un générateur est constitué de règles b qui décrivent les structures possibles à un niveau donné et de règles f qui décrivent le contenu des traits des structures générées. Il y a trois types de générateurs :

- morphologique : responsable de l'analyse des mots dans un texte ou de leur génération durant la synthèse du texte. Les morphèmes peuvent être libres, c'est-à-dire qu'ils peuvent apparaître comme mots entiers, ou liés, c'est-à-dire qu'ils apparaissent en cooccurrence avec d'autres morphèmes pour former un mot complet et incluent les suffixes, affixes^o et les stemmes,
- syntaxique : au cours de l'analyse d'un texte d'entrée, le niveau syntaxique est utilisé pendant le parsing pour bâtir une structure des constituants. En synthèse, il est utilisé pour autoriser les structures provisionnelles créées par le traducteur,
- relationnel : le but du niveau relationnel dépend de l'application. Dans le cas de la TA, ce niveau définit une structure d'interface d'une phrase. Deux langues sont alors en relation via cette structure d'interface plutôt que via leurs structures syntaxiques.

5.2.3.2. LES TRANSLATEURS

Il y en a de deux types :

- les translateurs normaux, notés :

@level (LEVEL1↔LEVEL2)

où LEVEL1 et LEVEL2 sont les noms de générateurs de niveau syntaxique et/ou relationnel. Le ↔ symbolise la bidirectionnalité du niveau traducteur. Un traducteur comprend des règles t et des règles tf qui définissent l'appariement des structures au niveau source à celles du niveau cible.

- les translateurs robustes, notés :

@robust (LEVEL1↔LEVEL2)

où LEVEL1 et LEVEL2 sont des niveaux morphologiques. Ils ne sont activés que si la traduction normale échoue et que le mode robuste est activé.

5.2.3.3. COMMUN

Dans la grammaire, un niveau commun est introduit par :

`@common(NAME)`

où NAME est une constante atomique indiquant le nom du niveau commun. Les règles définies à ce niveau sont supposées être partagées entre deux ou plusieurs niveaux. Si LEVEL1 et LEVEL2 ont un ensemble de règles identiques, on peut les retirer de LEVEL1 et LEVEL2 et les définir dans un niveau commun. LEVEL1 et LEVEL2 feront alors référence au niveau commun en incluant dans leurs définitions une directive `@call` au niveau commun.

```
Ex. : @level (csen/syntactic/english) %level CSEN
...
@call (common 3)
...
@level (csde/syntactic/german) % level CSDE
...
@call (common 3)
...
...
@common (common 3) % common level COMMON 3
...
```

Un grand nombre de niveaux peuvent appeler un niveau commun et un niveau peut appeler un grand nombre de niveaux communs. Un niveau commun peut même appeler d'autres niveaux communs. Un niveau commun ne peut s'appeler lui-même directement ou par le biais d'autres niveaux sinon il bouclerait en cherchant et en appliquant les mêmes règles. Un niveau donné ne peut appeler deux fois le même niveau commun. Les règles communes peuvent contenir des règles b, f, t et tf.

5.2.4. Les types de règles

Une grammaire CAT2 comprend 4 types de règles :

```
règles b }
règles f } pour les générateurs

règles t }
règles tf } pour les translateurs
```

`@rule (TYPE)` est utilisé pour définir un type de règles, où TYPE est b, f, t (les règles tf sont déclarées être du type f). Toutes les règles suivant la directive `@rule` sont considérées comme étant des règles du type déclaré jusqu'à la prochaine directive `@rule`, `@level` ou `@common` où jusqu'à ce que la fin du fichier soit atteinte. Toutes les règles de CAT2 ont un nom et sont de la forme NAME=RULE.

5.2.4.1. LES REGLES B

Les règles b définissent des parties d'arbres. Elles spécifient la racine de l'arbre et ses filles immédiates. Elles sont de la forme :

ROOT.[BODY]

où ROOT est la description d'un paquet de traits décrivant la racine de l'arbre, et BODY donne la liste des filles immédiates situées sous la racine. Il y a deux types de règles b : atomiques et non atomiques. Elles sont de deux types :

- * atomiques : elles n'ont pas de fille et sont de la forme : ROOT.[].

Elles décrivent des entrées lexicales, les éléments terminaux d'un arbre. Le BODY est une liste vide. Elles doivent obligatoirement contenir le trait *lex*.

- * non atomiques : elles décrivent des nœuds non-terminaux de l'arbre, elles correspondent à des règles de structures de phrase. Elles décrivent un arbre partiel, c'est-à-dire le nœud de la racine et ses filles immédiates. Une règle b non-atomique correspond à une règle de réécriture. La règle de réécriture :

S -> NP VP

se note en CAT2 sous la forme : rule = {cat=S}.[{cat=np}, {cat=vp}].

On peut utiliser divers opérateurs tels que ^ (optionnalité), * (0 ou plusieurs branches (ou séquences de branches)), + (1 ou plusieurs branches (ou séquences de branches)), ; (alternance).

5.2.4.2. LES REGLES F

Les règles f appliquent des contraintes de traits aux objets créés par les règles b. Elles ne peuvent ni modifier les structures syntaxiques ni effacer ou changer des traits existants. Elles sont de la forme :

ROOT.[BODY]

où ROOT est la description d'un paquet de traits et BODY décrit les filles sous la racine. Contrairement aux règles b, BODY peut décrire les arbres de filles jusqu'à n'importe quelle profondeur, chaque sous-arbre ayant la forme ROOT.[BODY]. Elles ne se limitent pas aux filles immédiates. Les règles f qui opèrent sur des objets atomiques spécifient un BODY vide, ROOT.[]. On peut utiliser les mêmes opérateurs ^, *, + et ; que pour les règles b. En plus ^, * et + peuvent être utilisés seuls pour représenter un constituant optionnel arbitraire. Ex. : frule = { }.[* , {cat=vp, ...}, *]. Les règles f sont appliquées à un objet dans l'ordre où elles sont définies dans la grammaire.

L'applicabilité des règles f est déterminée par la satisfaisabilité des contraintes conditionnelles apparaissant dans la règle. Si, dans la règle, toutes les parties de la condition sont satisfaisables, alors toutes les parties de la conséquence doivent l'être. Les descriptions de paquets de traits qui n'impliquent pas de contraintes conditionnelles sont traitées comme si elles avaient une partie condition sans partie conséquence, de la forme {...}>>{ }. On peut appliquer des valeurs par défaut à une structure en utilisant des règles f sans contrainte conditionnelle. Par exemple, pour assigner le trait 'wh=no' à

toutes les structures, on peut définir la règle f suivante : $wh_no=\{wh=no\}.[*]$. On peut obliger toutes les structures NP à avoir le trait 'wh=no' en utilisant la contrainte conditionnelle : $wh_no=\{cat=np\}\gg\{wh=no\}.[*]$. En utilisant * avec une conséquence disjonctive, on peut établir que les constituants suivant le verbe, s'il y en a, sont des combinaisons de NP et de non-NP, et s'il y a un NP, son paquet de traits doit être unifiable avec 'case=acc' :

$$accus2=\{cat=vp\}.\{\{cat=v\},*\{ \}\gg(\{cat=np,case=acc\};\{cat\sim np\})\}.$$

5.2.4.3. LES REGLES T

Les règles t transforment les structures du LEVEL1 au LEVEL2 et inversement. Elles peuvent être uni- ou bidirectionnelles et avoir l'une des formes suivantes :

$$LHS \Rightarrow RHS$$

$$LHS \Leftarrow RHS$$

$$LHS \Leftrightarrow RHS.$$

Quand on utilise un traducteur, l'un des niveaux est le niveau source, l'autre le niveau cible. La règle t établit que si l'objet source satisfait les contraintes du côté source, alors un arbre est construit qui doit satisfaire les contraintes du côté cible, être autorisé au niveau cible. Un arbre est autorisé s'il satisfait les contraintes d'une règle b et toutes les règles f applicables au niveau cible. Les contraintes conditionnelles sont inutiles dans les règles t puisque si une contrainte n'est pas satisfaisable, la règle t n'est pas appliquée. Les règles t ne sont applicables qu'aux niveaux syntaxique et relationnel, aucune règle t ne peut être appliquée au ou à partir du niveau morphologique.

5.2.4.4. LES REGLES TF

Les règles tf transfèrent des traits des structures sources vers les structures cibles. Elles sont aussi bidirectionnelles et sont de la forme :

$$ROOT1.[BODY1] \Rightarrow ROOT2.[BODY2]$$

$$ROOT1.[BODY1] \Leftarrow ROOT2.[BODY2]$$

$$ROOT1.[BODY1] \Leftrightarrow ROOT2.[BODY2]$$

où $ROOT_i$ et $BODY_i$ décrivent le paquet de la racine et les sous-arbres filles de l'objet source ou de l'objet cible provisoire, tout dépend de la direction de la traduction.

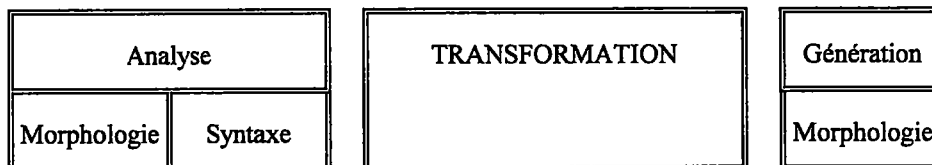
5.2.5. Processus utilisés dans l'interprétation des règles

On peut scinder le système de traduction CAT2 en différentes phases : analyse et génération qui comprennent les niveaux linguistiques MORPHOLOGIE, SYNTAXE et SEMANTIQUE, et transfert qui décrit concrètement le processus de traduction de la langue source dans la langue cible.

Analyse			TRANS- FERT	Génération		
Morphologie	Syntaxe	Sémantique		Sémantique	Syntaxe	Morphologie

TABLEAU 22 : PHASES DU PROCESSUS DE TRADUCTION ¹

Du point de vue du modèle de traitement de CAT2, il faut considérer la division de façon différente. On peut ainsi distinguer l'analyse morphologique, l'analyse syntaxique, la phase de transformations des arbres et la génération morphologique.

TABLEAU 23 : PHASES DE LA TRADUCTION DU POINT DE VUE DU MODELE DE TRAITEMENT ²

5.2.5.1. INPUT

Le processus d'input lit un texte à partir du terminal ou d'un fichier et le convertit en objet d'input à analyser ou à traduire. Durant l'input, le texte est segmenté en « unités de traduction » et en « token », qui, à la base, sont des phrases et des mots. Chaque unité de traduction devient un objet d'input, composé des descriptions de paquets de traits initiales pour chacun des « tokens » constituant l'unité. Cet objet est lu par l'analyseur morphologique, dans lequel les descriptions de paquets de traits sont unifiées avec les traits dans le dictionnaire pour devenir des objets morphologiques.

5.2.5.1.1. Les unités de traduction

Une unité de traduction est une séquence, de un ou plusieurs caractères, qui forme une unité simple et cohérente à traduire ; elle se termine par un des caractères suivants :

- * un point '.' suivi d'un blanc, d'une tabulation, de <CR> ou d'une fin de fichier,
- * un point d'interrogation '?' suivi d'un blanc, d'une tabulation, de <CR> ou d'une fin de fichier,
- * un point d'exclamation '!' suivi d'un blanc, d'une tabulation, de <CR> ou d'une fin de fichier,
- * un <CR> suivi immédiatement d'un autre <CR> ou d'une fin de fichier,
- * une fin de fichier.

Les trois premiers cas concernent, respectivement, des phrases déclaratives, interrogatives et exclamatives, le quatrième des titres ou des phrases isolées. Comme le point peut faire partie d'une abréviation, on ne peut pas considérer qu'il termine une unité de traduction, si le caractère qui suit le point et le blanc est un chiffre ou une minuscule.

¹ Tableau repris dans D. GESKE : *Das maschinelle Übersetzungssystem CAT2*, Sarrbrücken, 1996.

² *Ibid.*

5.2.5.1.2. Les tokens

Un token est une chaîne de un ou plusieurs caractères qui forment un mot, un nombre ou un signe de ponctuation. Un mot est une combinaison de symboles : lettres majuscules et minuscules, chiffres, points, apostrophes et soulignés. Un nombre est une chaîne de chiffres, incluant, de façon optionnelle, un point précédé et suivi d'au moins un chiffre. Un signe de ponctuation est représenté par n'importe quel autre caractère imprimable. Les tokens sont séparés par des blancs, des tabulations ou des <CR>.

5.2.5.2. L'ANALYSE MORPHOLOGIQUE

Elle prend comme input la liste des descriptions de paquets de traits initiales produites par le processus d'input, et produit pour chacune d'entre elles une liste de un ou plusieurs objets atomiques, où la valeur de string dans la description du paquet de traits s'apparie avec celle de ou des entrée(s) lexicale(s) au niveau morphologique correspondant. Il en résulte une liste qui devient un objet morphologique, ce dernier est examiné par le parser pour l'analyse sémantique.

Si l'anglais est la langue source, le niveau morphologique sera défini de la façon suivante :

@level(msen/morph/english)

CAT2 ne possède pas sa propre analyse morphologique. Les entrées dans le niveau morphologique de la grammaire forment plutôt un dictionnaire de formes pleines, dans lequel on peut sélectionner les structures de traits correspondantes.

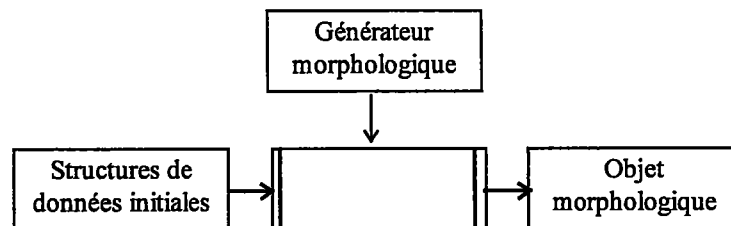


FIGURE 15 : ELABORATION D'UN OBJET MORPHOLOGIQUE AVEC UN DICTIONNAIRE DE FORMES PLEINES ¹

5.2.5.2.1. Les inputs numériques

Si un nombre apparaît dans le texte d'input, un objet atomique est créé, le trait *string* prend la valeur du nombre et le lexème prend la valeur fixe '@num'. On peut ajouter des traits au paquet de traits en utilisant des règles f qui permettront notamment de définir la catégorie syntaxique (par exemple CARD pour cardinal), ou l'accord en nombre dépendant de la valeur numérique.

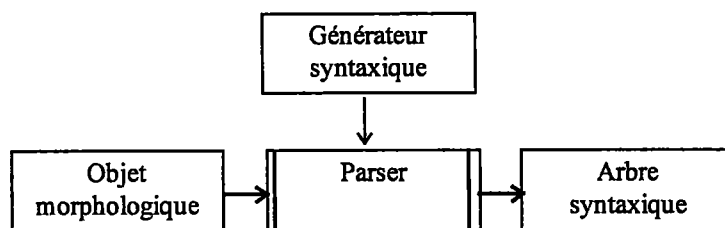
5.2.5.2.2. Les mots inconnus

Si aucune entrée morphologique n'est trouvée pour la chaîne d'input, un objet atomique est créé dont le trait *lex* prend la valeur atomique '@unknown'.

¹ La figure est extraite de *op. cit.* [Geske96].

5.2.5.3. L'ANALYSE SYNTAXIQUE

À partir d'un objet morphologique, pris comme input, le parsing (ou analyse syntaxique) construit une structure syntaxique. L'analyse est réussie quand au moins une structure syntaxique est construite et qu'elle contient tous les tokens de l'input.

FIGURE 16 : ANALYSE SYNTAXIQUE EN CAT2 ¹

Le parser travaille de façon ascendante, en combinant les mots dans des phrases, les phrases dans des phrases plus grandes, ou éventuellement en une seule phrase qui contient tous les mots de l'input d'origine. Cette structure syntaxique représente alors l'output du parser.

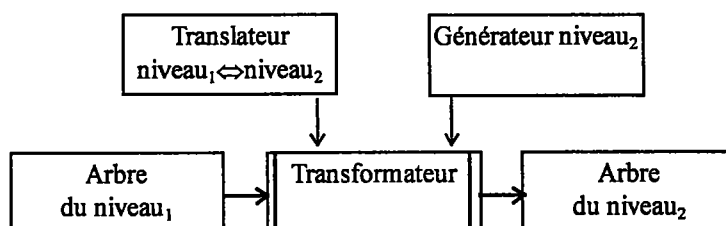
L'algorithme de CAT2 scanne essentiellement tous les items dans l'input et complète toutes les règles possibles qui contiennent l'item scanné. C'est pour ces deux raisons que la base de règles doit être minimale en CAT2.

5.2.5.4. LES TRANSFORMATIONS

Le processus de transformation a pour but de transformer les structures d'arbres entre les niveaux de représentation. Comme ces structures sont définies pas les générateurs morphologiques et relationnels, ce sont les seules qui puissent être touchées par les transformations.

Un objet du niveau syntaxique peut être transformé en un objet du niveau relationnel ou en un autre objet du niveau syntaxique.

Pour transformer un objet du niveau L_1 en un objet du niveau L_2 , on a besoin d'un traducteur qui connecte les deux niveaux (soit $L_1 \leftrightarrow L_2$, soit $L_2 \leftrightarrow L_1$), et le générateur pour le niveau L_2 .

FIGURE 17 : TRANSFORMATIONS D'ARBRES ²

¹ La figure est extraite de *op. cit.* [Geske96].

² *Ibid.*

5.2.5.5. LA SYNTHÈSE SYNTAXIQUE

La synthèse syntaxique correspond au processus de transformation d'une structure d'input, objet qui a été généré soit à un niveau relationnel soit à un niveau syntaxique, en une structure syntaxique définie un générateur du niveau syntaxique.

5.2.5.6. LA SYNTHÈSE MORPHOLOGIQUE

Les feuilles d'un arbre syntaxique contiennent les propriétés morphologiques des mots. Chaque feuille est un sous-objet atomique de l'arbre. Lors de la synthèse morphologique, il faut tout d'abord isoler les feuilles de l'arbre et en faire une liste, puis appliquer les règles à chacun des objets au niveau morphologique. Les règles ont pour fonction de déterminer la valeur de la chaîne de caractères ou les valeurs des objets atomiques, si elles ne sont pas encore disponibles.

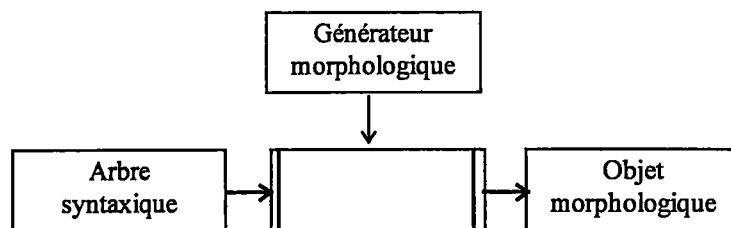


FIGURE 18 : GENERATION MORPHOLOGIQUE DANS CAT2 ¹

5.2.5.7. OUTPUT

Le processus d'output prend comme input la liste des paquets de traits produits par la synthèse morphologique, assigne un nombre à chaque paquet de traits, extrait la valeur de la clé externe de chacun d'entre eux et la fait apparaître sous forme de texte sur l'écran ou dans un fichier.

5.2.5.8. LE CHEMIN DE TRADUCTION

Le chemin de traduction représente la séquence de niveaux que l'objet source traverse jusqu'au niveau cible. Il guide la traduction en déterminant quel générateur et quel translateur activer. Il connecte les niveaux source et cible par le plus petit chemin possible, en sélectionnant le nombre minimum de translateurs et de générateurs qui connectent les deux.

5.2.5.9. TRADUCTION ROBUSTE

Dans une traduction classique, le chemin de traduction décrit les étapes qu'une phrase prise comme input traverse de la langue source à la langue cible. Si la traduction échoue, le système passe dans un mode robuste. Dans ce mode, chaque mot de l'input est traduit individuellement, en utilisant le même chemin de traduction que pour une traduction normale. Même si un mot ne peut pas être traduit, CAT2 produit un texte en sortie. Par exemple, si on veut traduire la phrase : *Der Rat verabschiedet den Beschluss*

¹ La figure est extraite de *op. cit.* [Geske96].

en anglais, mais que la règle t : $trule = \{lex=beschluss\}.\{ \} \Leftrightarrow \{lex=decision\}.\{ \}$.
manque, le résultat sera le suivant : **the council adopted the Beschluss*. De même si un mot a plusieurs traductions, elles apparaîtront entre $\langle \rangle$: **the <council/advice> adopted the Beschluss*.

Quand la traduction normale échoue, le système retrouve le premier (ou le seul) objet syntaxique créé à partir du texte source, extrait les feuilles (atomes) de l'arbre, et dresse la liste des atomes ressemblant à un objet morphologique. Si aucun arbre syntaxique n'est généré, l'objet morphologique produit par le processus d'input est utilisé pour la traduction robuste.

5.2.5.9.1. Niveau de traducteur robuste

On peut définir un niveau de traducteur robuste pour contrôler directement la traduction des mots du niveau morphologique source au niveau morphologique cible, en évitant, de ce fait, tous les niveaux syntaxiques et relationnels et les règles t qui les connectent.

Dans un niveau de traducteur robuste, les règles peuvent inclure des règles t , des règles tf et des directives $@call$ vers les niveaux de traducteurs communs.

5.2.5.9.2. Traduction robuste seulement

En utilisant la commande SET ROBUST ONLY, il n'y aura aucune analyse syntaxique sur le texte d'entrée. Le système passera immédiatement en mode robuste, en utilisant le niveau de traducteur robuste s'il est disponible ; sinon il suit le chemin de traduction normal. (processus de traduction)

5.3. PRESENTATION DU PROJET

Le but du programme est de traduire les prépositions spatiales françaises en allemand et inversement. Un corpus a été établi pour chaque langue. ¹

La principale difficulté réside dans le fait que bon nombre de prépositions de la langue cible peuvent se traduire par plusieurs prépositions dans la langue source. Durant le processus de traduction, le parser commencera par analyser le verbe, afin de savoir de quel type de verbe il s'agit (verbe de mouvement ou non, verbe d'extraction...), il scannera ensuite les différentes prépositions et le groupe nominal qui les suit, et c'est en fonction de ces trois éléments qu'il pourra déterminer quelle préposition, il devra utiliser. Dans certains cas, deux prépositions seront possibles, car le choix d'une seule préposition dépend, en fait, du contexte.

5.3.1. Récapitulatif sur les prépositions spatiales

Une étude des prépositions spatiales en français et en allemand nous a permis d'établir des règles d'emploi assez générales de celles-ci dans les deux langues, et selon le « domaine » dans lequel elles sont employées. Cependant, dans certains cas, nous

¹ Cf. en annexe.

avons été obligés de faire des règles plus spécifiques, car une généralisation semblait impossible.

Nous présenterons ci-dessous les résultats de nos investigations accompagnés d'un ou plusieurs exemples.

On peut d'ores et déjà noter que les indications « locatif » et « directif » n'ont aucune influence sur la préposition en français, alors qu'en allemand elles déterminent non seulement le choix de la préposition, mais également le cas qu'elle régit

5.3.1.1. LES LIEUX GEOGRAPHIQUES

5.3.1.1.1. Les continents

LOCATIF		DIRECTIF	
en + continent <i>en Amérique</i>	in + continent <i>in Amerika</i>	en + continent <i>en Amérique</i>	nach + continent <i>nach Amerika</i>
français : en + continent ¹			
allemand : in + continent (locatif), nach + continent (directif)			

5.3.1.1.2. Les pays

LOCATIF		DIRECTIF	
à + pays masculin <i>au Yémen</i>	in + pays avec article <i>im Yemen</i>	à + pays masculin <i>au Yémen</i>	in (+A) + pays avec article <i>in den Yemen</i>
ou	ou	ou	ou
à + pays masculin à initiale consonantique <i>au Luxembourg</i>	in + pays sans article <i>in Luxemburg</i>	à + pays masculin à initiale consonantique <i>au Luxembourg</i>	nach + pays sans article <i>nach Luxemburg</i>
en + pays féminin <i>en Suisse</i>	in (+D) + pays avec article <i>in der Schweiz</i>	en + pays féminin <i>en Suisse</i>	in (+A) + pays avec article <i>in die Schweiz</i>
ou	ou	ou	ou
en + pays avec voyelle <i>en Autriche</i>	in + pays sans article <i>in Österreich</i>	en + pays avec voyelle <i>en Autriche</i>	nach + pays sans article <i>nach Österreich</i>
français : à + pays masculin à initiale consonantique + Yémen en + pays féminin ou pays commençant par une voyelle (sauf Yémen)			
allemand : in (+ datif) + pays avec article, in + pays sans article (locatif) in (+ accusatif) + pays avec article, nach + pays sans article (directif)			

¹ Quand rien n'est précisé en français, c'est que la préposition est la même au locatif et au directif.

5.3.1.1.3. Les régions

LOCATIF		DIRECTIF	
en + région féminin <i>en Lorraine</i>	in + région sans article <i>in Lothringen</i> in + région avec article <i>in der Mongolei</i>	en + région féminin <i>en Lorraine</i>	nach + région sans article <i>nach Lothringen</i> in + région avec article <i>in die Mongolei</i>
dans + région masculin <i>dans le Tyrol</i>	in + région <i>in Tyrol</i>	dans + région masculin <i>dans le Tyrol</i>	nach + région <i>nach Tyrol</i>
français : en + région féminin dans + article + région masculin allemand : in (+ datif) + région avec article, in + région sans article (locatif) in (+ accusatif) + région avec article, nach + région sans article (directif)			

5.3.1.1.4. Les départements

LOCATIF		DIRECTIF	
en + département composé féminin <i>en Haute Garonne + en Guyane, en Savoie</i>	in + département avec déterminant <i>in der Haute-Garonne</i>	en + dép. composé féminin <i>en Haute Garonne + en Guyane, en Savoie</i>	in + département avec déterminant <i>in die Haute-Garonne</i>
dans + département pluriel <i>dans les Alpes Maritimes</i>	in + département pluriel <i>in den Alpes Maritimes</i>	dans + département pluriel <i>dans les Alpes Maritimes</i>	in + département pluriel <i>in die Alpes Maritimes</i>
dans + département masculin <i>dans le Jura</i>	in + département avec déterminant <i>in dem Jura</i>	dans + département masculin <i>dans le Jura</i>	in + département avec déterminant <i>in den Jura</i>
dans + dép. féminin simple <i>dans la Creuse</i>	in + département avec déterminant <i>in der Creuse</i>	dans + dép. féminin simple <i>dans la Creuse</i>	in + département avec déterminant <i>in die Creuse</i>
français : à + Paris en + Guyane, Savoie, département composé féminin dans + département pluriel, masculin, féminin simple allemand : in + département avec article (datif) ou département sans article (locatif) in + département avec article (acc.) ou nach + département sans article (directif)			

En allemand, l'usage veut que l'on garde le terme français mais également le genre. On pourra aussi ajouter le terme département (*ich fahre in den Departement Meurthe-et-Moselle*), ce qui permet, dans certains cas, de faire la distinction entre le département et le fleuve.

5.3.1.1.5. Les villes

LOCATIF		DIRECTIF	
à + ville <i>à Paris</i>	in + ville <i>in Paris</i>	à + ville <i>à Paris</i>	nach + ville <i>nach Paris</i>
français : à + ville ¹ allemand : in + ville (locatif), nach + ville (directif)			

¹ Nous ne tiendrons pas compte du cas particulier des villes commençant par *a* et pour lesquelles on peut utiliser la préposition *en*. L'usage n'est pas assez répandu.

