

# **Topologie et grammaires formelles de l'allemand**

**Thèse sous la direction d'Anne Abeillé  
pour l'obtention du diplôme de docteur de l'Université Paris 7 en Linguistique**

**présentée et soutenue publiquement par Kim Gerdes  
le 16 décembre 2002**

Ecole doctorale Science du langage, UFR de linguistique, Université Paris 7

membres du jury :  
Werner Abraham, Vienne  
Christian Boitet, Grenoble  
Igor Mel'čuk, Montréal  
Owen Rambow, New York



## **Remerciements**

*Certaines personnes trouvent les remerciements au début d'une thèse tout particulièrement intéressants (comme l'affirme par exemple Detmar Meurers 2000 au début de sa thèse) ; moi je ne fais pas partie de ces personnes et si vous décidez ici de tourner la page, je peux vous prévenir que vous ne ratez rien.*

*Bien sûr, les remerciements permettent de dévoiler quelques idées sur l'auteur, son entourage, les influences qu'il a subies. Souvent, il arrive que les remerciements influencent la lecture qui suit : savoir ce qui est à l'origine de chaque idée prend le dessus sur la question de la teneur en vérité de cette idée.*

*La deuxième raison pourquoi je n'aime pas les remerciements est que le post-scriptum « c'était dur mais sympa » ne correspond pas à l'expérience de ces quatre années passées : pour moi, c'était sans aucun doute les années les plus dures de ma vie, marquées par l'horrible maladie qui a touché ma mère juste six semaines avant que je m'inscrive en thèse. Plus marquant encore que la recherche que j'ai pu réaliser était le temps passé avec ma mère devenue gravement handicapée, inconsciente du temps et de son état, mais capable encore de corriger mes mauvaises citations des poèmes de Goethe ou mon français et même de donner son avis sur l'acceptabilité des mes exemples tordus en allemand.*

*Si j'ai réussi à mener à bien cette thèse malgré ces circonstances, cela a été possible grâce à l'aide de beaucoup de gens, et puisque c'est le moment et le seul endroit où je puisse le faire, je voudrais profiter de ces quelques pages pour les remercier, et ce en me mettant en contradiction avec mes idées de départ.*

*Je voudrais commencer par remercier ma tante, Imme Odenwälder pour se dévouer et s'occuper si bien tous les jours de ma mère. Sans elle, je n'aurais jamais trouvé de temps pour ma recherche.*

*Je voudrais remercier chaleureusement ma directrice Anne Abeillé, avant tout pour sa grande compréhension de ma situation et la confiance qu'elle m'a faite, malgré ma déviation par rapport au sujet de départ. Je la remercie également pour ses remarques, rares mais pertinentes et ciblées qui m'ont aidé à corriger un grand nombre d'erreurs dans ce présent travail.*

*Si Anne Abeillé est ma directrice, Sylvain Kahane est mon gourou. Il est sans doute la personne la plus influente sur mes travaux et beaucoup des idées présentes dans cette thèse sont nées lors de nos nombreuses discussions et journées de travail. Peut-être c'est grâce à l'approche commune de mathématiques convertis en linguistique qui nous a permis de travailler de manière si fructueuse ; peut-être c'est juste l'amitié qui s'est créée entre nous.*

*Mes bibles, par contre, usées aujourd'hui pour s'être trop souvent retrouvés entre mes mains durant ces quatre ans, ont été écrites par Stefan Müller 1999 et Andreas Kathol 1995. Un grand merci à Stefan Müller et surtout à Tilman Becker qui m'ont accueilli chez eux pendant mes fréquentes visites au DFKI ; leurs réponses à mes questions souvent naïves sur la syntaxe de l'allemand et sur les TAG m'ont donné l'impression d'en avoir compris quelque chose, surtout aux heures souvent tardives de nos discussions.*

*Outre-Rhin encore, Detmar Meurers et Laura Kallmeyer m'ont invité à Tübingen. Detmar, tout en me déconseillant de travailler sur la syntaxe verbale de l'allemand, a mis à ma disposition une pile d'articles à lire et une carte de photocopies. Ces articles, représentent la base de ma bibliographie.*

*Mes remerciements se dirigent cette fois-ci outre-Atlantique, à Igor Mel'čuk et Lydia Iordanskaja qui m'ont chaleureusement accueilli chez eux à Montréal. Bien qu'Igor ne voit toujours pas en moi un futur linguiste, je suis convaincu que les nombreuses discussions toujours passionnées et passionnantes que j'ai pu avoir avec lui m'ont approché de ce but auguste. Je le remercie d'avoir accepté d'être non seulement dans le jury mais également le rapporteur de thèse d'un mathématicien.*

*Je tiens également à remercier Aravind Joshi et ses adeptes, Anoop Sarkar, Chung-Hye Han, Carlos Prolo, Tonia Bleam, David Chiang et Alexandra Kinyon pour leur accueil et les nombreuses discussions sur les TAG, « la vie, l'univers et tout le reste » durant mes séjours à l'Institut de recherche en Sciences cognitives à Philadelphie et pendant leurs visites à Paris. Ma reconnaissance va tout particulièrement à Owen Rambow qui a accepté d'être mon rapporteur, malgré les contraintes de temps que je lui ai fait subir.*

*Werner Abraham m'a expliqué pendant deux soirées dans les montagnes inoubliables de Banff, les idées de la théorie GB ; il m'a presque convaincu. Je le remercie d'avoir accepté d'être dans mon jury de thèse et j'espère pouvoir le convaincre à mon tour du lieu d'être de ce qu'il appelait à l'époque ma « poubelle de mots ».*

*A Nancy, les discussions que j'ai pu avoir avec les membres du Loria m'ont aidé à développer l'aspect informatique de mon travail. Je tiens particulièrement à remercier Azim Roussanaly pour son accueil chaleureux à plusieurs reprises et aussi Bertrand Gaiffé pour son très bon support technique de son compilateur de méta-grammaires. Souvent, ses courriers électroniques étaient intitulés « Pas de panique ! »... Last but not least, il y a Patrice Lopez qui a souvent pris le train entre Sarrebruck et Nancy pour m'expliquer TAG, XML et d'autres raccourcis vers la vérité.*

*Au sud cette fois-ci, j'aimerais remercier Christian Boitet d'avoir soutenu ma candidature d'ATER à Grenoble et d'avoir accepté d'être dans mon jury de thèse.*

*A Paris, les séminaires sur la linguistique japonaise d'Irène Tamba et de Raoul Blin à l'EHESS et les discussions souvent enflammées qui suivaient étaient une bouffée d'air exotique m'apportant des contre-exemples dans cette langue étonnante pour toute affirmation linguistique imaginable.*

*Dans un entourage plus proche, je tiens à remercier les membres de l'équipe Talana. Pendant ces quatre années passées ensemble, entre engueulades, discussions scientifiques sérieuses, blagues et tests à l'aveugle de vins, Talana a été un second chez moi. Je voudrais tout particulièrement remercier Laurence Danlos de m'avoir accueilli et de m'avoir permis de m'inscrire en DEA de linguistique et Pascal Amsili pour ses précieux conseils lors de mes débuts dans l'enseignement. Et bien sûr, autres talaniens, collègues et amis, tout particulièrement Véronique Gendner, Lionel Clément, Adil El Ghali. Le dernier venu, Maître François Lareau a gentiment relu ma thèse et a introduit ça et là quelques québécoisismes.*

*Je tiens également à remercier Georges Boulakia du laboratoire de phonétique de notre UFR, pour sa générosité et son amitié : sans le prêt d'un des ordinateurs du laboratoire, je n'aurais jamais réussi à correctement faire tourner mes programmes.*

*Un étage plus haut, Muriel Geffriaud, Jeanine Mary m'ont soutenu et m'ont sauvé des paperasses incompréhensibles de l'administration française. Mon merci s'adresse également à Karen Ferret, autre occupante des locaux du 9<sup>ème</sup> étage, pour les discussions que nous avons eues sur la syntaxe de l'allemand.*

*J'espère que cette liste est suffisamment confuse pour éviter un classement facile et suffisamment longue pour que chacun s'y retrouve. Les rencontres avec des gens d'horizons linguistiques divers se sont passés comme mes voyages dans les pays « exotiques » : parti à la recherche de « l'exotisme », j'ai trouvé derrière une mince couche d'étrangeté des gens qui ont les mêmes idées et se posent les mêmes questions que moi.*

*Je ne crois pas à la « culture », ni d'un pays ni d'un courant linguistique, en particulier syntaxique.*

# Sommaire

<i>Introduction</i>	<i>1</i>
<i>L'allemand</i>	<i>7</i>
1 Les cas de l'allemand	8
2 Actant sémantique, actant syntaxique et argument syntaxique	10
3 Le passif	19
4 Prépositions et arguments syntaxiques	24
5 Le datif libre	27
6 La notion d'ordre libre des mots	32
<i>Topologie</i>	<i>51</i>
1 Les règles de base de la topologie	53
2 Différentes structures verbales de la topologie	57
3 Structure du complexe verbal	65
4 Complétives et Relatives	71
5 Sur les traces du modèle topologique	93
6 La notion de syntagme topologique	96
7 La place de la structure syntagmatique dans un modèle linguistique	102
<i>TAG</i>	<i>109</i>
1 Les grammaires d'arbres adjoints lexicalisées	111
2 Critiques et extensions	120
3 Les métagrammaires pour les TAG	150
4 DTAG	170
5 Remarques conclusives sur DTAG	208
<i>Formalisations et extensions de la topologie</i>	<i>211</i>
1 Formalisation d'une analyse topologique	212
2 Le problème du sujet dans un domaine enchâssé	219
3 Grammaires d'unification d'arbres (TUG)	228
4 LFG	236
5 L'ordre dans le Mittelfeld – un phénomène topologique ?	246
6 La topologie du syntagme nominal allemand	256
7 Topologie comparative : coréen, grec, français	261
<i>Conclusion</i>	<i>271</i>
<i>Références bibliographiques</i>	<i>275</i>
<i>Table des figures</i>	<i>285</i>
<i>Table des matières</i>	<i>290</i>



# Introduction

# Introduction

*Neque enim necesse est eas hypotheses esse veras, imo, ne verisimiles quidem; sed sufficit hoc unum, si calculum observationibus congruentem exhibeant.*<sup>1</sup>

## **Du Sens à la forme phonétique de la phrase**

Une idée qui est acquise aujourd'hui dans la communauté linguistique, est que les processus langagiers lient un niveau de représentation « intentionnel-conceptuel » (Chomsky 1995) et un niveau de représentation « articulatoire-perceptuel » (*ibid.*). Autrement dit, une des recherches essentielles en linguistique est de modéliser le passage entre ces deux niveaux de représentation, et pourtant, il n'y a aucun consensus sur le processus entre ces deux niveaux. En effet, on est loin de disposer d'un modèle unique, et finalement, très peu de notions et de définitions sont universellement reconnues.