5.3.1.1.6. Les îles

LOCATIF		DIRECTIF	
à + île sans déterminant <i>à Cuba</i>	auf + île <i>auf Kuba</i>	à + île sans déterminant <i>à Cuba</i>	nach + île <i>nach Kuba</i>
à + île au pluriel <i>aux Antilles</i>	auf + île au pluriel <i>auf den Antillen</i>	à + île au pluriel <i>aux Antilles</i>	nach + île au pluriel <i>auf die Antillen</i>
en + île féminin avec déterminant ¹ <i>en Corse</i>	auf + île <i>auf Korsika</i>	en + île féminin avec déterminant <i>en Corse</i>	nach + île <i>nach Korsika</i>
français : à + île sans déterminant, à + île au pluriel en + île féminin avec déterminant			
allemand : auf + île (locatif), nach + île (directif)			

5.3.1.1.7. Les autres lieux géographiques

5.3.1.1.7.1. Les déserts

LOCATIF		DIRECTIF	
dans + désert <i>dans le Sahara</i>	in + désert <i>in der Sahara</i>	dans + désert <i>dans le Sahara</i>	in + désert <i>in die Sahara</i>
français : dans + désert			
allemand : in + désert (locatif), in + désert (directif)			

5.3.1.1.7.2. Les montagnes

LOCATIF		DIRECTIF	
dans + montagne pluriel <i>dans les Alpes</i>	in + montagne ² pluriel <i>in den Alpen</i>	dans + montagne pluriel <i>dans les Alpes</i>	in montagne pluriel <i>in die Alpen</i>
dans + montagne singulier <i>dans le Caucase</i>	in + montagne singulier <i>in dem Kaukasus</i>	dans + montagne singulier <i>dans le Caucase</i>	in + montagne singulier <i>in den Kaukasus</i>
français : dans + montagne singulier ou pluriel			
allemand : in + montagne singulier ou pluriel (locatif), in + montagne singulier ou pluriel (directif)			

5.3.1.1.7.3. Les mers, les océans

LOCATIF		DIRECTIF	
sur + mer <i>sur la Mer Rouge</i>	auf + mer <i>auf dem Roten Meer</i>	sur + mer <i>sur la Mer Rouge</i>	auf + mer <i>auf das Rote Meer</i>
au bord de + mer <i>au bord de la Mer Rouge</i>	an + mer <i>an dem Roten Meer</i>	au bord de + mer <i>au bord de la Mer Rouge</i>	an + mer <i>an das Rote Meer</i>
français : sur + mer			
allemand : auf + mer (locatif), auf + mer (directif)			

¹ Attention, la préposition *en* s'emploie sans le déterminant.

Il semble que pour les noms d'îles masculins utilisés avec déterminant, on applique la règle des noms de pays (en fait, l'île se confond avec le pays : Japon, ce qui n'est pas le cas avec les noms féminins tels que la Corse...)

² En règle générale, les noms de montagnes sont précédées d'un déterminant en allemand, ce qui explique l'utilisation de la préposition *in*.

5.3.1.1.7.4. Les lacs

LOCATIF		DIRECTIF	
sur + lac <i>sur le Lac de Constance</i>	auf + lac <i>auf dem Bodensee</i>	sur + lac <i>sur le Lac de Constance</i>	auf + lac <i>auf den Bodensee</i>
au bord de + lac <i>au bord du Lac de Constance</i>	an + lac <i>an dem Bodensee</i>	au bord de + lac <i>au bord du Lac de Constance</i>	an + lac <i>an den Bodensee</i>
français : sur + lac allemand : auf + lac (locatif), an + mer (directif)			

5.3.1.1.7.5. Les rivières, les fleuves

LOCATIF		DIRECTIF	
sur + fleuve, rivière <i>sur la Seine</i>	auf + fleuve, rivière <i>auf der Seine</i>	sur + fleuve, rivière <i>sur la Seine</i>	auf + fleuve, rivière <i>auf die Seine</i>
au bord de + fleuve, rivière <i>au bord de la Seine</i>	an + fleuve, rivière <i>an der Seine</i>	au bord de + fleuve, rivière <i>au bord de la Seine</i>	an + fleuve, rivière <i>an die Seine</i>
français : sur + fleuve, rivière allemand : auf + fleuve, rivière (locatif), auf + fleuve, rivière (directif) français : au bord de + fleuve, rivière allemand : an + fleuve, rivière (locatif), an + fleuve, rivière (directif)			

5.3.1.2. LES BATIMENTS

5.3.1.2.1. Les bâtiments

LOCATIF		DIRECTIF	
dans / à l'école	in der Schule	à l'école	in die Schule ¹
dans le / au théâtre	im Theater	au théâtre	ins Theater
dans le / au cinéma	im Kino	au cinéma	ins Kino
dans / à l'église	in der Kirche	à l'église	in die Kirche
dans la / en prison	im Gefängnis	en prison	ins Gefängnis
français : dans / à + bâtiment masculin ou commençant par une voyelle dans la / en + bâtiment féminin à initiale consonantique allemand : in (+ datif) + bâtiment (locatif), in (+ accusatif) + bâtiment (directif)			

5.3.1.2.2. Cas particulier

LOCATIF		DIRECTIF	
à la poste	auf der Post	à la poste	auf die Post
à l'université	auf der Universität	à l'université	auf die Universität
à la banque	auf der Bank	à la banque	auf die Bank
français : à + bâtiment institutionnel allemand : auf (+ datif) + bâtiment ² (locatif), auf (+ accusatif) + bâtiment (directif) ³			

¹ On peut également dire *zur Schule* ou *zur Kirche gehen* si on considère non pas le bâtiment en tant que tel mais plutôt l'entité. Il existe également une troisième solution mais qui ne nous concerne pas dans notre étude puisqu'elle ne contient pas de préposition : *die Schule / die Kirche besuchen*.

² Bâtiment institutionnel sauf école.

³ Nous avons ici un cas particulier de l'utilisation de la préposition *auf*. En effet, nous sommes face à un emploi faussement spatial de *auf*, la détermination est abstraite et non spatiale.

5.3.1.3. LA MAISON

5.3.1.3.1. La maison

à	rentrer à la maison rester à la maison	nach Hause gehen zu Hause bleiben	nach + D zu + D
dans	dans la maison	im Haus(e)	in + D
de	chasser de la maison	aus dem Hause jagen	aus + D
devant	être devant la maison aller devant la maison	vor dem Haus stehen vor das Haus gehen	vor + D vor + A
français : à + maison allemand : zu + Hause (locatif), nach + Hause (directif), attention le mot Haus se met au datif et prend un e final Remarque : Pour les autres prépositions, il n'y a rien de particulier à signaler, il faut simplement vérifier si le verbe est ou non un verbe de mouvement en allemand afin d'utiliser le cas adéquat.			

5.3.1.3.2. Porte, fenêtre, mur

à	à la porte frapper à la porte au mur mettre au mur se mettre à la fenêtre	an der Tür an die Tür klopfen an der Wand an die Wand stellen sich ans Fenster stellen	an + D an + A
contre	jeter contre le mur	an die Wand werfen	an + A
dans	enfoncer un clou dans le mur	einen Nagel in die Wand schlagen	in + A
devant	être devant la porte	vor der Tür stehen	vor + D
par	par la porte regarder par la fenêtre sauter par la fenêtre	durch die Tür aus dem Fenster sehen aus dem Fenster springen	durch + A aus + D
français : à + porte, fenêtre, mur ¹ allemand : an (+ datif) + Tür, Fenster, Wand (locatif), an (+ accusatif) + Tür, Fenster, Wand (directif) ² Pour les autres prépositions, cf. remarque § 5.3.1.3.1.			

5.3.1.3.3. Le toit

sous	sous le toit	unter dem Dach unter das Dach	unter + D unter + A
sur	sur le toit	auf dem Dach auf das Dach	auf + D auf + A
Il ne se pose aucun problème pour la traduction de la préposition <i>sur</i> , en revanche, on constate que la préposition <i>sous</i> peut se traduire par <i>unter</i> (traduction habituelle) mais avec une construction particulière, ou par <i>in</i> , mais dans ce cas, le mot toit se traduit par <i>Haus</i> (la maison) et non pas par <i>Dach</i> .			

¹ On peut ajouter à ces exemples le mot tableau : *aller au tableau = an die (Wand-) Tafel gehen*

² On peut signaler que, dans les trois exemples, il y a toujours contact avec l'élément de localisation et que ce dernier représente une surface ; c'est ainsi que *contre* se traduit également par *an*.

5.3.1.3.4. Cas particuliers

Il semble préférable d'insérer ces exemples tels quels dans le dictionnaire, car il n'y pas équivalence entre les termes français et les termes allemands.¹

à	au grenier	unterm Dach	unter + D
à	au premier étage	im ersten Stock	in + D
à	au coin du feu	am Kamin	an
près	près du feu	am Feuer	an
après	une maison après l'autre	Haus an Haus	an
de en	aller de porte en porte	von Haus zu Haus gehen	von ... zu

5.3.1.4. LES MEUBLES

5.3.1.4.1. La table

à	se mettre à table	sich zu Tisch setzen	zu
de	se lever de table	vom Tisch aufstehen	von + D
sur	sur la table	auf dem Tisch	auf + D
	taper sur la table	auf den Tisch	auf + A
	mettre sur la table	auf den Tisch schlagen auf den Tisch bringen	auf + A auf + A
français : à + table allemand : bei + Tisch (locatif), zu + Tisch (directif) Pour les autres prépositions, cf. remarque § 5.3.1.3.1.			

5.3.1.4.2. L'armoire

dans	prendre qc dans une armoire	etwas aus einem Schrank nehmen	aus + D
	dans l'armoire	in dem Schrank	in + D
sous	sous l'armoire	unter dem Schrank unter den Schrank	unter + D unter + A
sur	sur l'armoire	auf dem Schrank	auf + D
		auf den Schrank	auf + A
Il n'y à rien de particulier à signaler si ce n'est que la préposition <i>dans</i> se traduira par <i>aus</i> quand le verbe sera un verbe d'extraction.			

5.3.1.4.3. Le tiroir

dans	prendre qc dans un tiroir	etwas aus einer Schublade nehmen	aus + D
	mettre qc dans un tiroir	etwas in eine Schublade stellen	in + A
	dans le tiroir	in der Schublade	in + D
Même remarque que pour le mot « armoire »			

¹ Nous ne nous occuperons pas des cas particuliers puisqu'il semble préférable de les mettre directement dans le dictionnaire, et par conséquent il n'y aura pas traitement de la préposition. Il en sera de même pour tous les paragraphes intitulés « cas particuliers ».

5.3.1.4.4. Le lit

à	aller au lit mettre au lit être / rester au lit se mettre au lit	zu Bett gehen zu Bett bringen im Bett liegen / bleiben sich ins Bett legen	zu zu in + D in + A
dans	se tourner et se retourner dans son lit	sich im Bett von einer Seite auf die andere wälzen	in + D
de	tirer qn du lit	jdn aus dem Bett holen / zerren	aus + D
français : à + lit allemand : in + Bett (locatif), zu + Bett (directif) Pour les autres prépositions, cf. remarque § 5.3.1.3.1.			

5.3.1.4.5. Les sièges

5.3.1.4.5.1. La chaise

sur	sur la chaise	auf dem Stuhl auf den Stuhl	auf + D auf + A
cf. remarque § 5.3.1.3.1.			

5.3.1.4.5.2. Le fauteuil

dans	s'effondrer dans un fauteuil s'enfoncer dans le fauteuil	sich in einen Sessel fallen lassen im Sessel versinken	in + A in + D
cf. remarque § 5.3.1.3.1.			

5.3.1.4.5.3. Le canapé

sur	sur le canapé	auf dem Sofa auf das Sofa	auf + D auf + A
cf. remarque § 5.3.1.3.1.			

5.3.1.4.5.4. Le divan

sur	sur le divan	auf dem Diwan auf den Diwan	auf + D auf + A
cf. remarque § 5.3.1.3.1.			

5.3.1.5. LES CONTENANTS

5.3.1.5.1. Cas général

LOCATIF		DIRECTIF	
dans la tasse	in der Tasse	dans la tasse	in die Tasse
français : dans + contenant allemand : in (+ datif) + contenant (locatif), in (+ accusatif) + contenant (directif)			

5.3.1.5.2. Les verbes de remplissage

en	mettre en bouteilles ¹ mettre en tonneaux	in Flaschen füllen in Fässer füllen	in + A
français : mettre (du liquide) en + contenant pluriel allemand : in (+ accusatif) + contenant pluriel + füllen			

en	mettre en sac	in einen Sack füllen	in + A
français : mettre en + contenant singulier allemand : in (+ accusatif) + contenant singulier + füllen			

5.3.1.5.3. Les verbes d'extraction

à	boire à la bouteille	aus der Flasche trinken	aus + D
dans	boire dans une tasse	aus einer Tasse trinken	aus + D
	boire dans un verre	aus einem Glas trinken	
	puiser dans un tonneau manger dans une assiette	aus einer Tonne schöpfen von einem Teller essen	von + D
français : verbe d'extraction + à + contenant avec une petite ouverture verbe d'extraction + dans + contenant avec une grande ouverture allemand : verbe d'extraction + aus + contenant (quelque soit la taille de l'ouverture) Il semble qu'il ne faut pas considérer l'assiette comme un contenant à proprement parler, car la préposition allemande est <i>von</i> qui marque l'éloignement plus que l'extraction.			

5.3.1.6. LES PERSONNES

5.3.1.6.1. après

après	l'un après l'autre	einer hinter dem andern	hinter + D
	être après qn	hinter jdm sein	
La prépositions <i>après</i> suivie d'un être humain se traduit par <i>nach</i> .			

5.3.1.6.2. chez

chez	se réfugier chez qn aller chez le médecin, chez Paul	zu jdm fliehen zum Arzt, zu Paul gehen	zu + D
chez	être chez le médecin, chez le coiffeur	bei dem Arzt, bei dem Frisoer sein	bei + D
chez	chez soi ² ne pas sortir de chez soi	zu Hause immer zu Hause sitzen	zu
chez	recueillir qn chez soi	jmanden ins Haus nehmen	in + A
de chez	de chez soi	von zu Hause	von zu
français : chez + humain allemand : bei (+ datif) + humain (locatif), zu (+ datif) + humain (directif) français : chez soi allemand : zu Hause (locatif), in das Haus (directif)			

¹ On peut noter qu'il existe également un verbe sans préposition pour le français, à savoir *embouteiller* et *entonner*.

² Les exemples avec des pronoms personnels en français ne sont pas traités car ils ne sont pas traduits par un pronom personnel en allemand.

5.3.1.6.3. devant / derrière

derrière	fermer la porte derrière soi	die Tür hinter sich zumachen	hinter
devant	devant le juge être / se mettre à genoux devant qn	vor dem Richter vor jemandem knien	vor + D
Il n'y a rien à signaler, les prépositions gardent leur traduction initiale, il faut uniquement vérifier si le verbe est ou non un verbe de mouvement.			

5.3.1.6.4. Corps humain

Pour les exemples ci-dessous, il n'y a aucun problème de traduction. La préposition allemande correspond à la préposition française et inversement. Il suffit de vérifier si le verbe est ou non un verbe de mouvement.

dans	dans la main s'enfoncer une épine dans le pied	in der Hand sich einen Dorn in den Fuß treten	in + D in + A
en	prendre qc dans la main	etwas in die Hand nehmen	in + A
sur	tomber sur le ventre sur le dos	auf den Bauch fallen auf dem Rücken	auf + A auf + D

5.3.1.6.5. Cas particuliers

Il semble préférable de mettre ces expressions telles quelles dans le dictionnaire, car il n'y a pas toujours équivalence entre les termes des deux langues.

dans	main dans la main	Hand in Hand	in
à	corps à corps	Brust an Brust	an
de à	de bouche à oreille	von Mund zu Mund	von zu
à même	à même la peau	auf der bloßen Haut	auf + D
dans	les yeux dans les yeux	Auge in Auge	in

5.3.1.7. DIVERS

5.3.1.7.1. La terre

Les exemples ci-dessous ne sont pas implémentés car il faut tenir compte du contexte pour obtenir une traduction correcte, or nous ne prenons pas le contexte en considération dans notre étude.

5.3.1.7.1.1. La terre comme sol

à	mettre à terre aller à terre (sport) jeter à terre	auf den Boden legen zu Boden gehen zu Boden werfen	auf + A zu + D
par	tomber par terre	zu Boden fallen	zu + D
de	sortir de terre	aus dem Boden schießen	aus + D

5.3.1.7.1.2. *La terre vue de la mer*

à	aller à terre mettre à terre	an Land gehen an Land bringen / setzen	an + A
sur	sur terre	zu Lande	zu + D
français : à + terre (directif), sur + terre (locatif)			
allemand : an + Land (directif), zu + Lande (locatif)			

5.3.1.7.1.3. *Cas particuliers*

Il semble préférable d'insérer ces exemples tels quels dans le dictionnaire, car ils correspondent à des expressions toutes faites, dans lesquelles les termes français et allemands ne sont pas équivalents.

en	porter en terre	zu Grabe tragen	zu + D
sur	sur mes / tes terres (et bienfonds)	auf eigenem Grund und Boden	auf + D
à	à ras de terre	dicht am Boden	dicht an
sur	sur la terre ferme	auf festem Boden	auf + D

5.3.1.7.2. *La mer*

5.3.1.7.2.1. *La mer comme lieu*

au bord de	au bord de la mer	an der See	an + D
en	en mer	auf See	auf
sur	sur la mer	auf der See	auf

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.2.2. *La mer par opposition au bateau*

Le problème qui se pose avec ces exemples réside dans le fait que, en français, les expressions signifient que l'entité se trouve à l'eau plutôt que sur le bateau, alors qu'en allemand, elles stipulent qu'elle est par-dessus bord sans toutefois préciser qu'elle est dans la mer, on ne traitera donc pas ces exemples.

à	à la mer tomber à la mer un homme à la mer	über Bord über Bord fallen Mann über Bord	über
à	à bord	an Bord	an
par-dessus	(jeter) par-dessus bord	über Bord (werfen)	über

5.3.1.7.2.3. *Cas particuliers*

Cf. remarque § 5.3.1.7.1.3.

par	par mer par terre et par mer	zur See zu Wasser und zu Lande	zu + D zu
sur	sur terre et sur mer	zu Wasser und zu Lande	zu

5.3.1.7.3. Le port

à	entrer au port	in den Hafen laufen	in + A
---	----------------	---------------------	--------

5.3.1.7.4. L'eau

à	sauter à l'eau jeter à l'eau se mettre à l'eau tomber à l'eau	ins Wasser springen ins Wasser werfen ins Wasser gehen ins Wasser fallen	{ in + A
sous	sous l'eau	unter Wasser	unter
sur	sur l'eau (flotter)	auf dem Wasser (treiben)	auf + D

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.5. Le sol

à	être au sol se plaquer au sol	am Boden liegen sich auf den Boden werfen	an + D auf + A
dans	s'enfouir dans le sol	sich in die Erde wühlen	in + A
sur	sur le sol	auf dem Boden, auf der Erde	auf + D

5.3.1.7.6. La boue

dans	traîner dans la boue	in den Schmutz ziehen	in + A
------	----------------------	-----------------------	--------

Il n'y a rien de particulier à signaler, la préposition française a son équivalent en allemand.

5.3.1.7.7. La campagne, la ville

5.3.1.7.7.1. La campagne

à	aller à la campagne être à la campagne habiter à la campagne	aufs Land gehen auf dem Lande sein auf dem Lande wohnen	auf + A auf + D auf + D
---	--	--	-------------------------------

Il n'y a rien de particulier à signaler, la préposition française a son équivalent en allemand. Toutefois, au locatif, il faut penser à mettre Land au datif et donc à rajouter un e.

5.3.1.7.7.2. La ville

de	expulser qn de la ville	jdn aus der Stadt weisen	aus + D
en	habiter en ville aller en ville	in der Stadt wohnen in die Stadt gehen	in + D in + A

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.8. La route

au bord de	au bord de la route	an der (Land-)Straße	an + D
------------	---------------------	----------------------	--------

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.9. La rue

à	jeter à la rue	auf die Straße werfen	auf + A
dans	dans la rue	auf der Straße	auf + D
	descendre dans la rue	auf die Straße gehen	auf + A
en	en pleine rue	auf offener Straße	auf + D
sur	les fenêtres de la chambre donnent sur la rue	die Fenster des Zimmers gehen auf die Straße	auf + A

Quelque soit la préposition française, la préposition allemande sera toujours *auf*, seul le cas changera en fonction du verbe (locatif ou directif).

5.3.1.7.10. Le marché

à	aller au marché	zum Markt	zu + D
	être au marché	auf dem Markt	auf + D
sur	lancer sur le marché	auf den Markt bringen	auf + A

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.11. La cour

dans	dans la cour	auf dem Hof	auf + D
sur	donner sur la cour	auf den Hof führen	auf + A

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.12. Le ciel

à	au ciel	im Himmel	in + D
	être au septième ciel	im siebten Himmel sein	in + D in + A
	aller au ciel	in den Himmel kommen	
dans	dans le ciel	am Himmel	an + D
	des nuages passent dans le ciel	Wolken ziehen am Himmel	an + D
de	du ciel	vom Himmel	von
entre	être suspendu entre ciel et terre	zwischen Himmel und Erde schweben	zwischen

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.¹

¹ Attention toutefois à l'expression : monter au ciel = gen Himmel fahren

5.3.1.7.13. Le sable

dans	se perdre dans le sable	im Sande verlaufen	in + D
sur	bâtir sur le sable	auf den Sand bauen	auf + A

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.14. Phénomènes météorologiques

sous	sous la pluie	in Regen	in + D
	sous un soleil de plomb	in glühender Sonne	
	sous une pluie battante	in strömendem Regen	

On est confronté ici à un cas particulier où la préposition *sous* suivie d'un phénomène météorologique se traduit par *in* (+ datif) en allemand, mais sans article.

5.3.1.7.15. Premier plan, arrière plan

à	être au premier plan	im Vordergrund stehen	in + D
	reléguer à l'arrière plan	in den Hintergrund drängen	in + A
	passer à l'arrière plan	in den Hintergrund treten	in + A
	en	être en arrière plan	im Hintergrund stehen

On peut noter que l'on utilise la préposition *à* dans tous les cas avec premier plan, *à* au directif avec arrière plan mais *en* au locatif. En allemand, la préposition usuelle est *in* suivie du cas correspondant.

5.3.1.7.16. La liste

de	rayer de la liste	aus der Liste streichen	aus + D
sur	être sur la liste	auf der Liste stehen	auf + D

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.17. La tombe

au bord de	être au bord de la tombe	am Rand des Grabe stehen	am Rand + G
dans	emporter dans la tombe	mit ins Grab nehmen	in + A
	se retourner dans sa tombe	sich im Grabe umdrehen	in + D
	avoir un pied dans la tombe	mit einem Fuß im Grabe stehen	

Il n'y a rien de particulier à signaler, les prépositions françaises ont leur équivalent en allemand.

5.3.1.7.18. La place

à	être à sa place remettre qn à sa place	an seinem Platz sein jdn an seinen Platz verweisen	an + D an + A
---	---	--	------------------

Il n'y a rien de particulier à signaler, la préposition française a son équivalent en allemand.

5.3.1.7.19. La poche

à	mettre la main à la poche	in die Tasche greifen	in + A
dans	tenir qc dans sa poche mettre qc dans sa poche	etwas in der Tasche haben etwas in die Tasche stecken	in + D in + A

Quelque soit la préposition française, la préposition allemande est *in*.

5.3.1.7.20. La chaîne, les fers

à	mettre qn aux fers mettre à la chaîne (chien)	jdn in Fesseln legen an die Kette legen	in + D pl an + A
---	---	--	---------------------

La préposition *à* se traduira pas *in* (+ datif) si elle suivie d'un mot pluriel, et par *an* (+ accusatif) si elle est suivie d'un mot singulier.

5.3.1.7.21. L'air

à	au grand air	in frischer Luft	in + D
en	en l'air regarder en l'air sauter en l'air	in die Höhe in die Luft gucken in die Höhe springen	in + A

Quelque soit la préposition française, la préposition allemande sera *in* suivie du cas adéquat.

5.3.1.7.22. Les points cardinaux

à	au nord	im Norden nach Norden	in + D nach + D
à	au sud	im Süden nach Süden	in + D nach + D
à	à l'est	im Osten nach Osten	in + D nach + D
à	à l'ouest	im Westen nach Westen	in + D nach + D
français : à + point cardinal allemand : in (+ datif) + point cardinal (locatif), nach (+ datif) + point cardinal (locatif)			

5.3.1.7.23. Cas particuliers

à	à l'abri, à couvert	unter Dach und Fach (fig)	unter
à	à l'abri d'un arbre	unter einem Baum	unter + D
à	mettre à l'abri	unter Dach und Fach bringen	unter
à	être à l'index	auf der schwarzen Liste stehen	auf + D
à	à 10 km à la ronde	auf 10 km (im) Umkreis	auf
dans	être plongé dans les livres	über den Büchern sitzen	über + D
de...à	du berceau à la tombe	von der Wiege bis zur Bahre	von...bis zu
dans	jeter un coup d'œil dans un livre	einen Blick in ein Buch tun	in + A
sur	mettre sur le papier	zu Papier bringen	zu
sur	cela n'existe que sur papier	das steht nur auf dem Papier	auf + D
à	à sa place	am rechten Ort	an + D
à	à l'endroit cité	am angegebenen Ort	an + D

Il semble préférable d'insérer ces exemples tels quels dans le dictionnaire, car ils correspondent à des expressions toutes faites. Ils ne sont pas traités dans notre étude.

5.3.1.7.24 Divers

Ces exemples ne sont pas pris en compte dans notre travail, car il semble difficile d'établir des règles générales, il faudrait apparemment écrire une règle par exemple, ce qui serait trop laborieux.

à	à la surface de	an / auf der Oberfläche	an / auf + D
à	tenir au jour / à contre-jour	ans / gegen das Licht halten	an / gegen + A
à	pendre au clou	an den Nagel hängen	an + A
à	suspendre à un clou	an einem Nagel aufhängen	an + D
à	aller au bal	auf den Ball gehen	auf + A
à	monter à cheval	aufs Pferd steigen	auf + A
à	porter à la scène	auf die Bühne bringen	auf + A
à	tomber à la renverse	auf den Rücken fallen	auf + A
à	au bureau	auf dem Büro	auf + D
à	tomber au champ d'honneur	auf dem Felde der Ehre fallen	auf + D
à	faire un trou à qc	ein Loch in etwas reißen	in
à	se mettre à couvert	in Deckung gehen (mil)	in
à	ramener à l'étable	in den Stall treiben	in + A
à	à la boutonnière	im Knopfloch	in + D
à	à la ligne	in einer Reihe	in + D
à	prendre à domicile	vom Hause abholen	von

5. Implémentation

après	courir après qc	nach etwas jagen	nach
après	l'un après l'autre	einer nach dem anderen	nach

au travers de	se jeter au travers de qc	sich einer Sache in den Weg stellen	in + A
---------------	---------------------------	-------------------------------------	--------

autour de	tourner autour de qc	um etwas laufen	um
autour de	s'enrouler autour de qc	sich um etwas winden	um

contre	élever une barrière contre qc	gegen etwas eine Schranke errichten	gegen
--------	-------------------------------	-------------------------------------	-------

dans	vivre dans la lune	auf dem Mond leben	auf + D
dans	dans un volume restreint	auf knappem Raum (fig)	auf + D
dans	insérer qc dans qc	etw in etw einarbeiten	in
dans	pénétrer dans qc	in etw fahren	in
dans	tomber dans un profond sommeil	in Schlaf sinken	in
dans	se faufiler dans qc	sich in etwas stehlen	in
dans	envelopper dans qc	in etwas wickeln	in
dans	monter dans le train	in den Zug einsteigen	in + A
dans	tomber dans un piège	in eine Falle geraten	in + A
dans	couper dans le vif	ins Fleisch schneiden	in + A
dans	verser dans le fossé	in den Graben fahren	in + A
dans	tomber / attirer dans une embuscade	in einen Hinterhalt geraten / locken	in + A
dans	s'inscrire dans la courbe	in die Kurve gehen	in + A
dans	attirer dans un piège	in eine Falle locken	in + A
dans	se mettre dans un bourbier	in einen Sumpf geraten	in + A
dans	tomber dans un guêpier	in ein Wespennest greifen (fig)	in + A
dans	se rouler dans une couverture	sich in eine Decke wickeln	in + A
dans	remettre qn dans ses droits	jdn in seine Rechte wiedereinsetzen	in + A
dans	dans le livre, le roman, le film	im Buch, Roman, Film	in + D
dans	chercher son salut dans la fuite	sein Heil in der Flucht suchen	in + D
dans	être pris dans l'engrenage	im Getriebe stecken (fig)	in + D
dans	rester dans l'ombre	im Hintergrund bleiben (fig)	in + D
dans	être dans les douleurs	in den Wehen liegen	in + D
dans	chercher dans tous les coins et recoins	in allen Ecken und Winckeln suchen	in + D
dans	laisser qn dans le doute	jdn im Zweifel lassen	in + D
dans	aller dans la lune	zum Mond fliegen	zu

5. Implémentation

de	sortir de qc	aus etw fahren	aus
de	retirer de la circulation	aus dem Umlauf ziehen	aus
de	de devant, de derrière, d'en haut, d'en bas, d'ici, de là, d'au-delà, de droite, de gauche	von vorn, von hinten, von oben, von unten, von hier, von dort, von drüben, von rechts, von links	von

en	avoir en magasin	auf Lager halten	auf
en	mettre en magasin	auf Lager nehmen	auf
en	en visite	auf / zu Besuch	auf / zu
en	en voyage	auf der Reise	auf + D
en	(aller) en voyage	auf Reisen / (auf die Reisen gehen)	auf + D / (A)
en	aller en vacances	in die Ferien gehen	in + A
en	mettre en selle	in den Sattel heben	in + A
en	aller en enfer	in die Hölle kommen	in + A
en	mettre en jeu	in Spiel bringen	in + A
en	être en lieu sûr	im sicheren Hafen sein (fig)	in + D
en	être bien en selle	fest im Sattel sitzen	in + D
en, à	en vol, à la volée	im Fluge	in + D

entre	s'interposer entre les combattants	sich zwischen die Streitenden drängen	zwischen + A
-------	------------------------------------	---------------------------------------	--------------

hors de	évasion hors de la réalité	Flucht in die Unwirklichkeit	in + A
---------	----------------------------	------------------------------	--------

par	envoyer par le fond	in den Grund bohren	in + A
par	par monts et par vaux	über Berg und Tal	über

sous	prendre qn sous son aile	jdn unter seine Fittiche nehmen (fig)	unter + A
sous	avoir qn sous sa coupe	jdn unter der Fuchtel haben	unter + D
sous	être sous les drapeaux	im Wehrdienst stehen	in + D

sur	avoir un fil sur son manteau	eine Fussel am Mantel haben	an + D
sur	jeter un regard (curieux) sur	einen (neugierigen) Blick werfen auf	auf
sur	mettre sur le feu	aufs Gas setzen	auf + A
sur	rejeter la faute sur qn	die Schuld auf jdn schieben	auf + A
sur	miser sur un cheval	auf ein Pferd setzen (course)	auf + A
sur	tout miser sur un coup	alles auf einen Wurf setzen	auf + A

sur	sur l'ensemble du front	auf der ganzen Front	auf + D
sur	être sur des charbons ardents	auf heißen Kohlen stehen	auf + D
sur	s'endormir sur ses lauriers	auf seinen Lorbeeren ausruhen (fig)	auf + D
sur	être sur le côté	auf der Seite liegen	auf + D
sur	se balancer sur les vagues	sich auf den Wellen wiegen	auf + D
sur	graver sur bois	in Holz schneiden	in
sur	jeter de l'huile sur le feu	Öl ins Feuer gießen (fig)	in + A
sur	noter sur son journal	ins Journal eintragen	in + A
sur	être sur le flanc (fam)	im Bett liegen müssen	in + D
sur	rouler sur l'or	im Golde schwimmen	in + D
sur	tenir sur les fonts baptismaux	über die Taufbecken halten aus der Taufbecken heben	über + A auf + D
sur	étendre sur le carreau	zu Boden strecken	zu
sur	monter sur les planches	zum / ans Theater gehen	zu / an + A

vers	diriger ses pas vers qc	seine Schritte nach etwas wenden	nach
------	-------------------------	----------------------------------	------

5.3.1.8. CONCLUSION

Cette étude sur les prépositions spatiales nous amène à constater qu'il est très rare de n'avoir qu'une traduction par préposition. Les deux tableaux ci-dessous récapitulent toutes les traductions possibles pour chaque préposition ; le premier a le français comme langue source, le second l'allemand. Les phrases données en exemple sont issues du corpus utilisé pour l'implémentation.