Une des rares notions qui continuent à subsister partout dans cette diversité théorique est l'idée de gouvernance syntaxique. Prenons comme exemple le modèle transformationnel de Chomsky. Cette notion joue un rôle primordial dans la définition de la structure X-barre, la colonne vertébrale de ce modèle. Or le terme de gouvernance telle qu'elle est définie par Chomsky regroupe non seulement la gouvernance syntaxique, mais aussi la gouvernance sémantique ainsi que le regroupement des constituants. Ceci peut donner lieu à des confusions et nous considérons qu'il convient à séparer ces différents éléments, tout en explicitant les liens qui existent entre eux.

Cette idée de séparation stricte entre les différentes structures renvoyant à des types d'information différents est d'autant plus essentielle si l'on veut construire un modèle tel qu'on l'entend dans des sciences comme la physique ou les mathématiques. Il est en effet étonnant de constater que la question de séparabilité de l'information n'est pas au cœur des discussions sur l'adéquation d'un modèle par rapport à un autre.

D'une manière concrète, dès que les analyses font surgir une différence entre d'un côté le lien qu'entretient un gouverneur sémantique avec son dépendant, et de l'autre le régime d'un gouverneur syntaxique, nous devons inclure deux structures dans notre modèle, tout en posant le lien qui existe entre ces deux structures. En effet, une représentation des structures de types différents dans une même structure ne ferait qu'embrouiller la compréhension du fonctionnement de l'ensemble.

A ma connaissance, le plus fervent défenseur de cette approche qui prône une séparation des structures est Igor Mel'čuk qui a développé et continue à le faire depuis quarante ans la théorie Sens-Texte. Mel'čuk pose comme base de son modèle la distinction des différentes structures

---

<sup>1</sup> « Pour être sûr, il n'est ni nécessaire que ces hypothèses soient vraies ni même qu'elles soient finalement probables. La seule chose qui nous suffit est que le calcul se montre congruent aux observations. » (notre traduction) Osiander von Nürnberg, introduction aux oeuvres de Copernic, *De revolutionibus caelestibus libri sex*, 1543



intermédiaires entre le « sens »<sup>2</sup> (le niveau « intentionnel-conceptuel » selon les termes de Chomsky 1995) et celui de « texte »<sup>3</sup> (le niveau « articulatoire-perceptuel » selon Chomsky 1995). De plus, Mel'čuk opte pour un modèle linéaire de la langue<sup>4</sup> où la structure d'entrée constituée par le sens est successivement transférée dans une (ou plusieurs) représentation(s) des niveaux syntaxiques, puis morphologiques, pour aboutir à une sortie finale qui correspond à la réalisation concrète de l'énoncé, traitée par un niveau phonétique.

sémantique ↔ syntaxe ↔ morphologie ↔ phonétique

**Figure 1 : les différentes structures de la théorie Sens-Texte**

De plus, les trois derniers niveaux sont dédoublés en une structure qui est plutôt orientée vers le sens, appelée « structure profonde », et en une structure orientée vers le texte, « la structure de surface »<sup>5</sup>. L'idée linéaire est conservée, comme on peut le voir dans la Figure 2.

sens ~ sémantique ↔ syntaxe profonde ↔ syntaxe de surface ↔ morphologie profonde  
 ↔ morphologie de surface ↔ phonologie profonde ↔ phonétique de surface ~ texte

**Figure 2 : les différents structures de la théorie Sens-Texte**

Notons que dans la théorie Sens-Texte, on considère le langage comme une production langagière et on se place donc plutôt du point de vue de la génération (synthèse), bien que toutes les correspondances qui sont établies entre les différents niveaux sont bidirectionnels, et peuvent donc s'appliquer aussi bien en génération qu'en analyse (compréhension).

Dans les structures sémantiques et syntaxiques, les éléments sont non ordonnés linéairement et ce n'est que lors du passage du niveau syntaxique au niveau morphologique qu'on considère les structures ordonnées. Du point de vue de la représentation, la structure sémantique et la structure syntaxique profonde sont représentés respectivement comme un graphe et un arbre de dépendance sémantique, où les lignes qui relient les nœuds indiquent, non leur ordre respectif, mais la relation qu'ils entretiennent entre eux. Une telle représentation diffère donc de l'arbre des constituants dont le rôle est d'indiquer justement l'ordre des différents éléments représentés dans l'arbre.

La question de l'ordre des éléments qui nous semble primordiale si l'on s'intéresse aux phrases que nous obtenons concrètement à la sortie du modèle reste pourtant le parent pauvre des grammaires de dépendance. On s'aperçoit que très souvent, les approches de linéarisation des arbres de dépendance sont simplistes dans le sens qu'on stipule la projectivité de l'arbre de dépendance, bien que ceci implique qu'on perde tout l'avantage d'expressivité que nous gagnons

<sup>2</sup> En fait, il pose encore un niveau « conceptuel » qui se place avant tous les autres niveaux et qui consiste en une représentation du concept même qui est indépendant de la langue. Pour Mel'čuk 1988, ce niveau ne fait pas partie de la langue et la linguistique s'occupe de ce qui est spécifique à chaque langue.

<sup>3</sup> Bien que la notion de texte renvoie souvent à l'idée d'un enchaînement cohérent de plusieurs phrases, notre analyse se limite à la phrase sans prendre en compte le contexte.

<sup>4</sup> Cette idée de linéarité de la langue nécessite une justification séparée, que l'on discutera dans la section II.7. Cette représentation de la langue n'exclut pas qu'il s'agit d'un traitement parallèle.

<sup>5</sup> Dans la section I.2, nous tenterons de justifier la nécessité de ce dédoublement des structures pour le niveau syntaxique.

en proposant un arbre de dépendance non linéaire. En effet, un arbre de dépendance projective avec sa projection comporte la même information qu'un arbre syntagmatique X-barre<sup>6</sup>.

Finalement, les grammaires de dépendance ne font que jeter le bébé avec l'eau du bain quand, dans leur refus de l'analyse X-barre, elles effacent toute idée d'arbre de constituants. Et pourtant, un arbre de constituants représente un outil puissant pour exprimer les règles de linéarisation d'une structure de dépendance, d'autant plus que les éléments de cet arbre de constituants sont des éléments concrets, « visibles » en tant que regroupements prosodiques à la surface.

## **La liberté d'ordre en allemand**

L'un des objectifs majeurs du travail que je présente ici est d'utiliser l'idée d'arbres de constituants dans une approche basée sur la dépendance. Nous nous intéresserons particulièrement au passage où une représentation d'arbre de dépendance non ordonnée se linéarise, en proposant notamment de remplacer le niveau morphologique profond habituel de la théorie Sens-Texte par un niveau que nous appelons la *structure topologique*. Nous développerons cette approche plus particulièrement pour l'allemand, une langue qui permet beaucoup de linéarisations pour une structure de dépendance donnée.

Afin d'éclaircir l'idée de « l'ordre libre » il est important de distinguer ce qui reste stable de ce qui varie. Bien que les différents ordres linéaire de l'allemand ont été sujet de beaucoup de travaux dans différentes approches linguistiques, il est difficile de trouver une étude complète sur l'ensemble des ordres linéaires possibles pour une structure profonde quelconque donnée, car la plupart des approches partent de la phrase pour aboutir à une analyse et ne se soucient guère de la démarche inverse, c'est-à-dire obtenir tous les ordres des mots à partir d'une même structure profonde. Les phénomènes d'ordres des mots de l'allemand constituent un défi pour la plupart des formalismes linguistiques. Par exemple dans le cadre de la Théorie du Gouvernement et de Liage, on a rencontré des problèmes de représenter certains ordres des mots, qui par la suite ont été traité comme des phénomènes à part sous les noms de *scrambling* (*brouillage d'arguments*), *VP-fronting* (*partiel*), et *l'intrapolition*. Ces appellations ont laissé leurs empreintes même sur des analyses qui se démarquent explicitement du cadre de la Théorie du Gouvernement et de Liage. Ainsi, les analyses traditionnelles dans des formalismes comme la Grammaire lexicale fonctionnelle (LFG), la Grammaire syntagmatique guidée par les têtes (HPSG) et la Grammaire d'arbres adjoints (TAG) traitent ces phénomènes aussi de manière isolée, en ne recherchant pas une explication globale. Ces différentes analyses font apparaître certains ordres des mots de l'allemand comme une entorse linguistique. Ces ordres problématiques ont ainsi inspiré plusieurs extensions des formalismes existants, en particulier pour les TAG (Rambow 1994) et pour HPSG.

Pourtant, il n'est pas clair « pourquoi l'anglais refléterait mieux que l'allemand la grammaire universelle quelque soit le sens qu'on donne à ce terme »<sup>7</sup>. C'est dans cet esprit que Kathol 1995 rétablit la notion de topologie en lui donnant une place prépondérante dans une grammaire formelle de l'allemand. Il propose une grammaire syntagmatique guidée par les têtes (HPSG) muni d'un système de traits permettant de traiter à part la structure syntaxique et l'ordre des mots. Bien que le cadre de travail de Kathol (HPSG) soit différent du nôtre (grammaires de dépendances), notre travail s'inspire largement de ses idées. Pour la représentation de l'ordre des mots, nous avons comme lui recours à la notion traditionnelle des champs, mais nous allons plus loin en introduisant une structure hiérarchique de ces champs, la structure topologique. Notre intérêt porte en particulier sur la description du lien qu'entretient la structure topologique qui est linéairement ordonnée avec une structure de dépendance non ordonnée, et c'est ce lien que nous

<sup>6</sup> Chomsky introduisait en 1957 le mouvement, justement à cause de l'insuffisance d'une grammaire de constituants immédiats.

<sup>7</sup> texte original « Mir will es nicht einleuchten, dass Englisch in irgendeinem interessanten Sinn die UG besser reflektieren soll als Deutsch ». Haider 1993:142, cité par Kathol 1995:1

décrivons et formalisons dans ce travail : autour de la question de l'ordre des mots en allemand, nous examinerons les données puis proposerons une formalisation en se basant sur la notion de topologie.

En fait, la Théorie Sens-Texte ne peut pas être considérée comme un formalisme en soi. En effet, cette théorie propose différents niveaux d'analyse ainsi que des critères permettant de distribuer l'information d'une phrase dans ces différents niveaux, mais ne cherche pas à donner une formalisation permettant la liaison entre ces différents niveaux<sup>8</sup>. D'ailleurs, il n'est même pas convenu si tous les transferts qui interviennent entre les différents niveaux se font avec les mêmes types d'outils formels<sup>9</sup>, ou s'il faut adapter la formalisation de chaque transfert aux différents types de structures qu'on a comme entrée et comme sortie.

C'est pourquoi nous avons cherché à comparer notre propre approche avec d'autres formalisations qui se basent sur les liens qui existent entre une structure de constituants et une structure de dépendance : nous nous sommes intéressés en particulier au formalisme de Grammaire d'Arbres Adjoints (TAG) de Joshi 1975 et au formalisme de Grammaire lexicale fonctionnelle développée par Bresnan & Kaplan 1982 (LFG).