5.3.1.8.1. Tableau français

à	an	Paul frappe à la porte = Paul klopft an die Tür
	auf	Paul est à la poste = Paul ist auf der Post
	aus	Paul boit à la bouteille = Paul trinkt aus der Flasche
	bei	Paul saisit le chien au collet = Paul packt den Hund bei dem Kragen
	in	Paul est à Paris = Paul ist in Paris
	nach	Paul va à Paris = Paul fährt nach Paris
	um	Paul tourne à un coin = Paul *herumläuft ¹ um eine Ecke
	unter	Paul se mêle aux coiffeurs = Paul mischt sich unter die Frisoer
	vor	Le bateau est à l'ancre = Das Schiff liegt vor Anker
	zu	Paul va à l'école = Paul geht zur Schule
après	nach	Paul est après le coiffeur = Paul ist nach dem Frisör
auprès de	dicht bei	Paul est auprès du poele = Paul ist dicht bei dem Ofen
	zu	Paul se rend auprès d'un malade = Paul begibt sich zu einem Kranken

¹ Le signe * indique la présence d'une erreur grammaticale : en effet, le programme n'a pas été conçu pour conjuguer les verbes à particule séparable en allemand, par conséquent tous les verbes seront conjugués selon le modèle traditionnel (la particule restera donc « collée » au mot).

5. Implémentation

autour de	um ... herum	Le chien court autour de la maison = Der Hund läuft um das Haus
chez	bei zu	Paul est chez le coiffeur = Paul ist bei dem Frisör Paul va chez le coiffeur = Paul geht zu dem Frisör
contre	an gegen	Paul jette un verre contre le mur = Paul wirft ein Glas an die Wand Paul nage contre le courant = Paul schwimmt gegen den Strom
à côté de	neben	Le verre est à côté de la tasse = Das Glas ist neben der Tasse
dans	an auf aus bei durch hinter in nach von zu	Paul est dans un endroit = Paul ist an einem Ort Paul est dans la rue = Paul ist auf der Strasse Paul boit dans un verre = Paul trinkt aus einem Glas Paul lit dans Molière = Paul liest bei Molière Paul passe dans Paris = Paul fährt durch Paris Paul est dans les coulisses = Paul ist hinter den Kulissen Le verre est dans l'armoire = Das Glas ist in dem Schrank Paul va dans le Tyrol = Paul fährt nach Tyrol Paul mange dans une assiette = Paul ißt von einem Teller Paul va dans la lune = Paul fliegt zu dem Mond
de	aus auf von von ... her	Paul expulse le coiffeur de la ville = Paul weist den Frisör aus der Stadt Paul est de ce côté = Paul ist auf dieser Seite Paul sort de la piste = Paul *abbringt von der Spur Le vent souffle du sud = Der Wind weht vom Süden her
au début de	am Anfang + G	Paul est au début de la rue = Paul ist am Anfang der Strasse
en deçà de	diesseits	Paul est en deçà de la frontière = Paul ist diesseits der Grenze
en dedans de	innerhalb + G innerhalb von + D	La tasse est en dedans de l'armoire = Die Tasse ist innerhalb des Schrankes La tasse est en dedans de l'armoire = Die Tasse ist innerhalb von dem Schrank
en dehors de	außerhalb + G außerhalb von + D	La tasse est en dehors de l'armoire = Die Tasse ist ausserhalb des Schrankes La tasse est en dehors de l'armoire = Die Tasse ist ausserhalb von dem Schrank
au delà de	jenseits + G	Paul est au delà de la frontière = Paul ist jenseits der Grenze
derrière	hinter	Le verre est derrière la tasse = Das Glas ist hinter der Tasse
en dessous de	unterhalb + G unterhalb von + D	Le verre est en dessous de l'armoire = Das Glas ist unterhalb des Schrankes Le verre est en dessous de l'armoire = Das Glas ist unterhalb von dem Schrankes
au dessus de	oberhalb + G oberhalb von + D	Le verre est au dessus de l'armoire = Das Glas ist oberhalb des Schrankes Le verre est au dessus de l'armoire = Das Glas ist oberhalb von dem Schrankes

devant	vor	Paul est devant la maison = Paul ist vor dem Haus
à droite de	rechts von	Le verre est à droite de la tasse = Das Glas ist rechts von der Tasse
en	an auf in nach	Le brouillard est en plusieurs endroits = Der Nebel ist an mehreren Stellen Paul est en mer = Paul ist auf See Paul est en Autriche = Paul ist in Österreich Paul va en Autriche = Paul fährt nach Österreich
à l'encontre de	D entgegen	Le coiffeur vient à l'encontre de Paul = Der Frisör kommt Paul entgegen
entre	in unter (gens) zwischen	Le verre est entre ses mains = Das Glas ist in seinen Händen Ils sont entre amis = Sie sind unter Freunden Paul lit entre les lignes = Paul liest zwischen den Zeilen
face à	angesichts + G	Paul est face à la maison = Paul ist angesichts des Hauses
en face de	gegenüber + D	Paul est en face de l'école = Paul ist gegenüber der Schule
à gauche de	links von	Paul est à gauche de la maison = Paul ist links von dem Haus
en haut de	oben auf oben in + D	Le verre est en haut de l'armoire = Das Glas ist oben auf dem Schrank Le verre est en haut de l'armoire = Das Glas ist oben in dem Schrank
hors de	außerhalb + G außerhalb von + D	Paul est hors de la maison = Paul ist ausserhalb des Hauses Paul est hors de la maison = Paul ist ausserhalb von dem Haus
jusque	bis an bis auf bis nach bis zu	Paul va jusqu'à la frontière = Paul fährt bis an die grenze Paul va jusqu'à la poste = Paul geht bis auf die Post Paul va jusqu'à Paris = Paul fährt bis nach Paris Paul va jusqu'à l'école = Paul geht bis zur Schule
loin de	weit von	Paul est loin de Paris = Paul ist weit von Paris
non loin de	unfern + G unfern von + D unweit + G unweit von + D	Paul est non loin de l'école = Paul ist unfern der Schule Paul est non loin de l'école = Paul ist unfern von der Schule Paul est non loin de l'école = Paul ist unweit der Schule Paul est non loin de l'école = Paul ist unweit von der Schule
le long de	(an+ D) entlang A entlang längs + G	Paul court le long du fleuve = Paul läuft an dem Fluss entlang Paul court le long du fleuve = Paul läuft den Fluss entlang Paul court le long du fleuve = Paul läuft längs des Flusses
par	aus durch in über	Paul saute par la fenêtre = Paul springt aus dem Fenster Paul regarde par la fenêtre = Paul sieht durch das Fenster Le bateau est par dix m de fond = Das Schiff ist in zehn M Tiefe Paul passe par Paris = Paul reist über Paris
parmi	aus	Paul choisit un endroit parmi cent = Paul wählt ein Ort aus Hunderten

par dessus	über	Paul saute par dessus la barrière = Paul *hintüberspringt über die Schranke
pour	auf	Paul part pour la campagne = Paul fährt auf das Land
	in	Paul part pour la guerre = Paul zieht in den Krieg
	nach	Paul part pour Paris = Paul reist nach Paris
près de	in der Nähe von	Paul est près de Paris = Paul ist in der Nähe von Paris
proche de	nahe + D	L'école est proche de la poste = Die Schule ist nahe der Post
sous	bei	Paul a la lettre sous la main = Paul hat den Brief bei der Hand
	hinter	Paul est sous les verrous = Paul ist hinter Schoss und Riegel
	in	Paul met la lettre sous enveloppe = Paul steckt den Brief in einen Umschlag
	unter	Le verre est sous l'armoire = Das Glas ist unter dem Schrank
sur	auf (loc)	Le verre est sur l'armoire = Das Glas ist auf dem Schrank
	bei	Paul a la lettre sur lui = Paul hat den Brief bei ihm
	nach	Paul tire sur le coiffeur = Paul schießt nach dem Friseur
	über	Le brouillard s'étend sur la vallée = Der Nebel breitet sich über das Tal
à travers	durch	Paul regarde à travers ses lunettes = Paul sieht durch seine Brille
	durch ... hindurch	Le bateau navigue à travers le brouillard = Das Schiff fährt durch den Nebel hindurch
en travers de	quer durch	La barrière est en travers de la rue = Die Schranke ist quer durch die Straße
vers	in Richtung auf	Paul va vers Paris = Paul fährt in Richtung auf Paris
	nach	Paul va vers Paris = Paul fährt nach Paris
via	über	Paul passe via Paris = Paul fährt über Paris
	via	Paul passe via Paris = Paul fährt via Paris
vis-à-vis de	gegenüber + D	L'école est vis-à-vis de la poste = Die Schule ist gegenüber der Post
en vue de	angesichts +G	Paul est en vue de la maison = Paul ist angesichts des Hauses

TABLEAU 24: CORRESPONDANCE ENTRE LES PREPOSITIONS FRANÇAISES ET ALLEMANDES

5.3.1.8.2. Tableau allemand

(an+ D) entlang	le long de	Paul court le long du fleuve = Paul läuft an dem Fluss entlang
am Anfang + G	au début de	Paul est au début de la rue = Paul ist am Anfang der Strasse
an	à	Paul frappe à la porte = Paul klopft an die Tür
	contre	Paul jette un verre contre le mur = Paul wirft ein Glas an die Wand
	dans	Paul est dans un endroit = Paul ist an einem Ort
	en	Le brouillard est en plusieurs endroits = Der Nebel ist an mehreren Stellen

5. Implémentation

angesichts + G	face à en vue de	Paul est face à la maison = Paul ist angesichts des Hauses Paul est en vue de la maison = Paul ist angesichts des Hauses
auf	à dans de en pour sur	Paul est à la poste = Paul ist auf der Post Paul est dans la rue = Paul ist auf der Strasse Paul est de ce côté = Paul ist auf dieser Seite Paul est en mer = Paul ist auf See Paul part pour la campagne = Paul fährt auf das Land Le verre est sur l'armoire = Das Glas ist auf dem Schrank
aus	à dans de par	Paul boit à la bouteille = Paul trinkt aus der Flasche Paul boit dans un verre = Paul trinkt aus einem Glas Paul expulse le coiffeur de la ville = Paul weist den Frisör aus der Stadt Paul saute par la fenêtre = Paul springt aus dem Fenster
außerhalb + G	en dehors de	La tasse est en dehors de l'armoire = Die Tasse ist ausserhalb des Schrankes
außerhalb + G	hors de	Paul est hors de la maison = Paul ist ausserhalb des Hauses
außerhalb + D	von en dehors de hors de	La tasse est en dehors de l'armoire = Die Tasse ist ausserhalb von dem Schrank Paul est hors de la maison = Paul ist ausserhalb von dem Haus
bei	à chez dans sous sur	Paul saisit le chien au collet = Paul packt den Hund bei dem Kragen Paul est chez le coiffeur = Paul ist bei dem Frisör Paul lit dans Molière = Paul liest bei Molière Paul a la lettre sous la main = Paul hat den Brief bei der Hand Paul a la lettre sur lui = Paul hat den Brief bei ihm
bis an bis auf bis nach bis zu	jusque jusque jusque jusque	Paul va jusqu'à la frontière = Paul fährt bis an die grenze Paul va jusqu'à la poste = Paul geht bis auf die Post Paul va jusqu'à Paris = Paul fährt bis nach Paris Paul va jusqu'à l'école = Paul geht bis zur Schule
dicht bei	auprès de	Paul est auprès du poele = Paul ist dicht bei dem Ofen
diessseits	en deçà de	Paul est en deçà de la frontière = Paul ist diessseits der Grenze
durch	dans par à travers	Paul passe dans Paris = Paul fährt durch Paris Paul regarde par la fenêtre = Paul sieht durch das Fenster Paul regarde à travers ses lunettes = Paul sieht durch seine Brille
D entgegen	à l'encontre de	Le coiffeur vient à l'encontre de Paul = Der Frisör kommt Paul entgegen
A entlang	le long de	Paul court le long du fleuve = Paul läuft den Fluss entlang
gegen	contre	Paul nage contre le courant = Paul schwimmt gegen den Strom
gegenüber + D	en face de vis-à-vis de	Paul est en face de l'école = Paul ist gegenüber der Schule L'école est vis-à-vis de la poste = Die Schule ist gegenüber der Post
hinter	dans derrière sous	Paul est dans les coulisses = Paul ist hinter den Kulissen Le verre est derrière la tasse = Das Glas ist hinter der Tasse Paul est sous les verrous = Paul ist hinter Schoss und Riegel

5. Implémentation

in	à dans en entre par pour sous	Paul est à Paris = Paul ist in Paris Le verre est dans l'armoire = Das Glas ist in dem Schrank Paul est en Autriche = Paul ist in Österreich Le verre est entre ses mains = Das Glas ist in seinen Händen Le bateau est par dix m de fond = Das Schiff ist in zehn M Tiefe Paul part pour la guerre = Paul zieht in den Krieg Paul met la lettre sous enveloppe = Paul steckt den Brief in einen Umschlag
in der Nähe von	Nähe près de	Paul est près de Paris = Paul ist in der Nähe von Paris
in Richtung auf	vers	Paul va vers Paris = Paul fährt in Richtung auf Paris
innerhalb + G	en dedans de	La tasse est en dedans de l'armoire = Die Tasse ist innerhalb des Schrankes
innerhalb von + D	en dedans de	La tasse est en dedans de l'armoire = Die Tasse ist innerhalb von dem Schrank
jenseits + G	au delà de	Paul est au delà de la frontière = Paul ist jenseits der Grenze
längs + G	le long de	Paul court le long du fleuve = Paul läuft längs des Flusses
links von	à gauche de	Paul est à gauche de la maison = Paul ist links von dem Haus
nach	à après dans en pour sur vers	Paul va à Paris = Paul fährt nach Paris Paul est après le coiffeur = Paul ist nach dem Frisör Paul va dans le Tyrol = Paul fährt nach Tyrol Paul va en Autriche = Paul fährt nach Österreich Paul part pour Paris = Paul reist nach Paris Paul tire sur le coiffeur = Paul schießt nach dem Frisoer Paul va vers Paris = Paul fährt nach Paris
nahe + D	proche de	L'école est proche de la poste = Die Schule ist nahe der Post
neben	à côté de	Le verre est à côté de la tasse = Das Glas ist neben der Tasse
oben auf	en haut de	Le verre est en haut de l'armoire = Das Glas ist oben auf dem Schrank
oben in + D	en haut de	Le verre est en haut de l'armoire = Das Glas ist oben in dem Schrank
oberhalb + G	au dessus de	Le verre est au dessus de l'armoire = Das Glas ist oberhalb des Schrankes
oberhalb von + D	au dessus de	Le verre est au dessus de l'armoire = Das Glas ist oberhalb von dem Schrankes
quer durch	en travers de	La barrière est en travers de la rue = Die Schranke ist quer durch die Straße
rechts von	à droite de	Le verre est à droite de la tasse = Das Glas ist rechts von der Tasse
über	par par dessus sur via	Paul passe par Paris = Paul fährt über Paris Paul saute par dessus la barrière = Paul *hinüberspringt über die Schranke Le brouillard s'étend sur la vallée = Der Nebel breitet sich über das Tal Paul passe via Paris = Paul fährt über Paris

um	à	Paul tourne à un coin = Paul *herumläuft ¹ um eine Ecke
um ... herum	autour de	Le chien court autour de la maison = Der Hund läuft um das Haus
unfern + G	non loin de	Paul est non loin de l'école = Paul ist unfern der Schule
unfern von + D	non loin de	Paul est non loin de l'école = Paul ist unfern von der Schule
unter	à parmi sous entre	Paul se mêle aux coiffeurs = Paul mischt sich unter die Frisoer Paul est parmi les coiffeurs = Paul ist unter den Frisoern Le verre est sous l'armoire = Das Glas ist unter dem Schrank Ils sont entre amis = Sie sind unter Freunden
unterhalb + G	en dessous de	Le verre est en dessous de l'armoire = Das Glas ist unterhalb des Schrankes
unterhalb von + D		Le verre est en dessous de l'armoire = Das Glas ist unterhalb von dem Schrankes
unweit + G	non loin de	Paul est non loin de l'école = Paul ist unweit der Schule
unweit von + D	non loin de	Paul est non loin de l'école = Paul ist unweit von der Schule
via	via	Paul passe via Paris = Paul fährt via Paris
von	dans de	Paul mange dans une assiette = Paul ißt von einem Teller Paul sort de la piste = Paul *abbringt von der Spur
von ... her	de	Le vent souffle du sud = Der Wind weht vom Süden her
vor	à devant	Le bateau est à l'ancre = Der Schiff liegt vor Anker Paul est devant la maison = Paul ist vor dem Haus
weit von	loin de	Paul est loin de Paris = Paul ist weit von Paris
zu	à auprès de chez dans	Paul va à l'école = Paul geht zur Schule Paul se rend auprès d'un malade = Paul begibt sich zu einem Kranken Paul va chez le coiffeur = Paul geht zu dem Frisör Paul va dans la lune = Paul fliegt zu dem Mond
zwischen	entre	Paul lit entre les lignes = Paul liest zwischen den Zeilen

TABLEAU 25 CORRESPONDANCE ENTRE LES PREPOSITIONS ALLEMANDES ET FRANÇAISES

5.3.2. Le programme

Cette étude s'inscrit dans le cadre de travaux de traduction automatique (TA) menés à l'*Institut für Angewandte Informationsforschung* (IAI) de Sarrebruck. En concertation avec l'IAI, nous avons décidé d'élaborer un programme indépendant, qui, après modification pourra être intégré au programme général. Il nous a semblé plus judicieux

¹ Le signe * indique la présence d'une erreur grammaticale : en effet, le programme n'a pas été conçu pour conjuguer les verbes à particule séparable en allemand, par conséquent tous les verbes seront conjugués selon le modèle traditionnel (la particule restera donc « collée » au mot).

de ne pas essayer de travailler directement sur le programme de Sarrebruck qui est en constante évolution, ce qui aurait généré beaucoup de problèmes d'adaptation.

Ce programme constitue un petit système de traduction. Il a été conçu pour la traduction des prépositions spatiales et ne s'intéresse pas aux autres problèmes linguistiques. C'est notamment pour cette raison que les verbes à particules séparables, en allemand, seront conjugués comme s'ils étaient des verbes simples, le programme n'étant pas établi pour un découpage morphologique. Il fonctionne pour un corpus de phrases-tests bien précis, puisque les entrées lexicales sont en nombre limité, et les règles syntaxiques ont été écrites pour des phrases simples. L'intérêt d'utiliser des phrases-tests est de limiter les règles de traduction aux phénomènes linguistiques du sujet. L'étude de ces phrases-test est un moyen pour nous d'identifier les contraintes linguistiques qu'il faut prendre en compte.

Les commentaires sont précédés par un ou plusieurs symboles de pourcentage (%).

5.3.2.1. LA STRUCTURE MORPHOLOGIQUE

5.3.2.1.1. Le dictionnaire morphologique

5.3.2.1.1.1. Les règles b

La structure morphologique se compose essentiellement de règles b aussi bien en allemand qu'en français. Elles auront quasiment la même forme dans les deux langues. Les traits qui n'apparaîtront que dans l'une des langues seront notés en italique. On pourra repérer les niveaux morphologiques grâce à :

@level(msfr/morph/french). pour le français et,
@level(msde/morph/german). pour l'allemand.

Dans notre cas, les dictionnaires morphologiques sont des fichiers séparés du programme. Les commandes

@include(dicofrsb). pour le français, et
@include(dicodesb). pour l'allemand

indiquent au compilateur de règles qu'il faut lire les données enregistrées dans les fichiers sus-nommés. Ces données sont interprétées de la même façon que si elles étaient notées dans le programme à la place de la directive @include.

Nous allons décrire les différentes entrées morphologiques par catégorie et donner un exemple pour illustrer cette description. Certains traits apparaîtront dans toutes les catégories mais ne seront décrits qu'une seule fois, il suffira de se reporter à la première occurrence de ce trait pour tout renseignement complémentaire. Les noms des traits et leurs valeurs ont été définis pour ce programme mais peuvent être modifiés ou complétés pour s'adapter à un autre programme.

Les dictionnaires morphologiques français et allemand se trouvent en annexe A et B.

5.3.2.1.1.1.1. La ponctuation

La règle est de la forme ci-dessous :

```
punct = {cat=punct,lex=LEX,string=STRING,mode=MODE}.[ ]
```

- **cat** représente la catégorie du mot et prend la valeur **punct** pour indiquer qu'il s'agit d'un signe de ponctuation,
- **lex** correspond à l'unité lexicale de l'entrée,
- **string** est la réalisation morphologique du lexème de l'entrée. Le signe de ponctuation est écrit entre apostrophes car il s'agit d'un caractère non alphanumérique,
- **mode** définit le mode de la proposition (**decl** = déclaratif).

ex : **punct** = {**cat**=**punct**,**lex**=**end**,**string**='.',**mode**=**decl**}.[]

5.3.2.1.1.1.2. Les déterminants

En ce qui concerne les déterminants, l'entrée morphologique a une forme un peu particulière : en effet, *lex* aura toujours la valeur *d* car les déterminants ne seront pas traduits d'une langue à l'autre. Ils seront supprimés, dans la langue source, entre les niveaux CS et IS puis régénérés, dans la langue cible, entre les niveaux IS et CS. La raison de ce processus est très simple, certains noms sont employés avec un déterminant en français et pas en allemand et inversement, donc le supprimant au niveau IS, on sera à même de le régénérer dans la langue cible, si besoin est. Dans ce type de règle, la partie située à gauche du symbole & sera invariable, alors que la partie droite variera en fonction des besoins. L'entrée morphologique sera donc de la forme suivante :

```
det = {lex=d,cat=det} &
      ({string=STRING,type=TYPE,agr={num=NUM,gen=GEN},case=CASE};
      ({string=STRING,type=TYPE,agr={num=NUM,gen=GEN},case=CASE}).[ ]
```

On aura autant de lignes que de « strings » différentes.

- **cat** correspond à la catégorie du mot et la valeur *det* au déterminant,
- **lex** correspond à l'unité lexicale de l'entrée et sert de clé de recherche dans le dictionnaire. Ce trait doit être présent dans chaque entrée lexicale avec une valeur atomique,
- **string** est la réalisation morphologique du lexème de l'entrée (ex. **string**=une) ; ce trait est censé être présent dans le dictionnaire morphologique. Sa valeur doit être atomique,
- **type** est utilisé pour définir le type du déterminant : **type**=**def** (défini), et **type**=**indef** (indéfini).
- **agr** peut contenir les attributs **num**, **gen** :
 - * **num** indique le nombre morpho-syntaxique avec **num**=**sing** pour le singulier et **num**=**plu** pour le pluriel,

- * **gen** indique le genre morpho-syntaxique avec **gen=fem** pour le féminin, et **gen=masc** pour le masculin,
- **case** indique le marquage morpho-syntaxique du cas et peut prendre quatre valeurs : **nom** = nominatif, **acc** = accusatif, **gen** = génitif, **dat** = datif. Ce trait est uniquement utilisé en allemand.

ex : **einen** = {**cat=det,lex=ein,string=einen,def=no,agr={num=sing,gen=masc},case=acc**}.[]

5.3.2.1.1.1.3. Les prépositions

L'entrée morphologique de la préposition est la plus complexe car c'est elle qui contient le plus d'informations :

prep = {**cat=p,lex=LEX,pos=POS,string=STRING,onset=ONSET,**
head={ cat=p,string=STRING,
ehhead={ cat=n,pform=PFORM,case=CASE,
pvalue={'T'=space,space=SPACE}}}}].[]

- **cat**, **lex**, **string** et **onset** ont la même signification que dans les paragraphes précédents, à la différence que **lex** n'aura pas une valeur constante,
- **pos** précise la position du constituant à l'intérieur de la structure de branches binaire. **pos=left** est identifié avec la branche gauche de la structure binaire et **pos=right** avec la branche droite. Ce trait peut être spécifié dans l'entrée lexicale pour faire la distinction entre préposition et postposition, notamment en allemand,
- **head** contient les traits sémantiques et syntaxiques :
 - **cat**, c'est-à-dire la partie de discours de l'entrée (**p** = préposition et postposition),
 - **ehhead** contient l'ensemble des traits de tête d'une structure donnée. Par définition, les traits **ehhead** sont un sous-ensemble du trait **HEAD**. Les traits de tête étendu sont généralement :
- * **cat**, c'est-à-dire la partie de discours de la projection étendue. **cat=n** pour les noms, déterminants, classificateurs et préposition, **cat=v** pour les verbes, complémenteurs, virgule et signe de ponctuation final d'une phrase, **cat=a** pour les adjectifs et les degrés, **cat=adv** pour les adverbes et les degrés.
- * **pform** indique la préposition du groupe prépositionnel, abstraction faite de la réalisation morphologique de celle-ci. Ainsi en allemand, les prépositions *an*, *am* et *ans* ont la même **pform** **pform=an**. Il en est de même en français, *à*, *au*, *aux* ont la même **pform** **pform=a**.
- * **case** indique le marquage morpho-syntaxique du cas des noms, déterminants et prépositions. À côté du marquage casuel de surface des prépositions contractées qui peut sans doute admettre que les prépositions soient conçues comme étant casuellement marquées, on ne peut pas trouver de support supplémentaire, ni de contre-indication pour le marquage casuel des prépositions.

- * **pvalue** est utilisé pour décrire la relation de modifiant entre une tête syntaxique et son modifieur. Cette valeur est contrôlée lors du transfert et est responsable de la génération correcte des prépositions et des complémenteurs des phrases. On distingue 6 types principaux de modification, mais on ne décrira, ici, que les prépositions spatiales. Il faut toutefois préciser que, dans notre cas, ce trait n'est pas utilisé, mais il sera utile quand ce programme sera intégré à un programme plus développé.

{'T'=space,space=_} est spécifié par la conjonction des attributs EXT, DYN et DIST (distance). Dist=coin si l'objet coïncide avec le point de référence, sinon, la distance est spécifié par *near* ou *far*.

- **pvalue={'T'=space,ext=_,space={dyn=_,dist={dist=_}}}**
 - ext=point : in Paris / à Paris
 - ext=path : über Paris /par Paris
 - dyn=state : in Paris / à Paris
 - dyn=dir (= direction) : nach Paris / à Paris
- **pvalue={'T'=space,space={dyn=_,dist={dist=coin,with=_}}}**
 - dist={dist=coin (coïncide)}
 - with=spot : am Auto / à la voiture
 - with=line : auf der Landstraße /sur la route nationale
 - with=square : auf dem Tisch /sur la table
 - with=cube : im Auto / dans la voiture
- **pvalue={'T'=space,space={dyn=_,dist={dist=(near;far),position=_}}}**
 - dist={dist=(near;far),position=hor,hor=facing} : vor dem Haus / devant la maison
 - dist={dist=(near;far),position=hor,hor=behind} : hinter dem Haus / derrière la maison
 - dist={dist=(near;far),position=hor,hor=left} : links vom Haus / du côté gauche de la maison
 - dist={dist=(near;far),position=hor,hor=right} : rechts vom Haus / du côté droit de la maison
 - dist={dist=(near;far),position=vert,vert=above} : über dem Haus / au dessus de la maison
 - dist={dist=(near;far),position=vert,vert=below} : unter dem Haus / en dessous de la maison
 - dist={dist=(near;far),position=center,center=line} : den Fluss entlang / le long du fleuve
 - dist={dist=(near;far),position=center,center=square} : um das Haus herum / autour de la maison
 - dist={dist=(near;far),position=center,center=cube} : ausserhalb des Hauses / en dehors de la maison
 - dist={dist=(near;far),position=between} : zwischen den Häusern / entre les maisons

ex : sous = {cat=p,role=gov,lex=sous,pos=left,onset=cons,string=sous,
 head={cat=p,string=sous,
 ehead={cat=n,pform=sous,
 pvalue={'T'=space,space={dyn=prov,
 dist={dist=near,position=vert,vert=below}}}}}.[]

5.3.2.1.1.1.4. Les noms

nom = { cat=n,lex=LEX,string=STRING,onset=ONSET,
agr={gen=GEN,num=NUM},compound=COMPOUND,
type=TYPE,case=CASE}.[]

- **cat**, **lex**, **string**, **agr**, **case** et **onset** ont la même signification que dans les paragraphes précédents,
- **type** représente le type de « catégorie » selon le schéma suivant :

type=abs	noms qui ne sont pas précédés d'un article
type=indef	noms précédés d'un article indéfini
type=def	noms précédés d'un article défini
type=defabs	noms précédés d'un déterminant défini mais utilisé sans déterminant
type=poss	noms précédés d'un article possessif
type=dem	noms précédés d'un article démonstratif

Aucune de ces valeurs n'est transférée dans la langue cible, puisque ces spécifications syntaxiques sont spécifiques aux catégories et sont susceptibles de changer pendant le transfert,

- **compound** est un trait exclusivement français qui prend les valeurs **compound=yes** si le terme est un mot composé et **compound=no** si le terme est un mot simple. Ce trait a un rôle important pour le choix de la préposition avec les noms de régions et de départements notamment.

ex : paris = {cat=n,lex=paris,string='Paris',onset=cons,
agr={gen=fem,num=sing},compound=no,type=abs}.[]

5.3.2.1.1.1.5. Les verbes

verbe = { cat=v,lex=LEX,string=STRING,agr={num=sing,per=3},
stative=STATIVE,extract=EXTRACT,enter=ENTER,
tense={tns=pres}}.[]

- **cat**, **lex**, et **string** ont la même signification que dans les paragraphes précédents,
- **agr** peut contenir les attributs num, per :
 - * **num** indique le nombre morpho-syntaxique avec **num=sing** pour le singulier et **num=plu** pour le pluriel,
 - * **per** indique la personne morpho-syntaxique des noms, déterminants et prépositions, avec **per=3** pour la troisième personne, **per=2** pour la deuxième, **per=1** pour la première,
- **stative** caractérise le verbe tel que **stative=yes** et **stative=no**. La classe des verbes statiques est spécifique à une langue ; un verbe peut être statique dans une langue, mais dynamique dans une autre. Toutefois dans les exemples qui nous concernent; il y a homogénéité entre les verbes français et allemands, par conséquent, ce trait est contrôlé lors du transfert,

- **extract** caractérise plus l'ensemble de la phrase que le verbe seul. Ce trait est fondamental puisqu'il va permettre de traduire des prépositions antonymes telles que *dans* et *aus* grâce notamment à **extract=yes** qui donne une indication d'extraction,
- **enter** précise si le verbe indique ou non un mouvement de pénétration,
- **tense** indique le marquage morpho-syntaxique du temps.

ex : `habiter = {cat=v,lex=habiter,string=habite,agr={num=sing,per=3},
stative=yes,extract=no,enter=no,tense={tns=pres}}.[]`.

5.3.2.1.1.2. Les règles f

Il n'y en a qu'une. Elle assigne la troisième personne à tous les noms et tous les déterminants.

5.3.2.1.2. L'analyse morphologique française

Pour montrer le déroulement de l'analyse morphologique et de toute la traduction nous prendrons comme exemple la phrase : *Paul est chez le coiffeur*, nous donnerons le détail puis le résultat de l'analyse.

Dans notre cas, nous aurons l'analyse suivante :

```
msfr:b: paul ....ENTRY FOUND.... >
msfr:f: per3 ok >
msfr:b: <Paul> analysing...>
msfr:b: <paul> analysing...>
msfr:b: e3tre ....ENTRY FOUND.... >
msfr:b: <est> analysing...>
msfr:b: chez ....ENTRY FOUND.... >
msfr:b: <chez> analysing...>
msfr:b: det ....ENTRY FOUND.... >
msfr:f: per3 ok >
msfr:b: <le> analysing...>
msfr:b: coiffeur ....ENTRY FOUND.... >
msfr:f: per3 ok >
msfr:b: <coiffeur> analysing...>
msfr:b: punct ....ENTRY FOUND.... >
msfr:b: <.> analysing ...>
Created object : msfr/1
```

Le parser prend les mots l'un après l'autre et vérifie leur présence dans le dictionnaire grâce aux règles b. A chaque fois qu'il a repéré un nom ou un déterminant, il applique la règle f per3 qui stipule que tous les noms et déterminants sont à la troisième personne.

Lorsque l'analyse est terminée le parser crée un objet au niveau MSFR qui est de la forme :

```
1 : Paul
2 : est
3 : chez
4 : le
5 : coiffeur
6 : .
```

et dont la liste de structures de traits va servir d'input au parser. Elle se présente ainsi :

History : msfr/1

WORD# 1

```
{ string = 'Paul',
  cat = n,
  lex= paul,
  agr = { gen = masc,
         num = sing,
         per = '3'}}
```

WORD# 2

```
{ string = est,
  cat = v,
  lex= e3tre,
  agr = { num = sing,
         per = 3},
  stative = yes,
  extract = no,
  enter = no,
  tense = {tns = pres}}
```

WORD# 3

```
{ string = chez,
  cat = p,
  lex= chez,
  pos = left,
  onset = cons,
  head = { cat = p,
          string = chez,
          ehead = { cat = n,
                   pform=chez,
                   pvalue = { T = space,
                              space = { dyn=1068,
                                         dyn ~= prov,
                                         dist = { dist = near,
                                                  position = center,
                                                  center = spot}}}}}}}
```

WORD# 4

```
{ string = le,
  cat = det,
  lex= d,
  type = def,
  agr = { gen = masc,
         num = sing,
         per = '3'}}
```

WORD# 5

```
{ string = coiffeur,
  cat = n,
  lex= coiffeur,
  onset = cons,
  type=def,
  agr = { gen = masc,
         num = sing,
         per = '3'}}
```

WORD# 6

```
{ string = '.',
  cat = punct,
  lex= end,
  mode = decl}
```

5.3.2.2. LES DICTIONNAIRES COMMUNS

Il en existe un pour chaque langue. Ils sont utilisés dans les niveaux CS et IS. Comme ils sont séparés du programme, ils sont consultés grâce aux directives :

```
@call(fr_lexicon). pour le français et,
@call(de_lexicon). pour l'allemand.
```

5.3.2.2.1. Les déterminants et les prépositions

Les entrées des déterminants et des prépositions sont identiques. On ne précise que le role, la valeur de lex et la catégorie.

```
ex : le = {role=gov,cat=det,lex=le}.[ ].
      sur = {role=gov,cat=p,lex=sur}.[ ].
```

5.3.2.2.2. Les noms

Les entrées des noms ont la même forme que celles des déterminants mais on ajoute des informations sémantiques. En ce qui concerne ces informations, on utilise la hiérarchie des types sémantiques, utilisés pour distinguer les entités, proposée par Cornélia Zelinsky-Wibbelt. (cf tableau p. XX), ainsi que le dictionnaire qu'Heinz Dieter Maas a composé dans le cadre de son étude morphologique mais qui permet d'utiliser des éléments qui n'apparaissent pas dans la classification de sa collègue. Les traits sémantiques seront partagés par les deux langues mais, dans certains cas, certains seront plus importants dans l'une ou l'autre langue.

- **sem** contient les traits sémantiques. Le système de traits sémantiques est essentiellement basé sur le système développé par Zelinsky-Wibbelt¹ [Zelinsky-Wibbelt 88], [Zelinsky-Wibbelt 89], qui a pour but une classification sémantique des noms. Son implémentation actuelle ne représente toutefois qu'un sous-ensemble de la structure de traits proposée par Zelinsky-Wibbelt, étendue à tous les éléments lexicaux. Une troisième différence par rapport au système d'origine développé par Zelinsky-Wibbelt réside dans le fait que les

¹ Cf. infra.

valeurs sémantiques ne sont pas des valeurs sémantiques lexicales pures, toutes assignées dans le dictionnaire et modifiées si nécessaire à des niveaux supérieurs de la projection. Une entrée lexicale peut être sous spécifiée par rapport au trait sémantique, mais cette information manquante peut être remplie par le contexte, ou, en l'absence d'un tel contexte, par un mécanisme par défaut.

- **anim** est (**nil**;{'T'=hum,sex=(male;female;nil)}), i.e. inanimé ou spécifié par la valeur de T. Ces entités peuvent être codées comme animées, qui peuvent être adressées verbalement pour changer une situation donnée (cf. [Lausberg 76], p. 11).
- **gran** est (**nil**;unique) où nil fait référence aux noms communs et unique aux noms propres.

Comme toute distinction sémantique, la valeur gran est strictement contrôlée au cours du transfert, évitant qu'un nom commun ne soit traduit par un nom propre et vice versa. Cette distinction unique/nil est également pertinente pour le composant monolingue, puisque les déterminants remplissent plusieurs fonctions selon qu'ils précèdent un nom commun ou un nom propre. Dans de nombreux cas le déterminant est obligatoire pour les noms propres et ne peut pas remplir de fonction sémantique. Une troisième valeur est 'part', pour les noms qui dénotent leur fonction ou leur existence comme partie d'un ensemble plus large (cf. [Zelinsky-Wibbelt 89]).

- **abstract** est (**nil**;{temp=TEMP}), i.e. il n'y a pas de valeur abstract ou la valeur de temp doit être spécifié ultérieurement. Un concept doit être classé comme abstract, s'il fait référence au temps (i.e. s'il a une valeur temporelle), s'il fonctionne comme prédicat ou si le concept n'a pas de réalisation matérielle.
- **concrete** spécifie des entités que l'on peut sentir grâce au toucher, ou qui font partie d'un ensemble de ces entités. À côté de cette perception tactile, on peut percevoir les choses de façon visuelle, olfactive ou auditive, mais aucun de ces derniers sens n'autorise, par lui-même, la description d'une entité comme concrète s'elle ne peut être touchée en même temps. Une entité que l'on peut sentir par le toucher peut également être explorée par un autre sens, sans que l'inverse soit vrai. La position centrale du sens tactile peut être expliquée par le degré assez élevé de corroboration du savoir transmis par ce sens. Pour tous les autres sens, une telle classification est plus difficile. Il n'y a pas de moyen de dire comment et quand il est possible de voir que « Hans lit un livre » ou de sentir que « Jean prépare un gâteau ».

Les mots qui font référence à une classe doivent être codés comme concrets comme ses éléments, puisque par la relation d'hyponymie, tout mot super-ordonné entre dans un (*SLOT*) d'un mot sous-ordonné (par ex. *Sie jagen Löwen, Sie jagen Säugetiere*).

La plupart des descriptions et implémentations linguistiques adoptent seulement une dimension sémantique qui caractérise l'entité exclusivement comme abstraite ou comme concrète. Selon cette stratégie, la valeur des

entités qui peuvent être abstraites ou concrètes n'est pas spécifiée à gauche, si bien qu'elle s'unifie avec les deux valeurs, abstrait et concret. La spécification sur la dimension 'abstrait' ou « concret » que nous utilisons peut donc sembler redondante. Mais pour les entités qui peuvent être abstraites et concrètes en même temps, un codage exclusif en abstrait ou concret peut conduire à des valeurs conflictuelles, e.g. quand les deux valeurs sont sous-catégorisées, comme dans *The book that Chomsky wrote on syntax is lying on the table*. Dans cette phrase le verbe *write...on...* sous-catégorise une entité abstraite (par ex. essai, poème, histoire). Le verbe *lies* + localisation concrète sous-catégorise une entité concrète. Les traits sémantiques du nom *livre* qui devient concret en étant sous-catégorisé par *lie* et le pronom relatif qui devient abstrait en étant sous-catégorisé par le verbe *write* doivent toutefois s'unifier pour permettre l'adjonction d'une proposition relative. Si un tel contexte déclenche des valeurs conflictuelles, on ne peut pas ajouter la proposition relative. Toutefois, si les deux dimensions 'abstrait' et 'concret' sont codées indépendamment de ces entités sémiotiques, la description sémantique peut s'adapter à différentes demandes.

Les objets sémiotiques, c'est-à-dire livre, lettre, image, ont toujours cette double nature, d'une part le contenu abstrait, d'autre part le support d'un contenu sémiotique, qui peut être abstrait ou concret comme dans les exemples précédents. Dans notre système cette double classification des entités selon la dimension 'abstrait' et la dimension 'concret' est importante dans ce sens que de nombreuses entités dans le monde informatique possèdent ce double statut, parmi lesquels on trouve fichier, répertoire, disque, cartouche, programme, mémoire tampon, pointeur, enregistrement.

- **abstract={temp=nil}** décrit les entités qui ont une signification abstraite mais non temporelle. Comme la signification temporelle implique la nature abstraite de l'entité, nous avons codé temp comme une sous-spécification de abstract. Comme exemples d'entités codées **abstract={temp=nil}**, on peut citer les institutions (e.g. la Commission, le Parlement), les objets sémiotiques (e.g. mémo, adresse), les régions cognitives (e.g. agriculture, mathématiques).
- **bound** est {'T'=quant} ou {'T'=cumul}, c'est-à-dire quantifié ou cumulé. Comme pour l'extension temporelle et l'abstrait, on suppose que les limites conceptuelles forment une catégorie applicable à toutes les parties du discours (cf. la théorie de Krifka sur l'homomorphisme des objets et des événements ([Krifka 91])). Nous adoptons la définition des limites conceptuelles de Quine (cf. [Quine 60]), selon laquelle les concepts font référence à un cumul si aucune somme du concept correspond au concept lui-même, par ex. *any sum of parts which are water is water* (p. 91). Sinon les concepts font référence à la quantification. Les noms communs utilisés au singulier et au pluriel sans article font référence au cumul (au moins en anglais). Les noms communs utilisés au singulier et au pluriel avec un article ou un autre quantifieur font référence à la quantification.

Une première version de la sémantique de CAT2 utilisait ici la distinction bound / mass proposée par Zelinsky-Wibbelt (cf. [Zelinsky-Wibbelt 89]).

Deux raisons nous ont toutefois fait préférer la distinction quantifié / cumulé. Premièrement, presque tous les noms de masse peuvent être mis au pluriel, ou apparaître avec un article ou un numéral si on parle d'unité de service (une bière, deux cafés), d'un type (une bière que j'ai eue en Allemagne), ou d'une instanciation (une guerre, deux guerres) (cf. [Bunt 85]). Ce point de vue a mené les sémanticiens, comme cela s'est produit pour l'Aktionsart, à se concentrer sur la nature compositionnelle de cette distinction sémantique, séparant la valeur de la classification lexicale selon la distinction mass/count (cf. [Verkuyt 72]). Deuxièmement, la distinction mass/count comme classification lexicale n'est pas pertinente pour le transfert. De nombreux exemples montrent que, les noms 'mass' peuvent être traduits par des noms 'count' et vice versa. Les exemples suivants sont tirés de [Mufwene 84] : fr-en : *nouvelle(s)* - *news*, *renseignement(s)* - *informations*, *conseil(s)* - *advice*, *connaissance(s)* - *knowledge*, *preuve(s)* - *evidence*, *habit(s)* - *clothes/clothings*. La distinction quantifié/cumulé doit être contrôlée lors du transfert afin d'assurer l'équivalence de traduction. Les concepts 'cumulé' doivent être traduits par des concepts 'cumulé' et les concepts 'quantifié' par des concepts 'quantifié'. Un avantage supplémentaire de la distinction quantifié/cumulé réside dans le fait qu'elle peut être étendue à d'autres catégories telles que les adjectifs et les verbes ayant un nombre d'implications aspectuelles intéressant.

{'T'=quant} peut être spécifié par quant=def ou quant=indef, selon que l'entité a subi une quantification définie ou indéfinie. Une quantification indéfinie peut être du type indef=many, indef=few, alors qu'une quantification définie peut être du type def=all ou def=1, def=2, def=3...

Nous proposons, page suivante, la figure récapitulative des traits sémantiques de C. Zelinsky-Wibbelt.

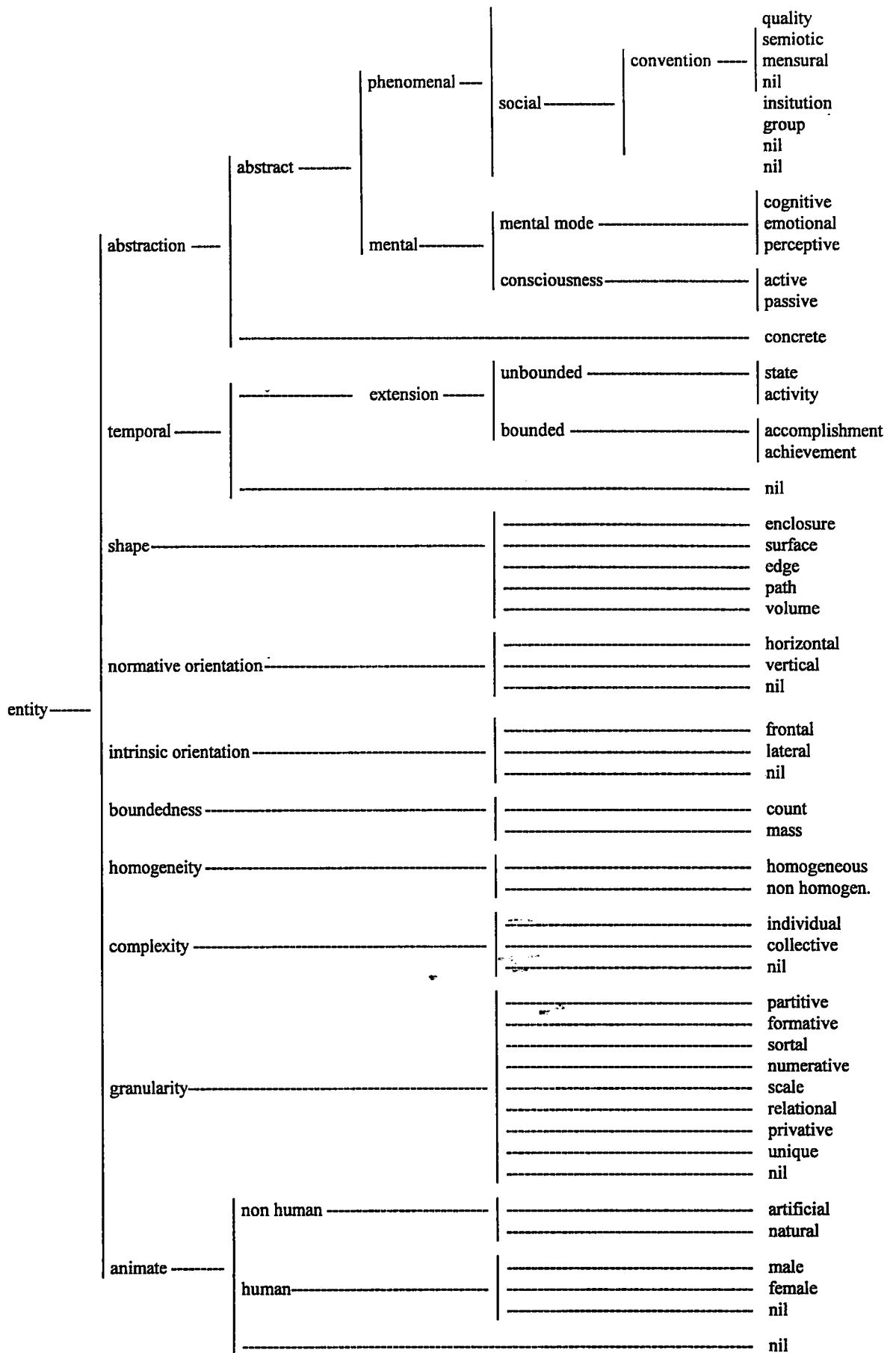


FIGURE 19 : LES TRAITS SEMANTIQUES PROPOSES PAR C. ZELINSKY-WIBBELT

Le dictionnaire ci-dessous a été composé par Heinz Dieter Maas dans le cadre de son analyse morphologique. Les traits qu'il utilise nous seront très utiles dans notre étude.

[hyper=agent]	agent
[hyper=agent,agent=ano]	membre d'un groupe
[hyper=agent,agent=anto]	auteur (d'une action)
[hyper=agent,agent=estro]	chef, maître
[hyper=agent,agent=ido]	parent (mère, fils...)
[hyper=agent,agent=isto]	profession
[hyper=agent,agent=ulo]	caractérisation grâce à des qualités
[hyper=animal]	animal
[hyper=animal,edible=no]	animal non comestible
[hyper=animal,edible=yes]	animal comestible
[hyper=box]	contenant, récipient
[hyper=coll]	collectif
[hyper=coll, coll=ano]	groupe de personnes
[hyper=coll, coll=isto]	groupe de travail
[hyper=coll, coll=ulo]	groupe de personnes ayant les mêmes caractéristiques
[hyper=disease]	maladie
[hyper=displace]	déplacement
[hyper=domain]	domaine abstrait
[hyper=domain,domain=isto]	domaine professionnel
[hyper=event]	événement
[hyper=family]	nom de famille
[hyper=fname]	prénom
[hyper=form]	forme
[hyper=game]	jeu
[hyper=gesetz]	loi, principe
[hyper=instrum]	instrument
[hyper=instr,complex=no]	outil simple
[hyper=instr,complex=yes]	outil complexe
[hyper=instr,instr=body]	partie du corps
[hyper=language]	langue
[hyper=loc]	locatif
[hyper=loc,loc=ag]	institution, commerce
[hyper=loc,loc=berg]	montagne
[hyper=loc,loc=berge]	groupe de montagnes
[hyper=loc,loc=bezirk]	zone, région
[hyper=loc,loc=city]	nom de ville
[hyper=loc,loc=country]	nom de pays
[hyper=loc,loc=departement]	département
[hyper=loc,loc=erdteil]	continent
[hyper=loc,loc=flaechе]	surface
[hyper=loc,loc=flaechе,up=an]	surface, préposition usuelle = an
[hyper=loc,loc=flaechе,up=auf]	surface, préposition usuelle = auf
[hyper=loc,loc=fluss]	fleuve
[hyper=loc,loc=gebaeude]	bâtiment
[hyper=loc,loc=gend]	nord, sud, est, ouest
[hyper=loc,loc=gelaende]	territoire

[hyper=loc,loc=insel]	île
[hyper=loc,loc=limit]	limite
[hyper=loc,loc=other]	autre
[hyper=loc,loc=rand]	bordure
[hyper=loc,loc=rand,up=an]	bordure, préposition usuelle = an
[hyper=loc,loc=rand,up=auf]	bordure, préposition usuelle = auf
[hyper=loc,loc=see]	mer, lac, océan
[hyper=loc,loc=way]	chemin
[hyper=loc,loc=wueste]	désert
[hyper=loc,loc=zimmer]	pièce
[hyper=material]	matériel / matière (indénombrable)
[hyper=material,edible=no]	matériel / matière non comestible
[hyper=material,edible=yes]	matériel / matière comestible
[hyper=medien]	radio, télévision
[hyper=money]	unité monétaire
[hyper=move]	déplacement
[hyper=phenomenon]	phénomène naturel (pluie...)
[hyper=plant]	plante, légume
[hyper=plant,edible=no]	plante non comestible
[hyper=plant,edible=yes]	plante comestible
[hyper=process]	processus, action
[hyper=quality]	qualité (humaine)
[hyper=quant]	quantité
[hyper=result]	résultat
[hyper=set]	ensemble
[hyper=set,set=instr]	ensemble (non humain)
[hyper=state]	état (au sens large)
[hyper=text]	loi, principe (mais moins juridique que gesetz)
[hyper=thing]	chose
[hyper=thing,edible=no]	chose non comestible
[hyper=thing,edible=yes]	chose comestible
[hyper=thing,thing=body]	partie du corps
[hyper=thing,thing=kleid]	vêtement
[hyper=thing,thing=moebel]	meuble
[hyper=time]	période
[hyper=time,time=duration]	durée
[hyper=time,time=month]	mois
[hyper=vehicle]	véhicule

ex : bouteille = {role=gov,cat=n,lex=bouteille,sem={ shape=enclosure,
hyper={hyper=box,open=small}}}.[.].

5.3.2.2.3. Les verbes

Les entrées des verbes seront identiques à celles des noms auxquelles on ajoute le trait **frame** qui contient la structure argumentale de l'entrée où arg1 représente le sujet, arg2 le groupe prépositionnel ou le complément d'objet direct s'il y a un arg3 qui représente alors le groupe prépositionnel. Pour chacun des arguments, on précise le rôle qu'il occupe : agent pour le sujet animé, nonagent pour le sujet non animé, patient pour le complément d'objet direct et loc pour le groupe prépositionnel.

ex : aller = {role=gov,cat=v,lex=aller,stative=no,extract=no,enter=no,
 frame={arg1={role=agent,cat=n,sem={anim=yes}},
 arg2={role=loc,stative=no}}}.[.].
 mettre = {role=gov,cat=v,lex=mettre,stative=no,extract=no,enter=no,
 frame={arg1={role=agent,cat=n,sem={anim=yes}},
 arg2={role=patient},
 arg3={role=loc,stative=no}}}.[.].

5.3.2.3. LA STRUCTURE SYNTAXIQUE FRANÇAISE

Par rapport à l'ancienne version de Cat2, un niveau a été rajouté entre ms et cs. Il permet de définir la structure des noms composés et des prépositions contractées grâce aux règles mw (multi-words). Par exemple, la règle au qui est de la forme :

au = {{string=au}} <==> [{lex=a2,string=a2,cat=p},
 {lex=d,string=le,cat=det,type=def,agr={gen=masc,num=sing}}].

stipule que la chaîne de caractères *au* se décompose en *a2* (préposition) et *le* (déterminant).

Le niveau syntaxique commence à partir de la directive @level(csfr/syntactic/french).

5.3.2.3.1. Les règles du niveau cs

Tout comme pour le niveau morphologique, les règles b sont plus nombreuses que les règles f.

5.3.2.3.1.1. Les règles b

Elles sont au nombre de huit et décrivent la structure de la phrase.

text = {cat=text,stative=S,extract=E,enter=P}.[{{cat=s,tense~=nil,stative=S,extract=E,
 enter=P,mode=M},{cat=punct,mode=M}}].
 s = {cat=s,stative=S,extract=E,enter=P}.[{{cat=np},{cat=vp,stative=S,extract=E,enter=P}}].
 np1 = {cat=np,agr=A}.[{{cat=n,agr=A,type=abs}}].
 np2 = {cat=np,agr=A}.[{{cat=det,agr=A,type=T},{cat=n,agr=A,type=T,type~=abs}}].
 vp1 = {cat=vp,agr=A,tense=T,stative=S,extract=E,enter=P}.[{{cat=v,agr=A,
 stative=S,extract=E,enter=P},{cat=pp,stative=S,extract=E,enter=P}}].
 vp2 = {cat=vp,agr=A,tense=T,stative=S,extract=E,enter=P}.[{{cat=v,agr=A,
 stative=S,extract=E,enter=P},{cat=np},{cat=pp,stative=S,extract=E,enter=P}}].
 pp = {cat=pp}.[{{cat=p},{cat=np}}].
 p_p = {cat=pp,stative=S,extract=E,enter=P}.[{{cat=p},{cat=pp,stative=S,extract=E,enter=P}}].