Concrètement, une partie importante de cette thèse est consacrée au développement et à l'implémentation d'une grammaire TAG lexicalisée électronique de l'allemand. On s'est intéressé tout particulièrement à ce formalisme car les TAG ont des propriétés computationnelles et linguistiques très intéressantes ; bien que leur algorithme d'analyse est d'ordre polynomial, les TAG arrivent à construire à la fois des structures de dépendance et de constituants même pour des phénomènes syntaxiques assez compliqués comme les dépendances de longue distance. De plus, la lexicalisation de la grammaire permet des analyses linguistiques d'une grande finesse et pose des questions théoriques intéressantes. Bien qu'il ait été montré par différents travaux (Becker et Rambow 1992, 1994) que pour l'allemand, une analyse TAG n'arrive pas toujours à faire le lien entre l'ordre linéaire et une dépendance convenable, nous avons cherché à regarder de plus près les problèmes qui se posent. Ceci était important pour connaître les limites pratiques et théoriques d'une TAG de l'allemand et de proposer, à partir de cette analyse, une extension de ce formalisme prenant en compte tous les problèmes rencontrés et proposant de meilleures analyses. En effet, il se trouve que lors de la construction de la grammaire TAG, nous nous sommes aperçu que ses avantages computationnels s'accompagnent d'une limitation dans l'articulation des différentes structures : la capacité générative des TAG ne permet ni d'obtenir ni de lier les structures de dépendance avec les structures de constituants que nous souhaitons générer. C'est ainsi que nous avons été amenés à proposer un formalisme appelé « grammaire d'unification d'arbres » (TUG), qui tente de préserver les avantages de TAG tout en permettant de lier une structure topologique à une structure de dépendance syntaxique<sup>10</sup>.

La structure topologique représente la notion centrale du présent travail. Elle se trouve au croisement de plusieurs approches linguistiques.

Premièrement, elle est née des analyses classiques et descriptives de l'allemand, qui introduisent les notions de « places » ou de « champs » dans les descriptions de l'ordre des mots en allemand, sans pour autant donner à la notion de place, un statut propre à l'intérieur d'une théorie.

---

<sup>8</sup> C'est la raison pour laquelle la question de « complexité de formalisme » n'a pas de sens dans la Théorie Sens-Texte. L'implémentation d'un système Sens-Texte par Wanner & Bohnet 2002 fait des choix sur l'expressivité d'une règle de transfert, mais le système reste le moins restreint possible, ce qui peut avoir des conséquences importantes sur la complexité du calcul des transferts que permet une grammaire donnée.

<sup>9</sup> Comme le propose Wanner & Bohnet 2002 et l'approche GUST de Kahane 2000.

<sup>10</sup> Les TUGs restent pour l'instant dans un stade théorique, et nous n'avons pas encore développé une réelle grammaire dans ce formalisme.

Deuxièmement, la structure topologique représente un outil de linéarisation d'une structure de dépendance et permet de contrôler précisément la « non-projectivité » d'un arbre de dépendance, et donc le degré de non correspondance entre une structure argument-prédicat et l'ordre des mots.

Troisièmement, la structure topologique est une structure de constituants avec cependant une tâche beaucoup moins lourde que son homologue de la théorie X-barre, la structure syntagmatique (II.7).

Enfin, la structure topologique se place dans une optique de l'analyse de la langue orale, et son utilité est partiellement justifiée par ce type d'analyse puisqu'elle peut jouer le rôle d'une des interfaces entre sens et parole : en effet, les constituants contenus dans la structure topologique sont essentiels pour le calcul des groupes prosodiques. Par ailleurs, cette structure peut être étendue à d'autres langues dites à ordres des mots libres comme le grec moderne ou le coréen.

## **Plan de la thèse**

L'objectif de ce travail est double. D'un point de vue théorique, nous proposons une analyse formelle de la question de l'ordre des mots en allemand en se basant sur la notion de topologie. Cette réflexion théorique nous débayera le terrain pour la possibilité d'une implémentation d'une grammaire électronique, d'une part sous le modèle des TAG et d'autre part selon le formalisme que nous mettons en place. L'avantage de cette grammaire électronique par rapport à celles qui existent déjà (cf. chapitre III) repose sur le fait que l'outil de métagrammaire permet de paramétrer de manière efficace cette grammaire.

Notre travail s'organise comme suit. Dans un premier temps, nous faisons une esquisse des points essentiels de la syntaxe de l'allemand. Nous serons amenés à définir les notions de cas, d'actant et d'argument et nous discutons des constructions qui découlent en quelque sorte de ces notions, tel que le passif (normal et datif), la sous-catégorisation des éléments (qu'elle soit verbale ou prépositionnelle) et l'ordre des mots (Chapitre 1).

Dans le deuxième chapitre, nous présentons la structure topologique ainsi que toutes les questions qui sont soulevées par une telle structure. Nous définissons plus particulièrement la notion de champs et de structure en soi, notamment en comparaison avec une structure syntagmatique d'une grammaire transformationnelle, et nous essayerons de cadrer notre structure à l'intérieur de la Théorie Sens-Texte. Nous donnerons les règles de base de cette structure puis nous discuterons quelques problèmes syntaxiques complexes, que la structure topologique réussit à analyser élégamment : nous évoquerons les questions de *partial VP fronting*, de la structure du complexe verbal et de la relative et du *pied piping*.

Le troisième chapitre constitue un chapitre à part, puisque nous y développons la TAG de l'allemand. Nous commençons par discuter les principes généraux d'une telle grammaire, ainsi que son applicabilité pour une langue comme l'allemand. Nous montrons également l'implémentation de l'ordre des mots en allemand qui a été réalisée dans ce formalisme.

Enfin, dans le quatrième et dernier chapitre, nous revenons au lien entre la dépendance syntaxique et la structure topologique pour proposer cette fois-ci une formalisation précises de ce lien, qui s'inspire de la grammaire TAG mais qui est plus puissante dans le sens qu'elle arrive à marier une structure de dépendances aux différentes linéarisations possibles. Nous discuterons également de quelques problèmes restants, notamment la question de l'ordre des éléments entre le verbe fini et le complexe verbal et celle de la topologie d'un syntagme nominal. Nous présentons également la structure topologique par rapport au formalisme LFG, et nous montrons comment, au delà de l'allemand, une telle structure peut être utile pour la description et l'analyse d'autres langues, notamment le grec moderne.

# Chapitre I

## **L'allemand**

### **Quelques notions de bases de la syntaxe allemande**

# L'allemand

## Quelques notions de bases de la syntaxe allemande

*Die deutsche Wortfolge ist nicht „frei“, sondern denkbedingt.*<sup>11</sup>

Dans ce chapitre, nous esquissons les aspects essentiels de la syntaxe de l'allemand, essentiels pour l'analyse topologique que nous introduisons dans le chapitre suivant et que nous traiterons par la suite dans le cadre d'une grammaire lexicalisée de cette langue. Notre esquisse sera guidée par les problèmes soulevés par l'existence de cas en allemand. Nous nous plaçons dans le cadre de la théorie Sens-Texte (Mel'čuk 1988) et discuterons de ce point de vue des notions d'actant et d'argument, du passif, de la sous-catégorisation des prépositions et du passif datif, puis d'*ordre libre*. Tous les points que nous examinons dans ce chapitre seront ensuite repris et analysés dans les chapitres suivants.

## 1 Les cas de l'allemand

### 1.1 Définition et conventions

Afin de pouvoir analyser et discuter la question de l'ordre des mots en allemand, il est nécessaire avant toute chose de définir le concept de « cas grammatical ». Par cas, on désigne une classe de formes nominales qui est, dans un premier temps, définie indépendamment de la fonction syntaxique qui est attribuée au nom sur lequel le cas s'attache. La définition que nous adoptons s'éloigne des analyses proposées par Kolb 1960, Admoni 1970 ou Eroms 1981 où le cas est défini à partir d'un sens présumé. Nous suivons plutôt les définitions de Heringer 1970 et Eisenberg 1999 qui considèrent le cas comme un marquage morphologique qui permet d'établir la relation syntaxique que le mot entretient avec son gouverneur syntaxique, et ce indépendamment de la position linéaire du mot à l'intérieur de la phrase ou de la cohérence sémantique de la phrase. Selon Eisenberg 1999:71, « en principe, les différents cas n'existent pas pour réaliser différentes significations, mais parce qu'il existe des verbes ayant plusieurs compléments qu'il faut distinguer »<sup>12</sup>. Mais comment peut-on, si nous suivons une telle approche, expliquer la similarité qui existe entre les différentes fonctions sémantiques qui sont réalisés par le même cas ? En fait, il s'agit simplement d'une question d'économie d'expression : en effet, si, pour les dépendants d'un nouveau sémantème, on devait choisir les cas qui les réalisent en surface, ce choix ne serait pas arbitraire, et reposerait sur des choix qui sont déjà faits pour des sémantèmes similaires. Mais cela ne veut aucunement dire que le cas porte lui-même un sens.<sup>13</sup>

---

<sup>11</sup> « L'ordre des mots en allemand n'est pas libre, mais conditionné par la pensée », Drach 1937, p. 26

<sup>12</sup> Texte original « Verschiedene Kasus gibt es im Prinzip nicht, weil damit verschiedene Bedeutungen realisiert werden sollen, sondern weil es Verben mit mehreren Ergänzungen gibt und diese Ergänzungen auseinandergehalten werden müssen » (ma propre traduction).

<sup>13</sup> Cette affirmation est strictement limitée au régime des cas en allemand. Il est clair que le génitif, quand il modifie un nom, porte un sens « d'appartenance ». De plus, en fonction de notre analyse sémantique de la phrase, il peut y avoir des situation où le cas peut servir à exprimer un contraste sémantique, ce que Mel'čuk 2002 (à paraître) appelle *l'utilisation quasi-significative des cas non significatifs*.

En allemand, il y a, comme le note la tradition, quatre cas : le nominatif (nom., 1), le génitif (gén., 2) le datif (dat., 3) et l'accusatif (acc., 4)<sup>14</sup>. Le cas est exprimé de façon synthétique avec le genre (féminin, masculin et neutre) et le nombre (singulier et pluriel). Traditionnellement, on dit que le cas est porté par le syntagme nominal (SN), car les cas apparaissent sur les trois parties du discours qui font partie du SN, les déterminants, les adjectifs et les noms (incluant les pronoms). Ces trois parties du discours font l'accord en cas, en genre et en nombre<sup>15</sup>. Pour une liste des différentes déclinaisons de l'allemand voir par exemple la grammaire du Duden:3.4,4.2, et pour une discussion des problématiques de l'accord à l'intérieur du SN voir par exemple Gerdes 1998. Regardons quelques exemples.