La règle text stipule qu'un texte se compose de phrases et de signes de ponctuation qui définissent le mode de la proposition (dans notre cas, il s'agira toujours du mode déclaratif). Elle précise également que les phrases doivent avoir un verbe conjugué (puisque tense~=nil signifie que le temps ne peut pas être égal à nil), ainsi que les traits *stative*, *extract* et *enter*.

La règle s décompose la phrase en un groupe nominal (np) et un groupe verbal (vp).

Nous avons ensuite deux types de règles qui peuvent s'appliquer aux groupes nominaux. La première np1 établit qu'un np se compose uniquement d'un nom, alors

que la seconde np2 impose la présence d'un déterminant dans le np et la même valeur pour le trait type du déterminant et du nom.

De la même façon nous avons également deux règles pour les groupes verbaux. La première vp1 décompose le vp en un verbe et un groupe prépositionnel (pp), et la deuxième vp2 le décompose en un verbe, un np et un pp. Dans tous les cas, les traits *stative*, *extract* et *enter* présents dans le vp, le seront au niveau du verbe ainsi que dans le pp et auront les mêmes valeurs.

Enfin nous utilisons deux règles pour les groupes prépositionnels : la règle pp scinde le groupe prépositionnel en une préposition et un groupe nominal et la règle p_p le scinde en une préposition suivie d'un groupe prépositionnel.

5.3.2.3.1.2. Les règles f

```
np_vpAgr = {cat=s,tense=T}.[cat=np,agr=A].>>{agr=A,tense=T,*].
declarative = {cat=s}>>{mode=decl}.[cat=np,*].
det_abs = {cat=np}>>{det=abs}.[+{cat~def}].
```

Trois règles f sont définies. La première np_vpAgr vérifie l'accord entre le groupe nominal et le groupe verbal, et la seconde declarative précise que si une phrase est déclarative, alors elle commence obligatoirement par un groupe nominal quels que soient les éléments qui le suivent, enfin la troisième det_abs vérifie l'absence du déterminant dans le groupe nominal.

5.3.2.3.1.3. Les règles kill

```
kill_jusque_pour = {cat=pp}.[cat=p,lex=jusque},{cat=pp}.[cat=p,lex=pour]>>{cat=nil,*]
... 1
```

Les règles kill sont des règles assassines. Elles permettent d'éliminer les traductions incorrectes. La règle kill_jusque_pour élimine la traduction d'une phrase si le GP se compose de la préposition *jusque* et d'un groupe prépositionnel dont la préposition est *pour*. Toutes les règles « kill » se lisent de la même façon.

5.3.2.3.2. L'analyse syntaxique

Pour l'analyse syntaxique, le parser utilise la liste de structures de traits résultant de l'analyse morphologique. Les différentes étapes de l'analyse figurent ci-dessous :

```
msfr<=>csfr:t: <Paul> ... copying input ... >
msfr<=>csfr:t: <est> ... copying input ... >
msfr<=>csfr:t: <chez> ... copying input ... >
msfr<=>csfr:t: <le> ... copying input ... >
msfr<=>csfr:t: <coiffeur> ... copying input ... >
msfr<=>csfr:t: <.> ... copying input ... >
fr_lexicon:b: paul FOUND >
csfr:b: np1 FOUND >
csfr:f: det_abs ok >
fr_lexicon:b: etre FOUND >
```

¹ ... remplace les règles suivantes. Elles se trouvent toutes en annexe E. Il en sera de même à chaque fois que l'on rencontrera ce symbole.

```

fr_lexicon:b: chez FOUND >
fr_lexicon:b: det FOUND >
fr_lexicon:b: coiffeur FOUND >
csfr:b: np2 FOUND >
csfr:b: pp FOUND >
csfr:b: vp1 FOUND >
csfr:b: s FOUND >
csfr:f: np_vpAgr ok >
csfr:f: declarative ok >
fr_lexicon:b: punct FOUND >
csfr:b: text FOUND >
Created object : csfr/1

```

Pour commencer, le parser copie la liste des structures de traits. Puis il commence à analyser le premier élément *paul*. La première règle b qui soit applicable est la règle b atomique *paul*, issue du niveau COMMON FR-LEXICON appelé au niveau CS. On obtient alors un sous-objet composé du seul mot *paul*. A ce moment, la règle *np1* qui construit un groupe nominal (*np*) à partir d'un nom commun sans déterminant peut s'appliquer puis suit la règle *det_abs* qui vérifie l'absence du déterminant dans le *np*, et comme aucune autre règle n'est applicable, l'analyse de cet élément est complète.

Le second élément est ensuite scanné. Seule une règle b du niveau commun est utilisée, aucune règle f n'est applicable, il en résulte un deuxième sous-objet pour le mot *être*. Il en sera de même pour les troisième, quatrième mots, à savoir *chez*, *le* et *coiffeur*. A chaque fois, un nouveau sous-objet est créé.

A partir de ce moment, le parser repère différentes règles b non atomiques qui peuvent être applicables. La première d'entre elles est la règle *np2* qui construit un groupe nominal avec un déterminant et un nom. Aucune règle f n'est applicable, donc le *np d coiffeur* est complet. Il n'y a plus de règle b pour les nps, l'analyse de cet élément est par conséquent terminée.

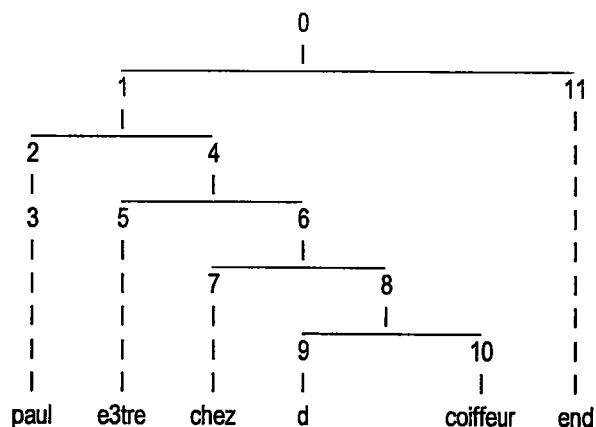
Le parser poursuit son travail et utilise la règle b *pp* définie pour les groupes prépositionnels et qui les décompose en une préposition suivie d'un groupe nominal. Il ne peut appliquer aucune règle f, ni aucune autre règle b pour les pps ; le groupe prépositionnel *chez d coiffeur* est complet, l'analyse de cet élément est achevée.

La règle b *vp1* qui scinde un groupe verbal en un verbe et un groupe prépositionnel est applicable. Le sous-objet *être chez d coiffeur* est construit. Il n'y a plus de règles b ou f utilisables, l'analyse de cet élément est finie.

L'avant-dernière règle b non atomique *s* est appliquée pour transformer un *np* suivi d'un *vp* en une phrase, qui sera le dernier sous-objet créé et qui aura la forme *paul être chez d coiffeur*. Le parser utilise à cet instant les deux règles f définies au niveau CS, la première *np_vpAgr* vérifie l'accord du *np* et du *vp* et la seconde *declarative* si la phrase est bien déclarative. Le parser ne peut employer aucune règle supplémentaire, l'analyse de cet élément est terminée.

Enfin, le dernier élément est scanné. La seule règle b qui soit applicable est la règle b atomique *punct*. A ce stade, tous les éléments ont été scannés. Le parser utilise, pour finir, la dernière règle b non atomique *text* qui décompose un texte en une phrase suivie d'un signe de ponctuation.

Comme output, le parser donne une structure arborescente pour text :



La structure de traits qui en résulte est la suivante : (elle sera utilisée pour construire la structure du niveau IS).

History : csfr1 <= msfr/1

```

{ cat = text,
  stative = yes,
  extract = no,
  enter = no}
  .[{ cat = s,
    stative = yes,
    extract = no,
    enter = no,
    tense = {tns = pres},
    mode = decl}
    .[{ cat = np,
      agr = {gen = masc,
        num = sing,
        per = '3'}
      type=abs,
      det=abs}
      .[{ string = 'Paul'
        cat = n,
        lex = paul,
        type=abs,
        agr = {gen = masc,
          num = sing,
          per = '3'}
        role = gov,
        sem = { gran=unique,
          anim = yes}}}],
    { cat = vp,
      agr = {num = sing,
        per = '3',
        gen = masc},
      tense = {tns = pres},

```



```

stative = yes,
extract = no,
enter = no}
  .[{ string = est,
      cat = v,
      lex = e3tre,
      agr = {num = sing,
            per = '3',
            gen = masc},
      stative = yes,
      extract = no,
      enter = no,
      tense = {tns = pres},
      role = gov,
      frame = { arg1=_1277,
                arg1 = ({role = agent,
                        cat = n,
                        sem = {anim = yes}}),
                ; {role = nonagent,
                    cat = n,
                    sem = {anim = no}}),
                arg2 = {role = loc,
                        stative = yes}}}],
{cat = pp,
 stative = yes,
 extract = no,
 enter = no}
  .[{ string = chez,
      cat = p,
      lex = chez,
      pos = left,
      onset = cons,
      head = {cat = p,
              string = chez,
              ehead = {cat = n,
                       pform = chez,
                       pvalue = {T = space,
                                 space = {dyn = _1165,
                                           dyn ~ = prov,
                                           dist = {dist = near,
                                                   position = center,
                                                   center = spot}}}}}],
role = gov},
{cat = np,
 agr = {gen = masc,
        num = sing,
        per = '3'}},
  .[{ string = le,
      cat = det,

```

```

lex = d,
type = def,
agr = {num = sing,
       gen = masc,
       per = '3'},
role = gov},
{ string = coiffeur,
  cat = n,
  lex = coiffeur,
  onset = cons,
  agr = {gen = masc,
         num = sing,
         per = '3'},
  role = gov,
  sem = {anim = yes}}]
{ string = '.',
  cat = punct,
  lex = end,
  mode = decl}]

```

5.3.2.4. TRANSFORMATION

Il y a transformation quand les structures arborescentes sont transformées entre des niveaux de représentation. Pour cela, il faut un traducteur entre les niveaux L_1 et L_2 ($L_1 \Leftrightarrow L_2$) et un générateur pour le niveau L_2 . Il y aura transformation entre les niveaux CSFR et ISFR, ISFR et ISDE, ISDE et CSDE. Dans ces différents niveaux, les règles peuvent se lire et s'appliquer dans les deux sens. Si le français est la langue source, elles se lisent et s'appliquent de gauche à droite, et dans le sens inverse si l'allemand est la langue source.

Le processus de transformation pourrait être décrit ainsi :

1. Unifier l'objet source avec l'objet cible de la règle t , en liant tout marqueur dans la règle t aux sous-objets adéquats à l'intérieur de l'objet.
2. Pour chaque objet lié à un marqueur, retourner à l'étape 1.
3. Construire un objet cible temporaire selon les spécifications précisées du côté cible de la règle t , en incluant tous les sous-objets qui ont été liés du côté source de la règle t et traduits successivement.
4. Localiser les règles b au niveau cible L_2 qui s'unifient avec l'objet cible temporaire.
5. Appliquer toutes les règles tf , dans un ordre séquentiel, dont le côté source s'unifie avec l'objet source et le côté cible avec l'objet cible temporaire.
6. Appliquer, dans un ordre séquentiel, les règles f définies au niveau L_2 , à l'objet cible temporaire. Si toutes les règles f applicables fonctionnent, l'objet temporaire est sauvegardé comme objet permanent et la transformation est complète.

Les règles de transformation entre le niveau CS et le niveau IS sont les suivantes :

5.3.2.4.1. Les règles t

@level(csfr=<=>isfr).

@rule(t).

text = {cat=text, stative=S, extract=E, enter=P}.[s:{tense=T},{lex=end}] <=>

{s:{role=proposition, tense=T, stative=S, extract=E, enter=P}}.

s1 = {cat=s}.[np:{cat=np},{cat=vp}.[v:{cat=v},pp:{cat=pp}]] <=>

{role=proposition}.[v:{cat=v},np:{cat=n},pp:{cat=p}}.

s2 = {cat=s}.[npA:{cat=np},{cat=vp}.[v:{cat=v},npB:{cat=np},pp:{cat=pp}]] <=>

{role=proposition}.[v:{cat=v},npA:{cat=n,role=agent},npB:{cat=n,role=patient},pp:{cat=p}}.

np1 = {cat=np}.[cat=n,lex=L,agr=A,type=abs] <=> {cat=n}.[cat=n,lex=L,agr=A,type=abs].

np2 = {cat=np}.[cat=det,agr=A,type=T],{cat=n,lex=L,agr=A,type=T,type~=abs} <=>

{cat=n}.[cat=n,lex=L,agr=A,type=T,type~=abs].

pp = {cat=pp, stative=S, extract=E, enter=P}.[p:{cat=p},n:{cat=np}] <=>

{role=loc, stative=S, extract=E, enter=P}.[p:{cat=p},n:{cat=n}}.

p_p = {cat=pp, stative=S, extract=E, enter=P}.[p1:{cat=p},p2:{cat=pp, stative=S, extract=E,

enter=P}] <=> {role=loc, stative=S, extract=E, enter=P}.[p1:{cat=p},p2:{role=loc,

stative=S, extract=E, enter=P}}.

atom = {cat~det, cat=C, lex=LU, string=LEX}.[] <=> {cat=C, lex=LU, string=LEX}.[]

La règle text lie le marqueur *s* au sous-objet qui sera une proposition. Les traits *stative*, *extract*, *enter* et *tense* sont transférés d'un niveau à l'autre. Le point est unifié avec la structure vide et n'est lié à aucun marqueur.

Les deux règles suivantes concernent la structure de la phrase. La règle *s1* s'applique aux phrases qui ne comportent pas de complément d'objet direct (COD). Elle lie le groupe nominal (GN) à *np*, le verbe à *v* et le groupe prépositionnel (GP) à *pp* et opère une translation des éléments entre les deux niveaux afin que le verbe passe en première position et soit suivi du GN et du GP. La règle *s2*, quant à elle, s'applique aux phrases qui incluent un COD. Elle lie alors le GN à *npA*, le verbe à *v*, le COD à *npB* et le GP à *pp* et opère les mêmes translations en insérant *npB* entre *npA* et *pp*.

Les règles *np1* et *np2* suppriment le déterminant entre les niveaux CS et IS, que le GN en possède un comme *c*'est le cas en *np2*, ou n'en possède pas comme en *np1*.

La règle *pp* lie le marqueur *p* à la préposition et *n* au GN et assigne le rôle *loc* au GP. La règle *p_p* lie le marqueur *p1* à la préposition et le marqueur *p2* au GP. Les traits *stative*, *extract*, *enter* et *tense* sont transférés d'un niveau à l'autre.

Enfin la règle *atom* transfère la valeur de *cat*, *lex* et *string* de CS à IS sauf pour les déterminants.

5.3.2.4.2. Les règles f

@rule(f).

npAgr = {cat=np}>>{agr=A}.[*] <=> {cat=n}>>{agr=A}.[*].

tense = {cat=v}>>{tense=X}.[*] <=> {cat=v}>>{tense=X}.[*].

Il y en a deux. La première copie les accords du groupe nominal du niveau CS au niveau IS, la seconde copie le temps du verbe.

5.3.2.5. LA STRUCTURE RELATIONNELLE FRANÇAISE

5.3.2.5.1. Les règles du niveau IS

Pour construire la structure relationnelle, le parser va utiliser en combinaison les règles de transformation de CS à IS et les règles du niveau IS. Le niveau relationnel commence à partir de la directive @level(isfr/relational/french).

5.3.2.5.1.1. Les règles b

```
@rule(b).
text = {role=text,stative=S,extract=E,enter=P}.[{role=proposition,cat=v,stative=S,extract=E,enter=P}].
pred = {cat=C,frame={arg1=A1,arg2=A2,arg3=A3}}. [{role=gov,cat=C,
    frame={arg1=A1,arg2=A2,arg3=A3}},^A1,^A2,^A3,^{role=mod}].
pred1 = {cat=C,frame={arg1=A1,arg2=A2,arg3=A3}}. [{role=gov,cat=C,
    frame={arg1=A1,arg2=A2,arg3=A3}},^A1,^A2,^A3,{role=loc,cat~n}].
```

Elles sont au nombre de trois : la première définit la structure du texte, la seconde celle des prédicats, la troisième la structure des prédicats dans un GP composé d'une préposition et d'un GP.

5.3.2.5.1.2. Les règles f

```
@rule(f).
frame1 = {frame={arg1=nil}}.[ ].
frame2 = {frame={arg2=nil}}.[ ].
frame3 = {frame={arg3=nil}}.[ ].
complete1 = {}. [{role=gov,frame={arg1={role=R}}},* ,{}>>{role=R},*].
complete2 = {}. [{role=gov,frame={arg2={role=R}}},* ,{}>>{role=R},*].
complete3 = {}. [{role=gov,frame={arg3={role=R}}},* ,{}>>{role=R},*].
stative1 = {role=loc,stative=yes,extract~yes,enter~yes}. [
    ( {cat=p}>>{lex=a2},
    {cat=n}. [ {cat=n,
        sem={gran=unique,shape=enclosure,
            hyper={hyper=loc,loc=city}}]);
    ( {cat=p}>>{lex=a2},
    {cat=n}. [ {cat=n,sem={shape=surface,
        hyper={hyper=loc,loc=gebiet}}]);
...].1
stative11 = ...
extract1 = ...
enter = ...
```

¹ Pour lire la totalité des règles stative1, stative11, extract1 et enter, se référer au programme en annexe.

Les règles *frame1*, *frame2* et *frame3* n'assignent aucun argument aux noms et aux prépositions.

Les règles *complete1*, *complete2* et *complete3* vérifient le cadre de sous-catégorisation du verbe c'est-à-dire si les rôles assignés dans la description argumentale du verbe sont bien les rôles présents dans les arguments.

Les quatre règles suivantes vérifient l'utilisation de telle ou telle préposition en fonction du type de verbe qui la précède et du type de GN qui la suit. *stative1* s'applique pour les verbes de localisation, *stative11* pour les verbes de mouvement, *extract1* pour les verbes d'« extraction » et *enter* pour les verbes d'« enfoncement ».

5.3.2.5.1.3. Les règles *kill*

killf = {role=proposition}.[cat=v,lex=aller},{role=agent,sem~={anim=yes}}>>{cat=nil},*,*].

Les règles *kill* sont des règles assassines. Elles permettent d'éliminer les traductions incorrectes. La règle *killf*, par exemple, détruit la traduction si le sujet de *aller* n'est pas animé.

5.3.2.5.2. L'analyse relationnelle

Les différentes étapes de cette analyse sont les suivantes :

```

0 $root*1 csfr=>isfr:t:text trying >
0/1 s*1 csfr=>isfr:t:s1 trying >
0/1/1 np*1 csfr=>isfr:t:npl trying >
0/1/1 np*1 csfr=>isfr:t:npl trying to build target object >
fr_lexicon:b:paul FOUND >
isfr:f:frame1 ok >
isfr:f:frame2 ok >
isfr:f:frame3 ok >
isfr:b:pred FOUND >
csfr=>isfr:f:npAgr ok >
0/1/1 np*1 csfr=>isfr:t:npl translation succeeded >
0/1/2/1 v*1 csfr=>isfr:t:atom trying >
0/1/2/1 v*1 csfr=>isfr:t:atom trying to build target object >
fr_lexicon:b:etre FOUND >
csfr<=>isfr:tf:tense ok >
isfr:f:frame3 ok >
0/1/2/1 v*1 csfr=>isfr:t:atom translation succeeded >
0/1/2/2 pp*1 csfr=>isfr:t:pp trying >
0/1/2/2/1 p*1 csfr=>isfr:t:atom trying >
0/1/2/2/1 p*1 csfr=>isfr:t:atom trying to build target object >
fr_lexicon:b:chez FOUND >
isfr:f:frame1 ok >
isfr:f:frame2 ok >
isfr:f:frame3 ok >
0/1/2/2/1 p*1 csfr=>isfr:t:atom translation succeeded >
0/1/2/2/2 n*1 csfr=>isfr:t:np2 trying >
0/1/2/2/2 n*1 csfr=>isfr:t:np2 trying to build target object >
fr_lexicon:b:coiffeur FOUND >
isfr:f:frame1 ok >
isfr:f:frame2 ok >
isfr:f:frame3 ok >
isfr:b:pred FOUND >
csfr=>isfr:f:npAgr ok >
0/1/2/2/2 n*1 csfr=>isfr:t:np2 translation succeeded >
0/1/2/2 pp*1 csfr=>isfr:t:pp trying to build target object >

```

```

isfr:b:pred FOUND >
isfr:f:stative1 ok >
0/1/2/2 pp*1 csfr=>isfr:t:pp translation succeeded >
0/1 s*1 csfr=>isfr:t:s1 trying to build target object >
isfr:b:pred FOUND >
isfr:f:complete1 ok >
isfr:f:complete2 ok >
0/1 s*1 csfr=>isfr:t:s1 translation succeeded >
0 $root*1 csfr=>isfr:t:text trying to build target object >
isfr:b:text FOUND >
0 $root*1 csfr=>isfr:t:text translation succeeded >
Created object : isfr/1

```

1. Pour commencer le parser localise la règle `t text` qui lie le marqueur `s` au sous-objet *Paul est chez le coiffeur* ; le point est unifié avec la description de structure de traits vide `{}` et n'est lié à aucun marqueur.
2. Ensuite il transforme le sous-objet lié à `s` grâce à la règle `t s` lie le sujet *Paul* au marqueur `np`, le verbe *est* au marqueur `v` et le groupe prépositionnel *chez le coiffeur* au marqueur `pp`.

A. Il transforme le sous-objet lié à `np`.

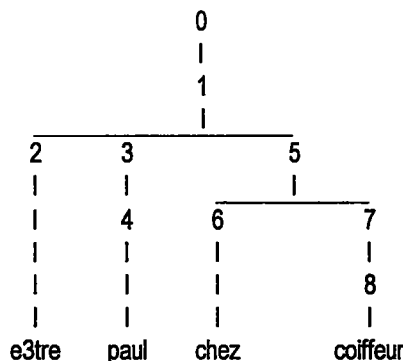
- i. Il localise la règle `np1` qui concerne les groupes nominaux sans déterminant.
- ii. Il repère l'entrée *paul* dans le dictionnaire sémantique français `fr_lexicon`. Il applique les règles `frame1`, `frame2` et `frame3` qui assignent zéro argument aux noms.
- iii. Il situe la règle `pred`. Le nom s'unifie avec le premier constituant et reçoit le trait `{role = gov}`, `cat` prend la valeur `n`, et `A1`, `A2` et `A3` la valeur `nil` comme résultat de l'application des règles `frame`.
- iv. La règle `npAgr` peut s'appliquer puisque l'objet source contient le trait `{cat=np}` et son body au moins un élément (noté `[*]`).
- v. Aucune règle `f` du niveau ISFR ne peut être appliquée, la transformation de l'objet est complète.

B. Il transforme les éléments du groupe verbal.

- i. Il repère le verbe et applique la règle `atom` qui unifie les variables avec la description des traits du verbe. Il localise l'entrée *etre* `fr_lexicon`, vérifie le temps et applique la règle `frame3` qui n'assigne aucun argument, puisque le verbe être ne comporte que deux arguments. Il ne peut plus appliquer d'autres règles, l'analyse de l'élément est terminée.
- ii. Il localise la règle `pp` qui découpe le groupe prépositionnel en une préposition et un groupe nominal. Il lie le marqueur `p` à la préposition et le marqueur `n` au groupe nominal.
 - a) Il repère la préposition et applique comme précédemment la règle `atom`. Il localise ensuite l'entrée *chez* dans `fr_lexicon`. Il applique les règles `frame1`, `frame2` et `frame3` qui n'assignent aucun argument à la préposition. Aucune règle supplémentaire n'est applicable, l'analyse de l'élément est finie.

- b) Il analyse ensuite le deuxième élément, à savoir le groupe nominal. Il applique la règle np2 qui transforme un GN composé d'un déterminant et d'un nom en un GN composé uniquement d'un nom. Il localise l'entrée *coiffeur* dans *fr_lexicon* et applique à nouveau les règles *frame1*, *frame2* et *frame3*. Puis il recommence les étapes 3., 4. et 5.
- C. Il essaie de construire le groupe prépositionnel grâce à la règle *pp* repérée précédemment, celui-ci a le trait {*role=loc*} ainsi que les traits *stative*, *extract* et *enter* correspondants au verbe. Il applique la règle *pred*. Il utilise ensuite la règle *stative1* qui détermine l'utilisation de telle ou telle préposition en fonction du GN qui la suit et du type de verbe qui la précède. Dans notre cas, il vérifie que l'élément qui suit la préposition a bien les traits {*hyper=agent,agent=isto*}. La règle est applicable, le GP est construit. Aucune règle *f* du niveau ISFR ne peut être appliquée, la transformation de l'objet est complète.
3. Un objet temporaire de catégorie *s* est construit, le marqueur *np* est lié au GN sujet, le marqueur *v* au verbe et enfin le marqueur *pp* au GP. Il trouve et applique la règle *pred*, unifiant la variable *C* au verbe, les variables *A1* au premier argument du verbe {*role=agent,cat=n*}, *A2* au deuxième argument {*role=loc,cat=p*}, et *A3* à nil. Le premier sous-objet adjacent au verbe s'unifie avec l'instanciation de *A1*, le second avec celle de *A2* et le troisième avec celle de *A3* et le nœud racine prend la valeur « *cat=v* ». Il n'y a plus de règle *tf* applicable. En revanche, il peut appliquer, au niveau ISFR, les règles *complete1* et *complete2* qui vérifient que les rôles assignés dans la description des arguments du verbe correspondent aux rôles présents à l'intérieur des arguments eux-mêmes. Comme tel est le cas, le sous-objet *e3tre paul chez coiffeur* est complet.
 4. L'objet temporaire pour la catégorie *text* est construit, il se compose de la translation de la structure marquée *s* sous un nœud racine vide. Le trait « *role=proposition* » est unifié avec l'unique nœud-fille. Il localise la règle *b text* au niveau ISFR et l'unifie avec l'objet cible temporaire.
 5. Il ne peut plus appliquer ni de règles *tf* ni de règles *f* au niveau ISFR. Toute la transformation est terminée.

On obtient la structure arborescente ci-dessous,



et la structure de traits correspondante.

History : isfr/1 <= csfr1 <= msfr/1

```

{role = text,
 stative = yes,
 extract = no,
 enter = no}
  .[ {role = proposition,
      cat = v,
      frame = { arg1 = { cat = n,
                        frame = { arg1 = nil,
                                  arg2 = nil,
                                  arg3 = nil},
                        agr = { gen = masc,
                               num = sing,
                               per = '3'},
                        role = _1458,
                        sem = {anim = _1448}},
      arg1 = ({ role = agent,
                cat = n,
                sem = {anim = yes}}
              ; {role = nonagent,
                  cat = n,
                  sem = {anim = no}})},
      arg2 = { role = loc,
                stative = yes,
                extract = no,
                enter = no,
                cat = p,
                frame = { arg1 = nil,
                          arg2 = nil,
                          arg3 = nil}},
      arg3 = nil},
 tense = {tns = pres},
 stative = yes,
 extract = no,
 enter = no}
  .[ {cat = v,
      lex = e3tre,
      string = est,
      role = gov,
      stative = yes,
      extract = no,
      enter = no,
      frame = { arg1 = { cat = n,
                        frame = { arg1 = nil,
                                  arg2 = nil,
                                  arg3 = nil},
                        agr = { gen = masc,
                               num = sing,

```



```

        per = '3'},
        role = _1458,
        sem = {anim = _1448}},
    arg1 = ({ role = agent,
             cat = n,
             sem = {anim = yes}}
           ; {role = nonagent,
             cat = n,
             sem = {anim = no}}),
        role = _1458,
        sem = {anim = _1448}},
    arg2 = { role = loc,
            stative = yes,
            extract = no,
            enter = no,
            cat = p,
            frame = { arg1 = nil,
                    arg2 = nil,
                    arg3 = nil}},
        arg3 = nil}},
    { cat = n,
      frame = { arg1 = nil,
                arg2 = nil,
                arg3 = nil},
      agr = { gen = masc,
             num = sing,
             per = '3'},
      role = _1458,
      sem = {anim = _1448}}
    &
    ({ role = agent,
      cat = n,
      sem = {anim = yes}}
     ; {role = nonagent,
      cat = n,
      sem = {anim = no}})
    .[ { cat = n,
        lex = paul,
        agr = { gen = masc,
              num = sing,
              per = '3'},
        type = abs,
        role = gov,
        sem = { gran = unique,
              anim = yes},
        frame = { arg1 = nil,
                  arg2 = nil,
                  arg3 = nil}},
    { role = loc,

```

```

stative = yes,
extract = no,
enter = no,
cat = p,
frame = { arg1 = nil,
          arg2 = nil,
          arg3 = nil}}
.[{ cat = p,
  lex = chez,
  string = chez,
  role = gov,
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil}},
 { cat = n,
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil},
 agr = { gen = masc,
        num = sing,
        per = '3'},
 role = mod}
.[{ cat = n,
  lex = coiffeur,
  agr = { num = sing,
        gen = masc,
        per = '3'},
  type = def,
  role = gov,
  sem = {hyper = {hyper = agent,
                  agent = isto}},
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil}}]]]]

```

5.3.2.6. LE TRANSFERT

Le niveau IS de la langue source que nous venons d'obtenir est ensuite transféré dans la langue cible. Le niveau de transfert est activé grâce à la directive @level(isfr=>isde).

5.3.2.6.1. Les règles t

Les règles t ont été placées dans un fichier séparé du programme. Elles sont lues grâce à la directive :

```

@level(isfr=>isde).
@include(tradsb).

```

Elles sont de la forme suivante :

```
@rule(t).
cons = {}.[+x] <=> {}.[+x].
unknown = {lex='@unknown',string=L}.[ ] <=> {lex='@unknown',string=L}.[ ].
atom = {lex=paul}.[ ] <=> {lex=paul}.[ ].
atom = {lex=coiffeur}.[ ] <=> {lex=frisoer}.[ ].
...
```

La règle `cons` copie, par défaut, une structure de la langue source vers une structure identique de la langue cible.

La règle `unknown` copie, par défaut, la valeur de `unknown` dans la langue cible. Elle est utilisée quand le parser rencontre des mots inconnus.

Les règles `atom` transfèrent le terme de la langue source vers le terme correspondant de la langue cible.

5.3.2.6.2. Les règles `tf`

```
@rule(tf).
cat = {}>>{cat=C}.[ ] <=> {}>>{cat=C}.[ ].
agr = {cat=n}>>{agr=A}.[*] <=> {cat=n}>>{agr=A}.[*].
tense = {cat=v}>>{tense=T}.[*] <=> {cat=v}>>{tense=T}.[*].
stat = {cat~=(n;p)}>>{stative=S}.[*] <=> {cat~=(n;p)}>>{stative=S}.[*].
state = {role=loc}>>{stative=S}.[*] <=> {role=loc}>>{stative=S}.[*].
extract = {cat~=(n;p)}>>{extract=E}.[*] <=> {cat~=(n;p)}>>{extract=E}.[*].
extracte = {role=loc}>>{extract=E}.[*] <=> {role=loc}>>{extract=E}.[*].
enter = {cat~=(n;p)}>>{enter=P}.[*] <=> {cat~=(n;p)}>>{enter=P}.[*].
entere = {role=loc}>>{enter=P}.[*] <=> {role=loc}>>{enter=P}.[*].
role = {}>>{role=R}.[*] <=> {}>>{role=R}.[*].
```

Elles copient les traits de la langue source vers la langue cible. La première copie la catégorie des atomes, la seconde les traits des arguments s'il y en a, la troisième les traits de temps, les quatrième et cinquième le trait « stative » respectivement pour les catégories autres que `n` et `p` et pour les catégories ayant le trait `{role=loc}`, les sixième et septième le trait « extract », les huitième et neuvième le trait « enter » et enfin la dixième les rôles sémantiques.

5.3.2.6.3. Les règles `kill`

```
kill_dans_hinter = {role=proposition}.[{cat=v},*,*,{cat=p,role=loc}.[{cat=p,lex=dans},{cat=n}.[{cat=n}]
<=> {role=proposition}.[{cat=v},*,*,{cat=p,role=loc}.[{cat=p,lex=hinter},
{cat=n}.[{cat=n,sem~={orient=vertical,concrete=yes}}]]].
...
```

A nouveau, ces règles assassines permettent d'éliminer les traductions incorrectes. La règle `kill_dans_hinter` supprime toutes les traductions de *dans* par *hinter* quand la valeur sémantique du nom qui suit la préposition est différente de `{orient=vertical,concrete=yes}`. Elle sera notamment utilisée pour annuler la traduction

de *Paul ist hinter den Kulissen* par *Paul est derrière les coulisses*. Toutes les règles suivantes se liront suivant le même procédé.

5.3.2.7. LA STRUCTURE RELATIONNELLE ALLEMANDE

5.3.2.7.1. Les règles du niveau isde

Pour construire la structure relationnelle, le parser va utiliser en combinaison les règles de transformation de ISFR à ISDE et les règles du niveau ISDE. Le niveau relationnel commence à partir de la directive `@level(isde/relational/german)`.

5.3.2.7.1.1. Les règles b

Elles sont identiques à celles de la structure française et se lisent par conséquent de la même façon. (cf. p. 258)

5.3.2.7.1.2. Les règles f

Elles sont également identiques à celles de la structure française. (cf. p. 258). Seules les valeurs de lex pour les prépositions dans les règles `stat1`, `stat11`, `extract1` et `enter1` changent et prennent les valeurs allemandes. A part cela, elles se lisent de la même façon.

Les règles `wasser_abs`, `wasser_def...` déterminent les cas dans lesquels le mot « wasser » s'emploie avec ou sans déterminant.

5.3.2.7.2. La synthèse relationnelle allemande

```
0 $root*1 isfr=>isde:t:cons trying >
0/1 x*1 isfr=>isde:t:cons trying >
0/1/1 x*1 isfr=>isde:t:atom trying >
0/1/1/1 x*1 isfr=>isde:t:atom trying to build target object >
de_lexicon:b:sein FOUND >
isfr<=>isde:tf:cat ok >
isfr<=>isde:tf:tense ok >
isfr<=>isde:tf:stat ok >
isfr<=>isde:tf:extract ok >
isfr<=>isde:tf:enter ok >
isfr<=>isde:tf:role ok >
isde:f:frame3 ok >
0/1/1/1 x*1 isfr<=>isde:t:atom translation succeeded >
0/1/2 x*2 isfr<=>isde:t:cons trying >
0/1/2/1 x*1 isfr<=>isde:t:atom trying >
0/1/2/1/1 x*1 isfr<=>isde:t:atom trying to build target object >
de_lexicon:b:paul FOUND >
isfr<=>isde:tf:cat ok >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
isde:f:frame1 ok >
isde:f:frame2 ok >
isde:f:frame3 ok >
0/1/2/1/1 x*1 isfr<=>isde:t:atom translation succeeded >
0/1/2 x*2 isfr<=>isde:t:cons trying to build target object >
isde:b:pred FOUND >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
0/1/2 x*2 isfr<=>isde:t:cons translation succeeded >
```

```

0/1/3 x*3 isfr<=>isde:t:cons trying >
0/1/3/1 x*1 isfr<=>isde:t:atom trying >
0/1/3/1 x*1 isfr<=>isde:t:atom trying to build target object >
de_lexicon:b:bei FOUND >
isfr<=>isde:tf:cat ok >
isfr<=>isde:tf:role ok >
isde:f:frame1 ok >
isde:f:frame2 ok >
isde:f:frame3 ok >
0/1/3/1 x*1 isfr<=>isde:t:atom translation succeeded >
0/1/3/2 x*2 isfr<=>isde:t:cons trying >
0/1/3/2/1 x*2 isfr<=>isde:t:atom trying >
0/1/3/2/1 x*2 isfr<=>isde:t:atom trying to build target object >
de_lexicon:b:frisoer FOUND >
isfr<=>isde:tf:cat ok >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
isde:f:frame1 ok >
isde:f:frame2 ok >
isde:f:frame3 ok >
0/1/3/2/1 x*1 isfr<=>isde:t:atom translation succeeded >
0/1/3/2 x*2 isfr<=>isde:t:cons trying to build target object >
isde:b:pred FOUND >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
0/1/3/2 x*2 isfr<=>isde:t:cons translation succeeded >
0/1/3 x*3 isfr<=>isde:t:cons trying to build target object >
isde:b:pred FOUND >
isfr<=>isde:tf:state ok >
isfr<=>isde:tf:extracte ok >
isfr<=>isde:tf:entere ok >
isfr<=>isde:tf:role ok >
isde:f:stat11 ok >
0/1/3 x*3 isfr<=>:t:cons translation succeeded >
0/1 x*1 isfr<=>isde:t:cons trying to build target object >
isde:b:pred FOUND >
isfr<=>isde:tf:tense ok >
isfr<=>isde:tf:stat ok >
isfr<=>isde:tf:extract ok >
isfr<=>isde:tf:enter ok >
isfr<=>isde:tf:role ok >
isde:f:complete1 ok >
isde:f:complete2 ok >
0/1 x*1 isfr<=>isde:t:cons translation succeeded >
0 $root*1 isfr<=>isde:t:cons trying to build target object >
isde:b:text FOUND >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
0 $root*1 isfr<=>isde:t:cons translation succeeded >
Created object : isde/1

de_lexicon:b:zu FOUND >
isfr<=>isde:tf:cat ok >
isfr<=>isde:tf:role ok >
isde:f:frame1 ok >
isde:f:frame2 ok >
isde:f:frame3 ok >
0/1/3/1 x*1 isfr<=>isde:t:atom translation succeeded >
0/1/3/2 x*2 isfr<=>isde:t:cons trying >
0/1/3/2/1 x*2 isfr<=>isde:t:atom trying >
0/1/3/2/1 x*2 isfr<=>isde:t:atom trying to build target object >
de_lexicon:b:frisoer FOUND >
isfr<=>isde:tf:cat ok >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
isde:f:frame1 ok >

```

```

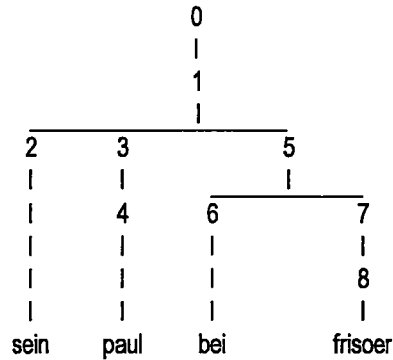
isde:f:frame2 ok >
isde:f:frame3 ok >
0/1/3/2/1 x*1 isfr<=>isde:t:atom translation succeeded >
0/1/3/2 x*2 isfr<=>isde:t:cons trying to build target object >
isde:b:pred FOUND >
isfr<=>isde:tf:agr ok >
isfr<=>isde:tf:role ok >
0/1/3/2 x*2 isfr<=>isde:t:cons translation succeeded >
0/1/3 x*3 isfr<=>isde:t:cons trying to build target object >
isde:b:pred FOUND >
isfr<=>isde:tf:state ok >
isfr<=>isde:tf:extracte ok >
isfr<=>isde:tf:entere ok >
isfr<=>isde:tf:role ok >
isde:f:stat11 failed >

```

1. Le parser utilise une première fois la règle cons, pour copier les éléments de la racine de la langue source. Il l'applique une seconde fois pour copier la structure de la phrase dans la langue cible.
2. Il analyse ensuite les différents éléments de la phrase :
 - A. Il applique la règle atom pour trouver un équivalent au terme « être ». Il repère l'entrée lexicale de *sein* dans le dictionnaire sémantique allemand de `_lexicon`. Il applique ensuite successivement les règles cat, tense, stat, extract, enter et role qui copient respectivement la catégorie, le temps, les traits « stative », « extract » et « enter » et les rôles et vérifie la correspondance entre les deux langues. Il utilise ensuite la règle frame3 qui n'assigne aucun argument à *sein* qui n'en comporte que deux. Aucune autre règle f ne peut être appliquée, le transfert de l'élément est complet.
 - B. Il applique la règle cons sur le GN sujet puis la règle atom et repère l'entrée lexicale de *paul* dans `_lexicon`. Il transfère la catégorie, les traits des arguments et le rôle et vérifie l'exactitude du transfert. Il applique ensuite les règles frame1, frame2 et frame3 qui n'assignent aucun argument aux noms. Aucune règle supplémentaire ne peut être utilisée, l'analyse du nom est complète. Il essaie ensuite de construire le GN, il applique les règles cons, pred, agr et role ; elles fonctionnent toutes correctement, l'analyse du GN est terminée.
 - C. Il analyse ensuite les éléments qui composent le GP, il utilise la règle cons, il lie x*1 à la préposition et x*2 au GN.
 - i. Il applique la règle atom, localise l'entrée lexicale de *bei* dans `_lexicon`, vérifie la catégorie et le rôle et applique les règles frame1, frame2 et frame3. L'analyse est finie, le transfert est effectué.
 - ii. Il recommence l'étape B. ci-dessus pour le mot *frisoer*.
 - iii. Lorsqu'il a fini d'analyser la préposition et le GN, il essaie de reconstituer le GP. Il applique la règle pred, vérifie les traits stative, extract et enter ainsi que le rôle et utilise la règle stat11 qui stipule que la préposition *bei* est suivie d'un GN ayant les traits {hyper=agent,agent=isto}. L'analyse est complète et le sous-objet est construit.
3. Le parser a analysé le verbe, le GN sujet et le GP. Il tente alors de reconstituer la phrase. Il applique la règle pred, puis les règles tense, stat, extract, enter et role pour

vérifier la correspondance des traits entre les deux langues. Il emploie ensuite les règles complete1 et complete2 qui concernent le cadre de sous-catégorisation du verbe. Aucune autre règle f ne peut être appliquée, l'analyse de cet élément est terminée. On obtient un sous-objet avec un nœud-racine ayant pour traits {role=proposition, cat=v...} dominant les translations du verbe, du GN sujet et du GP.

4. Pour terminer, il repère la règle b text au niveau ISFR, il applique les règles agr et role. Il ne peut plus appliquer aucune règle, le transfert est terminé. Un objet est créé au niveau ISDE, sa structure arborescente est présentée ci-dessus :



La structure de traits correspondante est la suivante :

History : isde/1 <= isfr/1 <= csfr1 <= msfr/1

```

{role = text,
 stative = yes,
 extract = no,
 enter = no,
 cat = n,
 agr = {num = _1531}}
.[{cat = v,
   frame = { arg1 = { role = agent,
                    cat = n,
                    frame = { arg1 = nil,
                              arg2 = nil,
                              arg3 = nil},
                    agr = { num = sing}}
   arg2 = { role = loc,
          cat = p,
          frame = { arg1 = nil,
                    arg2 = nil,
                    arg3 = nil}},
          stative = yes,
          extract = no,
          enter = no},
   arg3 = nil},
 tense = {tns = pres},
 stative = yes,
 extract = no,
 enter = no,

```

```

role = proposition}
.[{ lex = sein,
  role = gov,
  cat = v,
  stative = yes,
  extract = no,
  enter = no,
  frame = { arg1 = { role = agent,
                    cat = n,
                    frame = { arg1 = nil,
                              arg2 = nil,
                              arg3 = nil},
                    agr = { num = sing}},
  arg2 = { role = loc,
          cat = p,
          frame = { arg1 = nil,
                    arg2 = nil,
                    arg3 = nil},
          stative = yes,
          extract = no,
          enter = no},
          arg3 = nil},
  tense = _1324},
{ cat = n,
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil},
  agr = { num = sing},
  role = agent}
.[{ lex = paul,
  role = gov,
  cat = n,
  type = abs,
  agr = { gen = masc,
         num = sing},
  sem = { gran = unique,
         anim = yes},
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil}}],
{ cat = p,
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil},
  role = loc,
  stative = yes,
  extract = no,
  enter = no}
.[{ lex=bei,

```



```

role = gov,
cat = p,
pos = left,
frame = { arg1 = nil,
          arg2 = nil,
          arg3 = nil}},
{ cat = n,
  frame = { arg1 = nil,
            arg2 = nil,
            arg3 = nil},
  agr = { num = sing},
  role = mod}
.[{ lex= frisoer,
   role = gov,
   cat = n,
   type = def,
   agr = {num = sing,
          gen = masc},
   sem = {hyper = {hyper = agent,
                   agent = isto}},
   sem ~ = {bound = mass},
   sem ~ = { orient = vertical,
            concrete = yes},
   sem ~ = {orient = horizontal},
   sem = ( {anim = yes},
           {hyper = { hyper = agent,
                     agent = isto}}),
   frame = { arg1 = nil,
             arg2 = nil,
             arg3 = nil}}]]]]]]

```

Au niveau du transfert de ISFR à ISDE, la préposition française *chez* peut être traduite, en allemand, par *bei* et par *zu*. Quand l'analyse est terminée pour *bei*, il recommence la synthèse à partir du GP en utilisant la préposition *zu*. Il suit les étapes i. à iii., mais comme la règle stat11 ne peut pas être appliquée, l'analyse échoue et s'achève.

5.3.2.8. TRANSFORMATION

Les règles du niveau de transformation isde<=>csde sont identiques à celles du niveau de transformation csfr<=>isfr. On a simplement ajouté le trait *case* puisque l'allemand est une langue casuelle. Il faut seulement signaler qu'il a fallu rajouter des règles concernant les groupes prépositionnels puisque l'allemand possède des prépositions qui se situent à gauche du GN (règle pp), des postpositions qui se situent à sa droite (règle pp1), et des circompositions qui l'entourent. Elles sont de la forme :

```

pp = {role=loc, stative=S, extract=E, enter=P}.[p:{cat=p}, n:{cat=n}] <=>
      {cat=pp, stative=S, extract=E, enter=P}.[p:{cat=p, pos=left}, n:{cat=np}].
pp1 = { role=loc, stative=S, extract=E, enter=P}.[p:{cat=p}, n:{cat=n}] <=>
       {cat=pp, stative=S, extract=E, enter=P}.[p:{cat=p, pos=right}, n:{cat=np}].
pp2 = { role=loc, stative=S, extract=E, enter=P}.[p:{cat=p, lex=durch_hindurch, pos=left}, n:{cat=np}] <=>

```

```

{cat=pp,stative=S,extract=E,enter=P}.[cat=p,lex=durch,pos=left},{cat=n},
{cat=p,lex=hindurch,pos=right}].
pp3 = { role=loc,stative=S,extract=E,enter=P}.[cat=p,lex=von_herab,pos=left},{cat=np}] <=>
{cat=pp,stative=S,extract=E,enter=P}.[cat=p,lex=von,pos=left},{cat=n},
{cat=p,lex=herab,pos=right}].

```

Dans le cas des circompositions, il faut écrire une règle par circomposition. Les circompositions sont écrites, dans les dictionnaires, sous la forme d'un mot composé avec tiret, mais il faut découper ce mot pour en obtenir deux que l'on peut placer de part et d'autre du GN.

5.3.2.9. LA STRUCTURE SYNTAXIQUE ALLEMANDE

5.3.2.9.1. Les règles

Une nouvelle fois, les règles sont identiques à celles du français, on également ajouté le trait *case*, et des règles pour distinguer les prépositions des postpositions dans les groupes prépositionnels.

```

pp1 = {cat=pp,stative=S,extract=E,enter=P,case=C}.[cat=p,pos=left,stative=S,extract=E,enter=P,
head={cat=p,stative=S,extract=E,enter=P,ehead={case=C}}},{cat=np,case=C}].
pp2 = {cat=pp,stative=S,extract=E,enter=P,case=C}.[cat=np,case=C},{cat=p,pos=right,stative=S,
extract=E,enter=P,head={cat=p,stative=S,extract=E,enter=P,ehead={case=C,case~gen}}}.
pp3 = {cat=pp,stative=S,extract=E,enter=P}.[cat=p,pos=left, stative=S,extract=E,enter=P ,
head={cat=p, stative=S,extract=E,enter=P ,ehead={case=C}}},{cat=n}, {cat=p,pos=right,
stative=S,extract=E,enter=P ,head={cat=p, stative=S,extract=E,enter=P ,
ehead={case=C}}}.
p_p = {cat=pp,stative=S,extract=E,enter=P}.[cat=p ], {cat=pp, stative=S,extract=E,enter=P ,
case=C}].

```

5.3.2.9.2. La synthèse syntaxique allemande

Pour construire la structure syntaxique, le parser va utiliser en combinaison les règles de transformation de IS à CS et les règles du niveau CS. On obtient l'analyse suivante :

```

0 $root*1 isde<=>csde:t:text trying >
0/1 s*1 isde<=>csde:t:s1 trying >
0/1/1 v*1 isde<=>csde:t:atom trying >
0/1/1 v*1 isde<=>csde:t:atom trying to build target object >
de_lexicon:b:sein FOUND >
isde<=>csde:tf:tense ok >
0/1/1 v*1 isde<=>csde:t:atom translation succeeded >
0/1/2 np*1 isde<=>csde:t:np1 trying >
0/1/2 np*1 isde<=>csde:t:np1 trying to build target object >
de_lexicon:b:paul FOUND >
csde:b:np1 FOUND >
0/1/2 np*1 isde<=>csde:t:np1 translation succeeded >
0/1/3 pp*1 isde<=>csde:t:pp trying >
0/1/3/1 p*1 isde<=>csde:t:atom trying >
0/1/3/1 p*1 isde<=>csde:t:atom trying to build target object >
de_lexicon:b:bei FOUND >
0/1/3/1 p*1 isde<=>csde:t:atom translation succeeded >
0/1/3/2 n*1 isde<=>csde:t:np2 trying >
0/1/3/2 n*1 isde<=>csde:t:np2 trying to build target object >
de_lexicon:b:det FOUND >

```

```

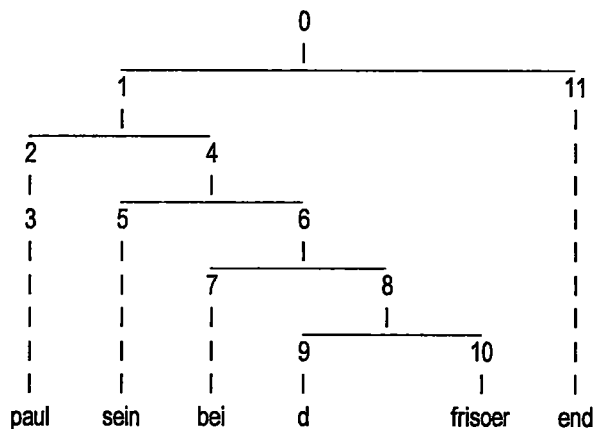
de_lexicon:b:frisoer FOUND >
csde:b:np2 FOUND >
0/1/3/2 n*1 isde<=>csde:np2 translation succeeded >
0/1/3 pp*1 isde<=>csde:pp trying to build target object >
csde:b:pp1 FOUND >
isde<=>csde:tf:tense ok >
0/1/3 pp*1 isde<=>csde:t:pp translation succeeded >
0/1 s*1 isde<=>csde:t:s1 trying to build target object >
csde:b:vp1 FOUND >
csde:b:s FOUND >
0/1 s*1 isde<=>csde:t:s1 translation succeeded >
0 $root*1 isde<=>csde:t:text trying to build target object >
de_lexicon:b:punct FOUND >
csde:b:text FOUND >
0 $root*1 isde<=>csde:t:text translation succeeded >
Created object : csde/l

0/1/3 pp*1 isde<=>csde:pp1 trying >
0/1/3/1 p*1 isde<=>csde:atom trying >
0/1/3/1 p*1 isde<=>csde:atom trying to build target object >
de_lexicon:b:bei FOUND >
0/1/3/1 $*1 isde<=>csde:t:atom translation succeeded >
0/1/3/2 n*1 isde<=>csde:t:np 2 trying >
0/1/3/2 n*1 isde<=>csde:t:np 2 trying to build target object >
de_lexicon:b:det FOUND >
de_lexicon:b:frisoer FOUND >
csde:b:np2 FOUND >
0/1/3/2 n*1 isde<=>csde:t:np2 translation succeeded >
0/1/3 pp*1 isde<=>csde:t:pp1 trying to build target object >
isde<=>csde:t:pp1 constraint on 'p' fails >
0/1/3 pp*1 isde<=>csde:t:pp1 fails-cannot build target object >

```

1. Il trouve la règle `text` qui lie le marqueur `s` à sein paul bei frisoer.
2. Il repère la règle `s1` qui lie le marqueur `v` au verbe, `np` au GN et `pp` au GP et analyse les différents éléments.
 - A. Il applique la règle `atom` et situe l'entrée lexicale de *sein* dans le dictionnaire sémantique allemand `de_lexicon` et vérifie le temps grâce à la règle `tf tense`. Il ne peut pas appliquer d'autres règles, l'analyse de cet élément est complète.
 - B. Il applique la règle `tf np1` et repère l'entrée lexicale de *paul* dans `de_lexicon`. Il utilise la règle `np1` du niveau CSDE pour construire un GN sans déterminant. Aucune règle supplémentaire ne peut être appliquée, l'analyse est terminée.
 - C. Il localise la règle `tf pp` qui, au niveau ISDE, lie `p` à la préposition et `n` au GN et essaie de construire le GP en analysant les différents éléments qui le composent.
 - i. Il applique la règle `atom` et situe l'entrée lexicale *bei* dans `de_lexicon`. Ce sont les deux seules règles applicables, l'analyse est par conséquent complète pour la préposition.
 - ii. Il essaie de transférer le GN de ISDE à CSDE grâce à la règle `np2`. Il repère les entrées lexicales pour *det* et *frisoer*, applique la règle `np2` du niveau CSDE qui décompose le GN en un nom précédé d'un déterminant. Il ne peut plus utiliser d'autres règles, l'analyse est finie et le GN *d frisoer* est construit.

- D. Il essaie ensuite de construire le GP, il emploie alors la règle pp1 du niveau CSDE. Comme aucune autre règle n'est applicable, l'analyse du GP *bei d frisoer* est terminée.
- E. Il tente de reconstituer l'élément s1 et applique successivement les règles vp1 (construction du groupe verbal) et s (groupe nominal suivi du groupe verbal). L'analyse de *paul sein bei d frisoer* est terminée.
- F. Il ne lui reste plus qu'à construire l'élément *text* en ajoutant le signe de ponctuation, qu'il repère dans le *lexicon*, à s. Il n'y plus rien à analyser ni aucune règle à appliquer. On obtient l'objet *paul sein bei d frisoer.* dont la structure arborescente est présentée ci-après :



La structure de traits correspondante a la forme suivante :

History : csde/1 <= isde/1 <= isfr/1 <= csfr1 <= msfr/1

```

{ cat = text,
  stative = yes,
  extract = no,
  enter = no,
  tense = _1586}
  .{ cat = s,
    tense = {tns = pres},
    stative = yes,
    extract = no,
    enter = no,
    mode = decl}
    .{ cat = np,
      agr = {num = sing,
             gen = masc},
      case = nom,
      tense = _1509}
      .{ cat = n,
        lex = paul,
        agr = {num = sing,
              gen = masc}
        type=abs,
        role = gov,
  
```

```

        sem = { gran=unique,
                anim = yes},
        case = nom}],
{cat = vp,
 agr = _1442,
 tense = {tns = pres},
 stative = yes,
 extract = no,
 enter = no}
  .[ {cat = v,
      lex = sein,
      string = _1401,
      role = gov,
      stative = yes,
      extract = no,
      enter = no,
      frame = { arg1 = ({role = agent,
                        cat = n},
                arg2 = {role = loc}},
      tense = {tns = pres},
      agr = _1442},
{cat = pp,
 stative = yes,
 extract = no,
 enter = no,
 case = _1310,
 tense = _1305}
  .[ {cat = p,
      lex = bei,
      string = _1284
      role = gov,
      pos = left,
      tense = _1269,
      stative = yes,
      enter = no,
      extract = _1254,
      head = {cat = p,
              stative = yes,
              enter = no,
              extract = _1254,
              ehead = {case = _1310}}},
{cat = np,
 agr = {num = sing,
        gen = masc},
 case = _1310,
 tense = _1183}
  .[ {cat = det,
      lex = d,
      agr = {num = sing,

```

```

        gen = masc},
    type = def,
    role = gov,
    case = _1310},
    {cat = n,
    lex = frisoer,
    agr = {num = sing,
          gen = masc},
    type = def,
    role = gov,
    sem = {hyper = {hyper = agent,
                    agent = isto}},
    case = _1310}]]],
{lex = end,
 cat = punct,
 mode = decl}]

```

G. Il recommence l'analyse à partir de C. mais en utilisant la règle pp1. Il réussit à recomposer le GN mais pas le GP et l'indique sous la forme :

```

isde<=>csde:t:pp1 constraint on 'p' fails,
en effet cette règle pp1 est employée pour une postposition ce qui n'est pas le
cas ici, l'analyse cesse donc à cet endroit.