- (1) die Maria schenkt dem Peter Blumen  
la Maria offre au\_Pierre fleurs
- (2) dem Peter schenkt die Maria Blumen  
au\_Pierre offre la Maria fleurs
- (3) Blumen schenkt dem Peter die Maria  
fleurs offre au\_Pierre la Maria  
'Marie offre des fleurs à Pierre'

Il convient tout d'abord de fixer les conventions que nous adoptons pour discuter des cas. Dans les phrases (1), (2) et (3), le sujet *die Maria* est au nominatif, l'objet direct *Blumen* à l'accusatif et l'objet indirect *dem Peter* au datif. Nous notons la relation syntaxique marquée par le nominatif par « sujet », et celle marquée par l'accusatif par « objet ». Par ailleurs, nous ne faisons pas la distinction traditionnelle entre le « casus rectus » (le nominatif) et les « cas obliques » (qui renvoient à tous les autres cas). De plus, nous n'utilisons pas le terme « d'objet indirect » : la relation qui existe entre les arguments au datif ou au génitif et leur gouverneur est indiqué par le nom du cas, et nous parlerons de « (objet) datif » et de « (objet) génitif ». Ainsi, dans les phrases ci-dessus, le verbe *schenkt* 'offre' a trois arguments syntaxiques, un sujet, un objet et un objet datif.

## 1.2 La raison d'être des cas

Le français est souvent décrit comme une langue à ordre des mots fixe. Or, si l'on suit les idées de Greenberg 1978, cet ordre qui fixe la place du sujet et de l'objet direct par rapport à un gouverneur entraîne que ces deux éléments n'ont pas besoin d'un marquage morphologique et peuvent ainsi garder leur forme de base. Si par contre deux éléments syntaxiques d'une phrase peuvent changer librement de place, il leur est difficile de préserver leur forme de base. Dans le cas du français, l'objet direct et l'objet indirect peuvent échanger leur place, et on doit donc trouver un moyen de les distinguer. Or, comme les noms du français n'ont pas de marquage casuel (à l'exception des pronoms), la stratégie choisie pour réaliser la différence au niveau syntaxique et de faire précéder l'objet indirect d'une préposition (dans le cas de l'exemple (4), il s'agit de la préposition *à*). Ainsi, l'objet indirect en français n'est plus exprimé par un syntagme nominal, mais un syntagme prépositionnel (sauf cas de cliticisation).

<sup>14</sup> Les chiffres entre parenthèses indiquent la notation abrégée traditionnelle de ces cas. Nous ne l'utilisons que dans la Grammaire d'Arbre Adjoints dans le chapitre 3.

<sup>15</sup> En fait, dire qu'on fait l'accord en genre entre un nom et un adjectif n'efface pas la hiérarchie qui peut exister entre ces deux éléments : en effet un nom a un genre fixe et un seul, et c'est lui qui impose son genre à l'adjectif.

- (4) Marie donne des fleurs à Pierre.
- (5) Marie donne à Pierre des fleurs.
- (6) Marie lui donne des fleurs.

Notons que ces deux objets diffèrent également dans leur comportement par rapport à la dislocation gauche : l'objet direct ne peut être disloqué à gauche (cf. (7)) sauf s'il est repris par un pronom dans la phrase principale (cf.(8)), tandis que l'objet indirect peut être disloqué sans aucune reprise pronominale (cf. (9)).

- (7) ?? Des fleurs, Marie offre à Pierre.
- (8) Des fleurs, Marie en offre à Pierre.
- (9) A Pierre, Marie offre des fleurs.

En allemand tous les noms porte un cas et donc un signe distinctif pour l'identification de sa relation syntaxique. Une conséquence de ce marquage est que l'ordre des mots de l'allemand est très libre par rapport à la relation syntaxique et peut donc coder d'autres sortes d'information, plutôt, comme nous le verrons dans la suite, de caractère communicatif (ou informationnel).

## 2 Actant sémantique, actant syntaxique et argument syntaxique

Le but de cette section est d'introduire dans les grandes lignes les notions qui apparaissent dans les différentes structures de dépendance et que nous utiliserons tout au long de cette thèse. Par ailleurs, nous tenterons de poser les différents niveaux d'abstraction que nous retenons et de définir le lien qui existe entre eux. Nous plaçons notre travail dans le cadre de la Théorie Sens-Texte (Mel'čuk 1988).

Dans un premier temps, nous définirons donc les termes d'actant à un niveau d'abstraction sémantique, puis nous introduirons à un niveau syntaxique (de surface, dans les termes de la Théorie Sens-Texte) le terme d'argument syntaxique et la notion de sous-catégorisation. Enfin, nous discutons la nécessité d'introduire un niveau intermédiaire appelé « dépendance syntaxique profonde » où apparaîtront des actants syntaxiques.

### 2.1 Actant et argument

Dans son livre *Éléments de syntaxe structurale* de 1959, Tesnière propose une définition de la notion d'actant qui aujourd'hui, est considérée comme une référence incontournable. Cette définition se base sur l'idée qu'un actant d'un concept fait partie de la définition du concept. Autrement dit, on ne peut ni expliquer ni même imaginer le concept sans mentionner les actants qui participent à ce concept. De la même manière, dans une mise en scène dont le rôle serait d'expliquer le concept en question, il serait impossible de ne pas placer ces participants, d'où le nom d'actant.

Prenons un exemple. Il est impossible d'expliquer la signification du verbe « donner » si on n'introduit pas ses actants qui sont au nombre de trois : le donneur, le récepteur et l'objet transmis. Ainsi, la phrase suivante constitue la paraphrase minimale que l'on pourrait proposer pour ce verbe : « un donneur transmet un objet à un récepteur ». Le verbe « donner » gouverne ses trois actants ; tous les autres éléments qu'il peut gouverner ne sont pas nécessaires pour la définition et n'en font pas partie. Tesnière appellent ces éléments des « circonstants » (en s'inspirant du terme allemand *Umstandswort*, « mot de circonstance »).