```

5.3.2.10. LA STRUCTURE MORPHOLOGIQUE ALLEMANDE

Une nouvelle fois, nous avons un niveau intermédiaire entre les niveaux CS et MS. Les règles sont également utilisées pour définir la structure des noms composés et des prépositions contractées :

```

beim = [{lex=beim,string=beim}] <=> [{cat=p,lex=bei,string=bei,head={cat=p,string=bei,
    ehead={cat=n,pform=bei,case=dat}},{cat=det,lex=d,string=dem,type=def,
    agr={gen=neut ;masc),num=sing}}].

```

5.3.2.10.1. Le dictionnaire morphologique allemand

Il comporte les mêmes éléments que le dictionnaire morphologique français, auxquels on a ajouté le trait *case* spécifique à l'allemand.

5.3.2.10.2. La synthèse morphologique allemande

```

csde<=>msde:t: <paul> ... copying input... >
csde<=>msde:t: <sein> ... copying input... >
csde<=>msde:t: <beim> ... found multiword... >
csde<=>msde:t: <frisoer> ... copying input... >
csde<=>msde:t: <.> ... copying input... >
msde:b:paul ....FOUND.... >
msde:f:per3 ok >
msde:b:sein ....FOUND.... >
msde:b:beim ....FOUND.... >
msde:b:frisoer ....FOUND.... >
msde:f:per3 ok >
msde:b:punct ....FOUND.... >
Created object : msde/1

```

Il transfère les structures de traits de CSDE à MSDE puis recherche les différents termes dans le dictionnaire morphologique allemand, il applique, pour chaque nom et déterminant, la règle per3 qui assigne la troisième personne à ces deux types d'éléments. Il trouve le cas à assigner dans la description de la préposition *bei* et le répercute sur le déterminant et sur le nom s'il s'agit d'un masculin faible ou si le nom se décline (Haus = nominatif, Hause = datif...).

La traduction de *Paul est chez le coiffeur* est la suivante :

1 : Paul
2 : ist
3 : beim
4 : Frisoer
5 : .

elle est suivie de la structure de traits correspondante.

History : msde/1 <= csde/1 <= isde/1 <= isfr/1 <= csfr1 <= msfr/1

WORD# 1

```
{ cat = n,
  lex= paul,
  agr = { num = sing,
         gen = masc,
         per = '3'},
  type = abs,
  role = gov,
  sem = { gran = unique,
         anim = yes},
  case = nom,
  string = 'Paul'}
```

WORD# 2

```
{ cat = v,
  lex= sein,
  string = ist,
  role = gov,
  stative = yes,
  extract = no,
  enter = no,
  frame = {arg1 = { role = agent,
                   cat = n},
           arg2 = {role = loc}},
  tense = {tns = pres},
  agr = { num = sing,
         per = '3'}
```

WORD# 3

```
{ cat = p,
  lex= beim,
```

```

string = beim}
                WORD# 4
{ cat = n,
  lex= frisoer,
  agr = { num = sing,
        gen = masc,
        per = '3'},
  type=def,
  role = gov,
  sem = {hyper = {hyper = agent,
                 agent = isto}},
  case = dat,
  string = 'Frisoer'
                WORD# 5
{ lex= end,
  cat = punct,
  string = '.',
  mode = decl}

```

5.4. CONCLUSION SUR L'IMPLEMENTATION

À la fin de ce travail d'implémentation, on constate qu'il est relativement aisé d'établir des règles générales pour tous les noms propres géographiques mais que les critères pris en compte ne sont pas les mêmes dans les deux langues. En revanche, ce n'est pas le cas pour tous les noms communs. Il est exact que dans certains cas (avec les noms de bâtiment, les institutions, ...) on peut essayer de généraliser les règles, mais il faut souvent faire des règles spécifiques, et ceci est d'autant plus vrai que l'on essaie de traduire des phrases au sens figuré. Dans le chapitre sur les prépositions spatiales, nous avons présentés les prépositions *à*, *de* et *en* dans la catégorie des prépositions vides. Pourtant si on se réfère uniquement au domaine spatial on remarquera que *de* exprime uniquement la provenance ou l'extraction et que *en* n'a que peu de traduction en allemand. En revanche *à* a beaucoup d'équivalent mais il est intéressant de noter que c'est également le cas de la préposition *dans* qui accepte dix traductions différentes.

La tâche la plus complexe est de définir les traits sémantiques. En effet ils doivent être pertinents et par conséquent précis, mais il ne faut pas trop en mettre pour éviter un ralentissement du processus de traduction. De plus si on insère trop de traits dans la structure sémantique des dictionnaires, il peut se poser un problème dans l'établissement des règles d'utilisation de telle ou telle préposition, car on risque d'avoir un même trait dans la contrainte positive et dans la contrainte négative : par exemple la règle

```

{cat=p}>>{lex=dans},
{cat=n}.[ {cat=n,sem={shape=enclature,hyper=moebel},sem~={shape=surface,hyper=moebel}}].

```

signifie que la préposition *dans* s'emploie dans le cas d'un meuble qui est un contenant (par ex. armoire) mais pas si le meuble est une surface (par ex. table). Cependant cette règle ne peut pas fonctionner car le trait {hyper=moebel} apparaît dans les deux contraintes imposées à la sémantique.

On est également obligé d'écrire une règle pour les verbes de mouvement et une pour les verbes statiques, puisqu'en allemand la préposition peut changer en fonction de son régime et du verbe qui la précède (*in die Schweiz fahren, nach Frankreich fahren*). De même, pour une même préposition le cas peut changer en fonction du verbe (*in der Schweiz sein, in die Schweiz fahren*).

Comme certains mots de la langue source peuvent avoir plusieurs traductions dans la langue cible, il faut écrire des règles « assassines » pour éliminer les fausses traductions. Plus il y a de termes avec plusieurs équivalents dans les dictionnaires, plus il faut écrire de règles « assassines » en faisant toutefois attention à ce qu'elles ne fassent pas interférence l'une avec l'autre.

Comme le programme ne comporte pas d'analyseur morphologique :

- * il nous est impossible de scinder un verbe à particule séparable en un verbe et une particule par conséquent nous devons traiter le verbe comme verbe simple,
- * on ne peut pas analyser la circomposition comme une combinaison de deux prépositions, il faut écrire une règle pour chaque cas,

Mais ces trois problèmes pourront être résolus dès lors que l'on intégrera notre programme à un programme plus important comprenant un analyseur morphologique.

6. CONCLUSION GENERALE

Les résultats obtenus à l'issue de l'implémentation montrent qu'il existe des points communs entre les prépositions allemandes et françaises :

- * la préposition se trouve en tête du groupe prépositionnel,
- * dans le choix de la préposition, il faut tenir compte du verbe qui la précède et du groupe nominal qui la suit,
- * il existe des prépositions monosémiques et polysémiques dans les deux langues,
- * on peut, dans certains cas, avoir contraction de la préposition et du déterminant.

Il est clair qu'elles présentent également des divergences :

- * il existe des postpositions et des circompositions en allemand mais pas en français,
- * les prépositions régissent un cas en allemand mais pas en français,
- * la distinction locatif / directif n'a aucune influence en français alors qu'elle est essentielle en allemand, tant pour le choix de la préposition que pour le cas qu'elle régit,
- * en ce qui concerne les noms, on ne prend en considération les mêmes critères : alors que le français tient compte de l'initiale du nom, de son genre et /ou de sa composition (mot composé ou mot simple), l'allemand se réfère à la présence ou à l'absence du déterminant,
- * il n'y a, apparemment, pas de prépositions vides en allemand,
- * la « portée » de la préposition peut également être différente : par exemple, dans la phrase allemande, *Paul wirft die Flasche in das Wasser*, on ne sait pas si la bouteille va entièrement ou partiellement sous l'eau, en français cette distinction sera marquée par l'utilisation de deux prépositions différentes : *dans*, si la bouteille s'enfonce complètement dans l'eau, *à* si elle reste à la surface.

Il est clair que cette étude a demandé beaucoup de travail de réflexion, et nous nous sommes heurtés à certains problèmes :

- * en français, par exemple, il a été très difficile de trouver avec précision le genre des noms propres (notre recherche n'a d'ailleurs pas abouti complètement), notamment pour les noms d'îles. En fait, ce problème s'est posé à nous car les règles d'utilisation des prépositions avec les noms d'îles proposées dans le Bon Usage de Grévisse, portaient du genre et de la localisation des îles. Nous avons vite remarqué que ces règles n'étaient pas valides en tant que telles et qu'il était nettement préférable de tenir compte de la présence ou non du déterminant. C'est ainsi que ce problème s'est en quelque sorte résolu.
- * en français, nous avons rencontré un problème au niveau de l'implémentation de certains noms féminins qui tantôt s'emploient avec un déterminant et tantôt sans. Il a fallu, pour chacun d'eux, rechercher tous les cas où ils s'utilisent avec

le déterminant et inversement ceux où ils s'utilisent sans celui-ci et écrire les règles de construction et les règles « assassines » correspondantes,

- * en allemand, la plus grosse difficulté a résidé dans le choix de la préposition avec un nom propre, car nous n'avons pas trouvé de règles préétablies. Pour ce faire, nous avons consulté diverses brochures d'agence de voyage, repéré les prépositions et les noms propres et construit nos propres règles,
- * il est parfois ardu de trouver des équivalents entre les deux langues et de construire les règles adéquates.

Il est vrai qu'il reste quelques problèmes en suspens :

- * notre programme ne permet pas la traduction de proverbes ou d'idiomes,
- * il ne traite pas non plus du cas où une des langues comprend une préposition et l'autre n'en a pas,
- * il ne prend en compte que les phrase composées d'un sujet, d'un verbe et d'un GP ou d'un sujet, d'un verbe, d'un COD et d'un GP, et ne s'intéresse pas aux prépositions présentes à l'intérieur d'un groupe nominal,
- * le parser ne peut pas générer des locutions prépositives du type à côté du, en face du, la contraction entre la préposition et le déterminant n'a pas lieu. Nous avons préféré travailler sur la traduction effective des prépositions (ce qui est le but de notre étude) et nous ne nous sommes pas attardés sur les phénomènes purement morphologiques. Il faut quand même signaler que la contraction est possible pour les prépositions simples.
- * certains cas particuliers ne sont pas traités pour diverses raisons :
 - il faut impérativement connaître le contexte pour utiliser les noms adéquats derrière la préposition (la mer vue par rapport à la terre ne se traduira pas de la même façon que la mer vue par rapport au bateau), et nous travaillons hors contexte,
 - il est impossible de faire une généralisation car chaque cas doit être étudié de façon indépendante, et nous avons décidé de ne pas les traiter dans l'immédiat car il est inutile d'écrire une règle par cas, ce qui va ralentir le processus de traduction et ce qui va à l'encontre du traitement automatique.

Cette étude nous a bien évidemment permis de tirer quelques enseignements :

- * la préposition français *en*, dans son sens spatial, ne peut s'employer sans déterminant, alors que dans son sens temporel c'est tout à fait possible.
- * même s'il semble possible de généraliser les règles d'utilisation des prépositions, en théorie, il est beaucoup moins facile de les généraliser en pratique. En effet, au niveau informatique, il faut écrire toute une série de règles de construction et de règles « assassines » et éviter qu'elles n'interfèrent les unes avec les autres. Nous avons constaté qu'il est nettement préférable de donner des noms différents aux règles, car, en cas de problème dans le processus de traduction, il est plus simple de repérer celle(s) qui pose(nt) problème lorsque l'on fait une trace

- * si on s'en tient uniquement du sens purement spatial des prépositions et plus précisément des locutions prépositives, on remarque qu'il existe quasiment toujours une traduction unique ; en revanche, quand les prépositions peuvent s'employer dans un sens spatial mais également dans un sens temporel ou abstrait, le nombre de traduction pour une préposition s'accroît,
- * il faut être très précis dans les dictionnaires. En effet, plus les dictionnaires sont complets, moins les règles de construction sont complexes, et les règles « assassines » s'en trouvent réduites d'autant,
- * sans contexte, il n'est pas possible de donner la préférence à une solution plutôt qu'à une autre. Dans le cas où une préposition française peut se traduire par deux prépositions en allemand et que le choix de celle-ci dépend du contexte, nous avons décidé de garder les deux solutions, et inversement.

Pour terminer nous pouvons dire que notre travail pourra ôter un doute à A.-M. Berthonneau et Pierre Cadiot qui ne sont pas certains que l'on puisse implémenter les prépositions (cf. supra p. 189).

Glossaire

Les mots qui se trouvent dans ce glossaire ont été signalés dans le texte par le symbole \diamond .

Affixe : LING. Élément qui se met au commencement (préfixe), à l'intérieur (infixe) ou à la fin (suffixe) d'un mot pour en modifier le sens ou la valeur grammaticale.

Anaphore : STYL. Reprise d'un mot ou d'un groupe de mots au début de phrases ou de membres de phrases qui se suivent, produisant un effet de renforcement, de symétrie.

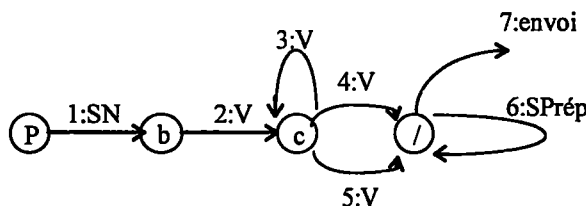
ATN¹ : Un réseau de transitions augmenté est un réseau de transitions récursif qui inclut les extensions suivantes : un ATN contient un nombre illimité de registres lui permettant de conserver de l'information en mémoire et d'accéder directement au contenu de chaque registre ; les arcs d'un ATN peuvent être étiquetés par des *conditions* et des *actions*. Une **condition** fait référence au contenu des différents registres et doit être satisfaite pour que l'arc qui la spécifie puisse être emprunté. Elle peut éventuellement - mais pas nécessairement - induire l'exécution d'une **action** qui, elle, modifie le contenu d'un registre. Les ATNs permettent la construction *en ligne* d'une structure profonde plutôt qu'une construction *post hoc* basée sur la dérivation initiale d'une structure de surface.

Grammaire

Base : $P \rightarrow SN SV$

$SV \rightarrow V_{avoir} V_{pp} SN$

$t_{passive} : X SN_1 V SN_2 Y \rightarrow S SN_2 \hat{E}tre V+é Y$



Cataphore : STYL. A l'inverse de l'anaphore, elle se caractérise par le fait que le terme repris précède le terme en expansion.

¹ Cette définition est extraite du livre de [Baschung92].

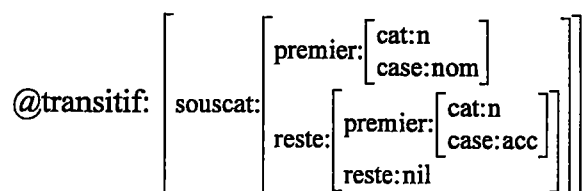
Compétence : Connaissance de la langue maternelle. Savoir sous-jacent du locuteur sur sa langue. Système formé par les règles et les éléments auxquels elles s'appliquent, intégré par l'usager d'une langue naturelle et qui lui permet de former un nombre indéfini de phrases grammaticales et de comprendre des phrases jamais entendues.

Compétence grammaticale : Faculté de former et interpréter des expressions linguistiques.

Compétence pragmatique : Faculté d'employer des expressions de manière appropriée pour arriver à l'effet communicationnel désiré.

Exophorique : Une relation est exophorique lorsque le présupposé se trouve hors du texte dans la situation qu'il évoque.

Forme-type : Une forme-type sert à décrire un ensemble de propriétés communes à plusieurs entrées lexicales ; elle permet de capturer des généralisations linguistiques s'appliquant à certaines classes de mots.



Graphe acyclique orienté : Le dag est orienté parce que l'ordre de parcours de ses nœuds est important, il est acyclique, car si deux arcs peuvent mener à un même nœud, les boucles complètes sont interdites.

Héritage / Héritage par défaut : Selon la relation d'héritage, définie par une hiérarchie de types (ou de classes d'objets), chaque objet comprend nécessairement, outre les propriétés caractéristiques de son type, toutes les propriétés des types supérieurs.

Héritage multiple : Il y a héritage multiple si un type donné peut avoir plusieurs sur-types non ordonnés.

Héritage partiel : Il y a héritage partiel si l'objet hérite seulement des propriétés des types supérieurs qui sont compatibles avec les siennes propres.

Isomorphisme : Identité formelle de deux ou plusieurs structures relevant de plans ou de niveaux sémiotiques différents, reconnaissable du fait de l'homologation possible des réseaux relationnels qui les constituent.

Onomasiologie : Étude sémantique des dénominations qui part du concept et recherche les signes linguistiques qui lui correspondent.

Performance : Mise en œuvre des connaissances pour la production d'énoncés. Comportement qui consiste à parler.

Prototype : La notion de prototype a été introduite pour mettre en cause la vision classique de la catégorisation (qui remonte à Aristote), aussi connue sous le nom de modèle ou théorie des conditions nécessaires et suffisantes. On considérait, jusque récemment, que pour qu'un objet appartienne à une catégorie, il fallait que cet objet satisfasse un certain nombre de conditions nécessaires et suffisantes. La théorie du prototype, apparue il y a une vingtaine d'années, substitue à cette vision rigide de la catégorisation une vision beaucoup plus souple selon laquelle un objet appartient ou n'appartient pas à une catégorie selon le degré de ressemblance qu'il entretient avec un exemplaire particulier de cette catégorie ou prototype.

Sémasiologie : Étude des significations, partant du mot pour en étudier le sens.

Sous-catégorisation : La tête lexicale d'un syntagme sous-catégorise syntaxiquement un certain nombre de compléments dont elle exige la présence (compléments obligatoires) ou bien dont elle détermine simplement la forme syntaxique. Un verbe *sélectionne* un sujet mais *sous-catégorise* un complément (sauf dans les grammaires de valence où le sujet fait partie de la valence d'un lexème, qui est l'équivalent de la sous-catégorisation dans cette théorie).

Jean donne un baiser à Marie

1 2

Jean interdit à Médor de sortir.

1 2

Structure de surface : Suite de sons ou de mots qui constituent la phrase.

Structure profonde : Composant sémique qui produit le sens de la phrase engendrée.

Taxinomie : Conçue traditionnellement comme la « théorie des classifications », la taxinomie s'applique actuellement à la classification elle-même, c'est-à-dire aux procédures d'organisation systématique des données observées et décrites.

t-uplet : Un triplet contient trois éléments. Un t-uplet est un ensemble qui contient t éléments.

Type : Un type est une sorte d'objet dénotée par une catégorie en logique intensionnelle. Les types de base sont *e* pour entité et *t* pour valeur de vérité.

INDEX DES FIGURES

Figure 1 : avant / après avec les ordres dynamiques.....	41
Figure 2 : avant / après avec les ordres statiques.....	41
Figure 3 : Les catégories « medium » de Hawkins	48
Figure 4 : Structure de ressemblance des familles de « medium »	52
Figure 5 : devant / derrière avec les ordres dynamiques.....	53
Figure 6 : devant / derrière avec les ordres statiques.....	53
Figure 7 : Le système des prépositions spatiales	60
Figure 8 : Le système 1 de P. Marcq	60
Figure 9 : Organisation du modèle LFG	121
Figure 10 : Organisation du modèle GPSG	128
Figure 11 : Organisation du modèle HPSG	140
Figure 12 : Organisation hiérarchique des signes linguistiques dans HPSG	143
Figure 13 : Organisation simplifiée de quelques signes lexicaux (mots non fonctionnels)	143
Figure 14 : Architecture du système de TA dans le projet KIT-FAST	181
Figure 15 : Élaboration d'un objet morphologique avec un dictionnaire de formes pleines	210
Figure 16 : Analyse syntaxique en CAT2	211
Figure 17 : Transformations d'arbres	211
Figure 18 : Génération morphologique dans CAT2	212
Figure 19 : Les traits sémantiques proposés par C. Zelinsky-Wibbelt	249

INDEX DES TABLEAUX

Tableau 1 : Combinaisons possibles entre adverbes et prépositions	15
Tableau 2 : Contractions possibles entre prépositions et articles	16
Tableau 3 : Les principales constructions des prépositions.....	19
Tableau 4 : Trinôme institutionnel de Mantchev.....	36
Tableau 5 : Utilisation des articles et des prépositions en français	37
Tableau 6 : Les îles : genre, déterminant, localisation, superficie.....	40
Tableau 7 : Tableau récapitulatif du système 1	61
Tableau 8 : Tableau récapitulatif du système 2	64
Tableau 9 : Tableau récapitulatif du système 3	66
Tableau 10 : Les cas dits locaux en finnois	83
Tableau 11 : Tableau récapitulatif sur les cas.....	95
Tableau 12 : Les prépositions allemandes : positions et rections.....	99
Tableau 13 : Les notations utilisées dans UCG.....	112
Tableau 14 : Les catégories dans UCG (version française FDP).....	115
Tableau 15 : Les fonctions grammaticales dans LFG.....	123
Tableau 16 : Notation traditionnelle vs notation LFG.....	125
Tableau 17 : La notation GPSG selon K. Baschung.....	137
Tableau 18 : Notation traditionnelle vs notation GPSG	138
Tableau 19 : Notation traditionnelle vs notation HPSG.....	146
Tableau 20 : Notation traditionnelle vs notation TAG	157
Tableau 21 : Tableau comparatif des modèles de grammaires basées sur l'unification	163
Tableau 22 : Phases du processus de traduction.....	209
Tableau 23 : Phases de la traduction du point de vue du modèle de traitement.....	209
Tableau 24: Correspondance entre les prépositions françaises et allemandes.....	231-234
Tableau 25 : Correspondance entre les prépositions allemandes et françaises.....	234-237

INDEX DES NOMS PROPRES

A

Abeillé (A.) : 139; 149; 150; 188; 192; 196
Ajdukiewicz (K.) : 110; 111
Alsina : 122
André (E.) : 74; 78
Andrew (A.) : 179
Arnolds : 198
Aurnague : 72
Austin : 122; 124

B

Bajcsy : 80
Baker : 185
Barclay : 12
Bar-Hillel (Y.) : 110; 111
Bartsch (R.) : 22
Baschung (K.) : 134; 137
Bennett : 43
Berthonneau (A.-M.) : 168
Bierwisch : 43
Blache (P.) : 186; 186
Block (H.-U.) : 195
Boer (C. de) : 5; 7; 30; 31; 94
Boggess : 72
Boguraev : 93
Borillo (A.) : 43; 72
Bransford : 12
Bréal : 80
Bredenkamp (A.) : 183
Bresnan (J. W.) : 102, 120, 121, 122, 124;
176; 198
Briffault (X.) : 70
Brinkman (H.) : 7
Brøndal (V.) : 36; 84
Brun (C.) : 176
Bruneau (Ch.) : 30
Brunot (F.) : 30
Butt (M.) : 121; 176

C

Cadiot (P.) : 59; 168
Calder (J.) : 185-186
Carpenter (B.) : 182

Chomsky (N.) : 103; 122; 191; 197
Church : 189
Colmerauer : 103
Confais (J.-P.) : 68; 80
Cooper : 43
Covington (M.-A.) : 187; 195
Cowie : 190
Cuyckens : 24; 50

D

Daille (B.) : 192
Dalrymple (M.) : 176
Damourette : 4
Dauzat : 5
Davis : 80
Doerre : 185
Döpke : 18; 22
Doran : 188
Dorna (M.) : 196
Dörre (J.) : 196
Dunker (G.) : 143

E

Egedi : 190
Ehrlich (U.) : 174
Eisele : 185
Emele (M.) : 186
Engels : 99

F

Fauconnier (G.) : 12
Feuillet : 94
Fillmore (C.) : 85; 86; 88; 96; 97
Fitschen (A.) : 193
Fodor (J.-A.) : 168
Fortmann (C.) : 176
Franckel (J.-J.) : 34; 35
Franklin : 13
Franks : 12
Frege (G.) : 110; 111
Frei (H.) : 33
Freksa : 72

G

Garham : 13
Garrod : 43
Gazdar (G.) : 103; 128; 199
Geach (P.-T.) : 110
Gehrke (M.) : 195
Glasgow : 13; 74
Godard (D.) : 188
Gopal : 74
Gougenheim : 31; 35
Gregory (H.) : 186
Grévisse (M.) : 31; 38; 53; 57
Grimes : 93; 94
Guillaume (G.) : 20; 35; 40; 84; 85; 98
Guimier (E.) : 188
Gungordu (Z.) : 179

H

Habel : 74
Haegeman : 190
Hahn (von) : 80
Heusinger (K. von) : 108; 111
Hanks : 189
Harbusch : 78
Hasida (K.) : 184
Hawkins : 43; 47; 49; 50
Herskovits : 10; 21; 23; 43; 52
Herweg : 25; 43; 48; 49
Herzog (G.) : 74
Heusinger : 111
Hjelmslev (L.) : 97; 108
Hornby : 189
Hottenroth : 11; 24; 26; 43; 48
Hubrich (J.) : 193
Humphreys (K.) : 185
Hunze (R.) : 195
Husserl (E.) : 110
Hußmann : 80
Husson (A.) : 192

J

Jackendoff (R.) : 21; 76
Jaye (J.-H.) : 164
Johnson (M.) : 24; 185; 198
Johnson-Laird : 13; 43; 48
Joshi (A.) : 102; 122; 149
Jouannaud (M.-P.) : 188
Jung (M.) : 193

K

Kanerva 122
Kaplan (R.) : 108; 120; 124; 127; 176; 177;
178; 198
Karp : 189
Kasper (R. T.) : 198
Kaufmann : 11
Kay (M.) : 12; 102; 109; 198
King (T.) : 121; 179
Kintsch : 13
Klatzky : 74
Klein (E.) : 128; 186
Kodric (S.) : 185
Krieger (H.-U.) : 186
Kroch (A.) : 149
Kroeger : 121
Kuipers (J.-P.) : 72; 73
Kullbach (B.) : 193

L

Lakoff : 24
Lambek (J.) : 110; 111
Landau : 76; 164
Lang : 10; 11; 24
Langacker : 10; 21
Lappin (S.) : 186
Larousse : 3; 4; 6
Le Bidois (G.) : 7
Le Bidois (R.) : 7
Lebaud (D.) : 34; 35
Leech : 43
Leeman (D.) : 59
Leiser : 74
Leniewski : 110
Létoublon (F.) : 97
Lindkvist (C.) : 43
Lüth (T.-G.) : 78
Lutzeier : 43

M

Maas (H.-D.) : 250
Mackin : 190
Madvig : 80
Manning : 127
Mantchev : 36
Martin : 190
Maxwell (J.) : 127; 179; 177; 178
Mayo (B.) : 178

Mchombo : 122
Miller 43; 48; 188
Mohanan (K.-P.) : 121
Moignet (G.) : 30; 40
Moilanen (M.) : 43; 71
Montague (R.) : 105; 129; 135
Morin (J.-Y.) : 196
Mounin (G.) : 6; 9
Müller (S.) : 184

N

Namer (F.) : 188; 192
Neumann : 80
Nino (M.-E.) : 176
Noord (G. van) : 184
Nordlinger : 121

P

Papadias : 13; 74
Paul (H.) : 8
Penman : 171
Penn (G.) : 182
Pereira (F.-C.-N.) : 103
Phelizon (J.-F.) : 6
Pieper (S.) : 193
Pieton : 4
Pinchon (W.) : 4
Pollard (C.-J.) : 139; 145; 187; 188; 198
Popowich (F.) : 185
Pottier (B.) : 11; 32; 34; 41; 90; 91; 92;
167; 168
Pribbenow : 11; 24; 43
Pullum (G.-K.) : 103; 128
Pustejovsky : 167

R

Rastier (F.) : 164
Reape (M.) : 185
Recourcé (G.) : 188; 192
Retz-Schmidt : 78
Richarz (D.) : 193
Ridder (H.) : 193
Ringenson : 40
Ritzke (J.) : 172
Robert (P.) : 3; 5; 90
Robichaud (B.) : 188; 192
Rohr : 74
Rohrer (C.) : 176
Rosner : 185

Rounds-Kasper : 182
Russel : 111
Ruwet (N.) : 32

S

Sabah (G.) : 164
Sag (I.) : 128; 139; 145; 152; 187; 188; 198
Saint-Dizier (P.) : 165
Samvelian (P.) : 192
Sanford : 43
Scaliger : 80
Schabes (Y.) : 149; 155; 156; 189;
Schachtl (S.) : 195
Schanen (F.) : 68
Schank (R.) : 88
Schauer : 194
Scheffe : 80
Schieber (S.) : 149; 155
Schirra : 11; 74; 78
Schmid (L.-A.) : 195
Schwarze (C.) : 18; 22
Searle (J.-R.) : 27
Segond (F.) : 176
Sharp (R.) : 198
Shieber (M.-S.) : 156
Simmons : 88; 96
Simpson : 127
Sirai (H.) : 184
Slatka : 56
Smith : 74
Sowa : 164
Spang-Hanssen : 31
Spark-Jones : 93
Stolzenburg (E.) : 187
Stopp (E.) : 78

T

Talmy : 21
Tombe (des) : 198
Tsuda (H.) : 184
Tversky : 13

U

Umbach (C.) : 179

V

Vandeloise (C.) : 10; 21; 29; 32; 33; 41;
42; 43; 48; 49; 53
Van Dijk : 13

Vieu : 43; 72
Vijay-Shanker (K.) : 149; 153; 155; 156
Vogel (C.) : 185
Volk (M.) : 193

W

Wagner (A.) : 193; 194
Wahlster : 77
Warren (D.-H.-D.) : 103
Wartburg (W. von) : 31
Webster : 50
Wehrli (E.) : 196
Weir (D.) : 149; 155; 156
Weissenborn (J.) : 97
Weisweber : 179
Wilks : 164
Wirt (M.) : 193
Wood (W.) : 93
Wunderlich (D.) : 26; 43; 48

Z

Zaenen (A.) : 122
Zeevat (H.) : 186
Zelinsky-Wibbelt (C.) : 23; 165; 171; 248
Zilbershatz : 74
Zumthor (P.) : 31
Zünkler (C.) : 195

INDEX DES PROJETS

ALE (HPSG) : Attribute Logic Engine 182
ALEP (HPSG) : Advanced Linguistic Engineering Platform 183
BABEL (HPSG) : 184
CITYTOUR : 74
CUF : 196
FDP (UCG) : 116
GPSG parser (GPSG) : 179
GPSG-Tools (GPSG) : 179
GULP : 195
HAM-RPM 80
HDRUG (HPSG) : 184
HPSG-PL (HPSG) : 185
JPSG (HPSG) : Japanese Phrase Structure Grammar 184
KAMRO : 75
Kit-Fast (GPSG) : 180
KoGiTum : 195
Konstanz LFG Workbench (LFG) : 178
LANDSCAN : 80
LFG PARGRAM (LFG) : 176
LKP : 195
MERCATOR : 80
NAOS : 80
Parser LFG pour le turc (LFG) : 179
PLEUK (HPSG) : 185
Projet de Grammaire Syntagmatique du français (HPSG) : 188
SOAS HPSG Ellipsis (HPSG) : 186
SOCCER : 75
STORM (UCG) : 176
SWYSS : 80
SYSLID (UCG) : 174
TDL : Type Description Language (HPSG) : 186
TFS : Typed Feature Structures System 6.1 (HPSG) : 186
TOUR 72, 73
UBS : UnifikationsBasierte Sprache (HPSG) : 187
VITRA : 74
Xerox LFG Grammar Writer's Workbench (LFG) : 178
XTAG (TAG) : 188

INDEX DES NOTIONS

A

à côté de : 76
à droite de : 76
à gauche de : 76
à : 19; 20; 30; 31; 37; 38; 39; 40; 56; 57;
58; 76; 92; 214; 215; 216; 217; 218;
219; 220; 221; 223; 224; 225; 227
ablatif : 84; 94
abseits : 20
accusatif : 83; 84; 90; 91; 92; 94; 98
acteur : 88
adjonction : 163; 164; 165
agent : 87; 93; 96
agentif : 86; 90; 92; 94; 96
an : 61; 62; 63; 65; 69; 70; 71; 218; 223;
224; 225
ancrage sémantique : 153
après : 41; 42; 221
arbre :
 arbre auxiliaire : 152
 arbre de dérivation : 150; 152
 arbre dérivé : 150; 152
 arbre élémentaire : 149
attributs fixes : 204
au bord de : 223; 225; 226
au-dedans de : 54
au-dehors de : 54
au-dessous de : 53
au-dessus de : 53; 76
auf : 61; 65; 67; 69; 216; 217; 224; 225
aus : 60; 61; 69; 72; 221; 224
außer : 69
außerhalb : 67; 69; 71
avant : 41; 42

B

bei : 65; 66; 68; 71; 219; 221
bénéfactif : 90; 92
bénéficiaire : 93; 96
but : 87; 88; 93; 96

C

cas : 82-96
 cas conceptuel : 91
 cas d'emploi : 85

cas de discours : 98
cas de fonction : 84
cas de langue : 98
cas direct : 84
cas linguistique : 92
cas oblique : 84
cas profond : 85; 86
cas sémantique : 84
cas syntaxique : 84
casualisation : 83
causal : 90; 92
causatif : 92
chemin : 93; 96
chez : 30; 54; 221
chrono-expérience : 93
chronologie : 92
coindexation : 142
concaténation : 119
constante : 202
contrainte :
 contrainte concaténative : 203
 contrainte conditionnelle : 203
 contrainte disjonctive : 203
 contrainte existentielle : 203
 contrainte négative : 203
 contrainte positive : 203
contre : 20; 54
contre-agent : 87

D

dans : 19; 20; 25; 30; 43-53; 56; 57; 58;
215; 217; 220; 221; 224; 225; 226
datif : 84; 86; 90; 92; 98
DCG : 103; 108-109
de : 19; 20; 30; 34; 40; 58; 92; 224; 226
dépendances :
 dépendances à longue distance : 153
 dépendances locales : 152
 dépendances moyennes : 153
derrière : 19; 20; 25; 40; 53; 76; 221
descriptif : 94; 96
dessus : 20
destination : 93
devant : 19; 25; 40; 53; 76; 222
diesseits : 67

directif : 94
disjonction : 110
dominance casuelle : 90
durch : 60; 61; 63; 64; 65; 70; 71; 72

E

eingangs : 20
en bas de : 53
en dedans de : 54
en dehors de : 54
en haut de : 53
en : 19; 30; 34; 37; 38; 39; 40; 57; 58; 214;
215; 217; 221; 223; 224; 225; 227
en-dessous de : 76
entgegen : 63; 64
entlang : 64; 70
entre : 30; 54; 76; 225
ergatif : 90; 91
espace : 10
espaces mentaux : 12
extension : 107; 134

F

factitif : 86
FAS (Fonctor-Agreement-Structure) : 180
final : 90; 92
flexion casuelle : 84
focus : 123
fonctions non sous-catégorisables : 123
fonctions sous-catégorisables : 123
force : 93; 96
formalisme :
 formalisme déclaratif : 104
 formalisme inductif : 104
 formalisme informatif : 103
 formalisme surfaciste : 103
FUG : 102; 109

G

gegen : 63; 64; 65; 69
gegenüber : 62; 70
généralisation 107
générateur : 199; 205
 générateur morphologique : 205
 générateur relationnel : 205
 générateur syntaxique : 205
génitif : 84; 98
GPSG : 103; 128-138; 180
grammaire arborescente : 149

grammaire syntagmatique : 149
graphe conceptuel : 159
GTU : 193

H

héritage : 142
héritage multiple : 143
hinter : 59; 61; 62; 65; 69; 70
hors (de) : 19; 43
HPSG : 103; 139-147; 192; 193

I

in : 60; 61; 63; 66; 68; 69; 71; 214; 215;
216; 217; 220; 221; 224; 226; 227
innerhalb : 67; 71
instrument : 87; 88; 93; 96
instrumental : 84; 86; 90; 92
instrumental-sociatif : 94

J

jenseits : 67
jusqu'(e) : 55

L

landmark : 11
längs : 64; 70
LFG : 102; 120-130; 192
lieu : 88; 93
localisation : 24; 26
 localisation actantielle : 22
 localisation dynamique : 24
 localisation propositionnelle : 22
 localisation spatiale : 21
 localisation statique : 24
locatif : 84; 86; 92; 94

M

matériel : 93
médiatif : 94; 96

N

nach : 64; 65; 66; 68; 70; 214; 215; 216;
218; 221
neben : 59; 61; 62; 64; 65; 69; 70
niveau : 204
niveau commun : 206
nombre : 210
nominatif : 83; 84; 90; 91; 92; 94; 98
non-compositionnalité : 154

O

oberhalb : 67
objectif : 86; 94; 96
objet : 87; 93; 96; 201
orientation : 10; 42

P

paires prépositionnelles : 40
par : 30; 55; 57; 59; 92; 223
par-dessus : 223
parmi : 55
passage : 72
 passage imperfectif : 71
 passage perfectif : 71
patient : 87; 93; 96
PATR-II : 119-120
position : 24
pour : 20; 30; 55; 58; 92
préposition :
 préposition à complément zéro : 20
 préposition abstraite : 31
 préposition casuelle : 30
 préposition demi-vidée : 30
 préposition incolore : 31
 préposition non-casuelle : 30
 préposition perlativale : 70
 préposition pleine : 30
 préposition projective : 24
 préposition semi-casuelle : 30
 préposition vidée : 30; 31; 105
près de : 76
primitives :
 primitives basiques : 113
 primitives non basiques : 113
principes :
 principes associatifs dépendants : 47
 principe d'accès : 12
 principe d'ancrage lexical (TAG) : 157
 principe d'association central : 47
 principe d'ordre des mots (HPSG) : 146
 principe d'unicité (LFG) : 125
 principe de cohérence (LFG) : 125
 principe de complétude (LFG) : 125
 principe de consistance sémantique (TAG) : 157
 principe de cooccurrence prédicat-arguments (TAG) : 157
 principe de cooccurrence : 104

principe de non-compositionnalité (TAG) : 157
principe de parcimonie : 12
principe de partage des traits (HPSG) : 144
principe de propagation des traits : 104
principe de restriction sur la cooccurrence des traits : 136
principe de sous-catégorisation (HPSG) : 145
principe de spécification de traits par défaut (GPSG) : 136
principe des traits de Pied (GPSG) : 134
principe des traits de Tête (GPSG) : 134
principe des traits de Tête (HPSG) : 144
principe des traits non locaux (HPSG) : 146
principe du contrôle et de l'accord (GPSG) : 135
principe du trait Marque (HPSG) : 146
principe du trait Spec (HPSG) : 146
principe sémantique (HPSG) : 145
projection : 133

Q

quasi-arbre : 150

R

réceptif : 94; 96
référent : 93; 96
régime : 18; 19; 20
règles : 201; 206
 règles b : 207
 règles b atomiques : 207
 règles b non atomiques : 207
règles de dominance immédiate : 130
règles de précédence linéaire : 130
règles de réécriture : 135
règles f : 207
règles ID : 130
 règles ID lexicales : 130
règles t : 208
règles tf : 208
relation :
 relation d'ordre : 72
 relation de connexion : 72
 relation directive : 68
 relation horizontale : 71
 relation locative : 23; 68

relation verticale : 71
relations linéaires : 71
relations non linéaires : 71
résultat : 87; 93
rôles d'orientation : 93
rôles de changement : 93
rôles spécifiques : 93

S

schémas de dominance immédiate : 144
signe de ponctuation : 210
situatif : 94; 96
slash : 149
sociatif : 90
source : 87; 88; 93; 96
sous : 19; 20; 25; 30; 40; 42; 43; 224; 226
sous-catégorisation : 125; 168
statif : 94; 96
structure :
 structure complexe : 106
 structure de traits : 101; 105; 148
 structure de traits complexe : 119
 structure de traits cyclique : 105
 structure de traits réentrantes : 109
 structure non réentrante : 106
 structure récursive : 106
 structure réentrante : 106
 structure c : 122; 123; 124
 structure f : 122; 123; 124
subsomption : 107
substitution : 149; 151; 152
sujet : 93
sur : 19; 20; 25; 27; 28; 30; 40; 42; 43; 56;
 223; 224; 225; 226

T

TAG : 102; 149; 192
TAG « synchrones » : 156
temps : 88
thème : 88
token : 213
topic : 125
trait : 105
 trait activé : 154
 trait inhibé (ou virtualisé) : 154
 traits « d'accord » : 154
 traits « de pied » : 154
 traits « de tête » : 154
 traits afférents : 151; 152

 traits aval : 151; 152
 traits inhérents : 154
trajectoire : 42
trajector : 11
transfert casuel : 90; 91
transformation : 211
translateur : 199
 translateurs normaux : 205
 translateurs robustes : 205

U

über : 59; 61; 63; 64; 65; 69; 70
UCG : 103; 110; 115; 117
um : 63; 64; 65; 69; 70; 71
unification : 101; 119; 130; 143
union prioritaire : 108
unter : 59; 61; 63; 64; 65; 69; 70; 224
unterhalb : 67

V

valeurs fixes : 204
variable : 202
véhicule : 93; 96
verbes de déplacement : 29
verbes de mouvement dirigé : 28
verbes de mouvement non dirigé : 28
verbes de mouvement : 28; 29
verbes duratifs : 28
verbes inchoactifs : 28
verbes perfectifs : 28
vers : 30; 55
vocatif : 84
von : 61; 225
vor : 59; 61; 62; 64; 65; 69; 70
voran : 64

Z

zu : 65; 66; 67; 68; 218; 219; 220; 221;
 223; 225
zwischen : 62; 64; 70; 225

- [Abeillé93] A. ABEILLE : *Les nouvelles syntaxes : grammaires d'unification et analyse du français*, Colin, Paris, 1993.
- [André86 et al.] E. ANDRE, G. BOSCH, G. HERZOG, T. RIST : « Characterizing Trajectories of Moving Objects Using Natural Language Path Descriptions », in *Proc. of the 7th ECAI*, vol. 2, pp. 1-8, Brighton, UK, 1986.
- [Balibar-Mrabti90] A. BALIBAR-MRABTI : « Analyse d'adverbes en *dans* », in *Sur les compléments circonstanciels, Langue Française* 86, pp. 65-74, Larousse, Paris, 1990.
- [Baschung91] K. BASCHUNG : *Grammaires d'unification à traits et contrôle des infinitives en français*, Adosa, Clermont-Ferrand, 1991.
- [Baylon95] C. BAYLON, X. MIGNOT : *Sémantique du langage, Initiation*, Nathan Université, Paris, 1995.
- [Berthonneau91] A.-M. BERTHONNEAU, P. CADIOT : « Prépositions, représentations, référence », in *Langue Française*, 32, Larousse, Paris, 1991.
- [Berthonneau93] A.-M. BERTHONNEAU, P. CADIOT : *Les prépositions : méthodes d'analyse*, Presses Universitaires de Lille, Lille, 1993.
- [Borillo90] A. BORILLO : « À propos de la localisation spatiale », in *Sur les compléments circonstanciels, Langue Française* 86, pp. 75-84, Larousse, Paris, 1990.
- [Briffault95] X. BRIFFAULT, M. DENIS : *Analyse d'un Corpus de Dialogues de Navigation à Bord d'un Véhicule Automobile*, LIMSI-CNRS, Orsay, 1995.
- [Butler92] C.-S. BUTLER : *Computer & Written Texts*, Blackwell, Oxford (United Kingdom), Cambridge (USA), 1992.
- [Cadiot91] P. CADIOT : « "Train" et ses prépositions. Modes de donation du référent et principes cognitifs », in *Cahiers de Lexicologie* 58, pp. 63-79, Didier Édition, Paris, 1991.
- [Cervoni90] J. CERVONI : « Prépositions et compléments prépositionnels », in *Sur les compléments circonstanciels, Langue Française* 86, pp. 85-89, Larousse, Paris, 1990.
- [Cervoni91] J. CERVONI : *La préposition. Étude sémantique et pragmatique*, Duculot, Paris / Louvain-la-Neuve, 1991.
- [Clas94] A. CLAS, P. BOUILLON : *TA-TAO : Recherches de pointe et applications immédiates*, Actes du Colloque de Montréal 1993, Montréal, 1994.
- [Cloutier96] Y. CLOUTIER : *Les terminauteries câblées. Bienvenue aux terminantes*, dictionnaire en ligne sur Internet, dernière mise à jour le 17 juin 1996.
- [Courtois90] B. COURTOIS : « Un système de dictionnaires électroniques pour les mots simples du français », in *Dictionnaires électroniques du français, Langue Française* 87, pp. 11-22, Larousse, Paris, 1990.
- [Creissels95] D. CREISSELS : *Éléments de syntaxe générale*, Presses Universitaires de France, Paris, 1995
- [Dirven89] R. DIRVEN, R.-A. GEIGER : *A User's Grammar of English : Word, Sentence, Text, Interaction*, Part B, The Structure of Sentences, Peter Lang Verlag, Frankfurt am Main, 1989.

- [Duden72-4] DUDEN : *Der Grosse Duden*, Band 4, Bibliographisches Institut AG, Mannheim, 1972.
- [Duden72-9] DUDEN : *Der Grosse Duden*, Band 9, Bibliographisches Institut AG, Mannheim, 1972.
- [Duden89] DUDEN : *Deutsches Universal Wörterbuch A-Z* Dudenverlag Mannheim, Wien, Zürich, 1989.
- [Dugas90] A. DUGAS : « La création lexicale et les dictionnaires électroniques », in *Dictionnaires électroniques du français, Langue Française* 87, pp. 23-28, Larousse, Paris, 1990.
- [Egli91] U. EGLI : *(In)definite Nominalphrase und Typentheorie*, Arbeitspapier 27, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1991.
- [Egli91] U. EGLI, R. EGLI-GERBER : *Sprachsysteme - logische und historische Grundlagen der erweiterten Phrasenstrukturgrammatik*, Arbeitspapier 28, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1991.
- [Egli92] U. EGLI, K. VON HEUSINGER : *Das Patr-II Format in Anwendung auf das Deutsche*, Arbeitspapier 42, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1992.
- [Fauconnier 96] G. FAUCCONNIER, E. SWEETSER : *Spaces, worlds and grammar*, University of Chicago Press, London, 1996.
- [Fries88] N. FRIES : *Präpositionen und Präpositionalphrasen im Deutschen und im Neugriechischen*, Max Niemeyer Verlag, Tübingen, 1988.
- [Fuchs92] C. FUCHS, P. LE GOFFIC : *Les Linguistiques Contemporaines, Repères théoriques*, Hachette Supérieur, Paris, 1992.
- [Fuchs93] C. FUCHS : *Linguistique et traitements automatiques des langues*, Hachette, Paris, 1993.
- [Gardent94] C. GARDENT, K. BASCHUNG : *Techniques d'analyse et de génération pour la langue naturelle*, éd. Adosa, Clermont-Ferrand, 1994.
- [Gazdar85] G. GAZDAR, E. KLEIN, G. K. PULLUM, I. SAG : *Generalized Phrase Structure Grammar*, Basil Blackwell, Oxford, 1985.
- [Geske96] D. GESKE : *Das maschinelle Übersetzungssystem CAT2*, Saarbrücken, 1996.
- [Grappin89] P. GRAPPIN : *Grand dictionnaire français-allemand*, Larousse, Paris, 1989.
- [Grévisse93] M. GREVISSE : *Le Bon Usage*, Duculot, treizième édition refondue par André Goosse, Paris / Louvain-la-Neuve, 1993.
- [Greimas93] A.-J. GREIMAS, J. COURTES : *Sémiotique, dictionnaire raisonné de la théorie du langage*, Hachette Supérieur, Paris, 1993.
- [Gross75] M. GROSS : *Méthodes en syntaxe. Régime des constructions complétives*, Hermann, Paris, 1975.
- [Günthner95] S. GÜNTNER, T. LUCKMANN : *Asymmetries of Knowledge in Intercultural Communication : The Relevance of Cultural Repertoires of*

- Communicative Genres*, Arbeitspapier 72, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1995.
- [Heusinger91] K. VON HEUSINGER : *Kategoriale Unifikationsgrammatik*, Arbeitspapier 37, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1991.
- [Heusinger93] K. VON HEUSINGER : *Eselssätze und ihre Pferdefüße. Probleme der Nominalphrasensemantik und ihre Repräsentation*, Arbeitspapier 64, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1994.
- [Heusinger94] K. VON HEUSINGER : *Definite Kennzeichnungen, Anaphora und Salienz*, Arbeitspapier 66, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1994.
- [Hörnig96] R. HÖRNIG, B. CLAUS, D. DURSTEWITZ, E. FRICKE, U. SCHMID, K. EYFERTH : *Object Access in Mental Models under Different Perspectives Induced by Linguistic Expressions*, Kit-Report 126, Berlin, 1996.
- [Jayez82] J.-H. JAYEZ : *Compréhension automatique du langage naturel : le cas du groupe nominal en français*, Masson, Paris, 1982.
- [Kruypers96] J.-P. KRUYERS, G. BRAND : *Dictionnaire de néologie Internet*, version 2.1.6.1., dernière révision le 22 juillet 1996.
- [Larousse82] LAROUSSE : *Grand Dictionnaire Universel du XIXème siècle*, Slatkine, Genève, Paris, 1982.
- [Larousse86] *Grand Larousse de la Langue Française en 7 volumes*, tome 5 : O-PSI, Librairie Larousse, Paris, 1986.
- [Larousse97] *Le petit Larousse*, Larousse, Bordas, Paris, 1997.
- [Lauvray92] J. LAUVRAY : *Les verbes pronominaux : étude théorique et appliquée à la Traduction Assistée par Ordinateur*. Expérimentation sur trois langues, Thèse de doctorat de l'Université de Metz, Metz, 1992.
- [Le Bidois71] G. & R. LE BIDOIS : *Syntaxe du français moderne. Ses fondements historiques et psychologiques*, 2^e édition revue et complétée, édition Picard, Paris, 1971.
- [Le Goffic93] P. LE GOFFIC : *Grammaire de la phrase française*, Hachette Supérieur, Paris, 1993.
- [Létoublon90] F. LETOUBLON : « Sur les genoux des Dieux », in *Sur les compléments circonstanciels*, *Langue Française* 86, pp. 22-29, Larousse, Paris, 1990.
- [Lindqvist94] C. LINDQVIST : *Zur Entstehung von Präpositionen im Deutschen und Schwedischen*, Max Niemeyer Verlag, Tübingen, 1994.
- [Lyons70] J. LYONS : *Linguistique générale. Introduction à la linguistique théorique*, Larousse, Paris, 1970.
- [Maillot93] C. MAILLOT, M. DORVAUX : *Traduction automatique des mots composés allemands en français : étude théorique et implémentation*, projet de recherche du D.E.S.S. « Industries de la Langue », Metz, Sarrebruck, 1993.
- [Malchow93] A. MALCHOW : *Syntax and Semantics of Construction*, Arbeitspapier 56, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1993.
- [Marcq72] P. MARCQ : *Prépositions spatiales et particules « mixtes » en allemand*, Vuibert, 1972.

- [Maurel90] J.-P. MAUREL : « Examen circumstantiarum », in *Sur les compléments circonstanciels, Langue Française* 86, pp. 30-35, Larousse, Paris, 1990.
- [McConnel95] S. MCCONNEL : *PC-PATR Reference Manual*, Dallas, 1995.
- [Miller-Torris90] P. MILLER & T. TORRIS : *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Hermès, Paris, 1990.
- [Moilanen79] M. MOILANEN : *Statische lokative Präpositionen im heutigen Deutsch. Wahrheits- und Gebrauchsbedingungen*, Max Niemeyer Verlag, Tübingen, 1979.
- [Mounin74] G. MOUNIN : *Dictionnaire de la linguistique*, Presses Universitaires de France, Paris, 1974.
- [Perousseaux95] Y. PEROUSSEAUX : *Manuel de typographie française élémentaire*, Atelier Perousseaux éditeur, Reillanne, 1995.
- [Perousseaux96] Y. PEROUSSEAUX : *Mise en page et impression, notions élémentaires*, Perousseaux, Reillanne, 1996.
- [Phelizon76] J.-F. PHELIZON : *Vocabulaire de la linguistique*, édition Roudil, Paris, 1976.
- [Platt71] J.-T. PLATT : *Grammatical Form and Grammatical Meaning, a tagmemic view of Fillmore's deep structure case concepts*, North-Holland Publishing Company, Amsterdam, London, 1971.
- [Pottier62] B. POTTIER : *Systématique des éléments de relation*, Klincksieck, Paris, 1962.
- [Pottier92] B. POTTIER : *Théorie et analyse en linguistique*, Hachette Supérieur, Paris, 1992.
- [Ramm91] W. RAMM : *Grundlagen für eine wissensbasierte Maschinelle Übersetzung*, IAI Working Paper N° 22, Saarbrücken, 1991.
- [Ramm 92] W. RAMM : *Conceptual Information and the Interpretation of Spatial Prepositions : A Comparison Between Approaches to Systemic-Functional Text Generation and Cognitive Grammar Based Machine Translation*, IAI Working Paper N° 25, Saarbrücken, 1992.
- [Rastier94 et al.] F. RASTIER, M. CAVAZZA, A. ABEILLE : *Sémantique pour l'analyse, De la linguistique à l'informatique*, Masson, Paris / Milan / Barcelone, 1994.
- [Reboul90] S. REBOUL : « A la Guadeloupe/en Guadeloupe », in *Le lexique : construire l'interprétation, Langue Française* 103, pp. 68-78, Larousse, Paris, 1990.
- [Robert92] LE ROBERT : *Dictionnaire Historique de la langue française*, Paris, 1992.
- [Robert66] P. ROBERT : *Dictionnaire alphabétique et analogique de la langue française*, Tome 5, Le Robert, Paris, 1966.
- [Rothkegel89] A. ROTHKEGEL : « Polylexicalität. Verb-Nomen-Verbindungen und ihre Behandlung in EUROTRA », in *IAI Working Papers* N° 17, Saarbrücken, 1989.
- [Sabah90] G. SABAH : *L'intelligence artificielle et le langage, Représentations des connaissances*, Vol. 1, Hermès, Paris, 1988, 2ème éd. 1990.
- [Schanen89] F. SCHANEN, J.-P. CONFAIS : *Grammaire de l'allemand. Formes et fonctions*, Nathan, Paris, 1989.

- [Schleier75] I. E. H. SCHLEIER : *Das Funktionssystem der Präpositionalkonstruktion in der deutschen Gegenwartssprache*, H. Buscke Verlag, Hamburg, 1975.
- [Schmitz93] B. SCHMITZ, J. J. QUANTZ : *Defaults in Machine Translation*, Kit Report 106, Technische Universität Berlin, 1993.
- [Schwarze81] C. SCHWARZE : *Analyse des prépositions*. III ème Colloque franco-allemand de linguistique théorique du 2 au 4 février 1981 à Constance, Max Niemeyer Verlag, Tübingen, 1981.
- [Schweisthal71] K. G. SCHWEISTHAL : *Präpositionen in der maschinellen Sprachbearbeitung*, Ferd Dümmlers Verlag, Bonn, 1971.
- [Schwertel93] U. SCHWERTEL : *Probleme der Pluralsemantik*, Arbeitspapier 58, Fachgruppe Sprachwissenschaft, Universität Konstanz, 1993.
- [Sharp94] R. SHARP : *CAT2 Reference Manual*, IAI Working Paper 27, Saarbrücken, 1994.
- [Steiner86 et al.] E. STEINER, U. ECKERT, B. WECK, J. WINTER : « The Development of the EUROTRA-D System of Semantic Relations », in *IAI Working Papers N° 2*, Saarbrücken, 1986.
- [Stopp94 et al.] E. STOPP, K.-P. GAPP, G. HERZOG, T. LÄNGLE, T.-C. LÜTH : « Utilizing Spatial Relations for Natural Language Access to an Autonomous Mobile Robot », in B. Nebel, I. Dreschler-Fischer (eds.) : *Advances in Artificial Intelligence*, 18th German Annual Conference on Artificial Intelligence, pp. 39-50, Springer Verlag, Berlin, Heidelberg, 1994.
- [Streiter95] O. STREITER : *Linguistic Reference Manual for the CAT2 Machine Translation System*, Saarbrücken, 1995.
- [Streiter96] O. STREITER : *Linguistic Modelling for Multilingual Machine Translation*, Shaker Verlag, Aachen, 1996.
- [Taylor89] J.-R. TAYLOR : *Linguistic categorization. Prototypes in Linguistic Theory*, Clarendon Press, Oxford, 1989.
- Trésor de la Langue Française. Dictionnaire de la langue du XIXème et du XXème siècle*. Centre National de la Recherche Scientifique. Tome 13, Gallimard, Paris, 1988.
- [Uszkoreit89] H. USZKOREIT : « Word Order and Constituent Structure in German », in *Language* 65(1), pp. 141-149, 1989.
- [Vandeloise88] C. VANDELOISE : « L'expression du mouvement », in *Langue Française* 76, Larousse, Paris, 1988.
- [Volk88] M. VOLK : *Parsing German with GPSG : The Problem of Separable Prefix Verbs*, Thesis Submitted to the Graduate Faculty of The University of Georgia, Athens, Georgia, 1988.
- [Weisweber94] W. WEISWEBER : *Installations- und Benutzungshandbuch des experimentellen MÜ-Systems des Projekts Kit-Fast*, Kit-Report 116, Berlin, 1994.
- [Willmet72] M. WILMET : *Gustave Guillaume et son école linguistique*, Nathan, Paris, Labor, Bruxelles, 1972.

[Zelinsky-Wibbelt93] C. ZELINSKY-WIBBELT : *The Semantics of Prepositions. From Mental Processing to Natural Language Processing*. Mouton de Gruyter, Berlin, New York, 1993.