Tesnière 1959:128 propose de distinguer les actants des circonstants selon deux critères : la définition sémantique des termes et la différence de forme qui existe entre ces éléments. Reprenons le deuxième critère : selon la définition de Tesnière, en français, tout syntagme nominal et tout syntagme nominal au datif, c'est-à-dire un syntagme prépositionnel introduit par la préposition *à*, obtient le statut d'actant et tous les autres syntagmes celui de circonstant<sup>16</sup>. Or, un tel critère, qui découle de la non distinction entre une dépendance sémantique et une dépendance syntaxique, ne peut être retenu pour distinguer le terme d'actant de celui de circonstant. Même Tesnière (1959:127) reconnaît que la distinction n'est pas claire : « à première vue la limite entre actants et circonstants est nette. Mais à y regarder de près, on s'aperçoit qu'elle est délicate à fixer avec précision ». De plus, une telle définition ne peut être appliquée à des langues comme l'allemand et il nous semble complètement arbitraire de dire que les syntagmes au nominatif, à l'accusatif et au datif (ce qui correspond aux syntagmes nominaux et au syntagme prépositionnel introduit par « à » du français) sont des actants alors que ceux au génitif et tous les autres syntagmes prépositionnels sont des circonstants. D'un autre côté, si l'on adopte une définition purement sémantique, chacun des cas peut apparaître soit comme actant, soit comme circonstant. Nous illustrons ces phénomènes dans les exemples (10) à (13) où chaque syntagme nominal a été repris une fois comme actant (phrases en a.) et l'autre fois comme circonstant (phrases en b.) :

- (10) a. Mein Gott ist größer als dein Gott  
mon dieu (nominatif) est plus\_grand que ton dieu  
'Mon dieu est plus grand que ton dieu.'
- b. Mein Gott ist der blöd  
mon dieu (nominatif) est celui-là sot  
'Dieu qu'il est bête'
- (11) a. Ich gedenke eines wundervollen Abends  
je se\_rappelle d'un merveilleux soir(génitif)  
'Je me rappelle d'un merveilleux soir.'
- b. Eines schönen Tages reißen wir die Zäune von den großen Gärten ein.  
un joli jour (génitif) mettons nous les grillage de des grands jardins à\_bas (particule)  
'Un jour, nous mettrons à bas les grillages des grands jardins.'
- (12) a. Peter hilft mir  
Pierre aide à\_moi (datif)  
'Pierre m'aide.'
- b. Das ist mir zu blöd  
cela est à\_moi (datif) trop sot  
'Je trouve cela trop bête.'
- (13) a. Ich schlage nächsten Montag vor.  
Je propose prochain lundi (accusatif) particule  
'Je propose lundi prochain.'

---

<sup>16</sup> Tesnière ne considère donc pas les syntagmes prépositionnels avec *sur* (Pierre compte *sur* Marie) ou avec *de* (Pierre parle de sa thèse) comme des actants.

- b. Nächsten Montag begeben sich die Intellektuellen erstmal auf die "Suche nach der neuen Mitte"<sup>17</sup>.  
 prochain lundi (accusatif) mettent se les intellectuels d'abord sur la « recherche après le nouveau milieu »  
 'Lundi prochain, les intellectuels vont d'abord se mettre à « la recherche du nouveau centre »'

Malgré son caractère flou, c'est donc le critère sémantique que nous adoptons comme seul critère permettant de distinguer l'actant du circonstant. Mais si l'on se tourne maintenant vers la syntaxe, on s'aperçoit que les relations de gouvernance et de modification ressemblent à la relation qui existe entre gouverneur et actant sémantique, sans pour autant coïncider parfaitement. Dans les phrases simples en a. des exemples (10) à (13), les sémantèmes prédicatifs qui se réalisent en tant que verbes fléchis<sup>18</sup>, assignent, au niveau de la réalisation de surface, un cas à ses actants. Dans ce cas précis, les actants sémantiques coïncident avec des entités qui dépendent syntaxiquement du verbe, et que nous appelons « arguments syntaxiques » (ce qui correspond à *l'argument de syntaxe de surface* dans la notation habituelle de la Théorie Sens-Texte). Il est difficile de définir le terme d'« argument syntaxique » mais nous considérons qu'il est indispensable de faire la distinction entre d'une part l'actant sémantique et l'argument syntaxique car la coïncidence n'est souvent pas obligatoire.

D'abord, un actant peut ne pas se réaliser en tant qu'argument syntaxique au niveau de la réalisation de surface. Prenons comme exemple le lexème 'schreiben' 'écrire' qui comporte (au moins) deux actants, la personne qui écrit (actant 1) et l'objet écrit (actant 2). Comme il apparaît dans les exemples en (14), le sens de *schreiben* 'écrire' reste le même bien que certains éléments syntaxiques soient omis, soit parce que la situation d'énonciation permet de retrouver très facilement les actants, soit parce qu'on cherche à exprimer une généralité. Ainsi, dans la phrase a. les deux actants sont présents, dans b. c'est l'actant 1 qui manque, dans c. c'est l'actant 2 qui est omis, et enfin dans d. on ne trouve plus aucun actant dans la réalisation de surface. Pour ces différents exemples, on parle d'ellipse sémantique.

- (14) a. Der Peter schreibt ein Buch.  
 Le Pierre écrit un livre.  
 'Pierre écrit un livre.'
- b. Es ist wichtig ein Buch zu schreiben.  
 Il est important un livre à écrire.  
 'Il est important d'écrire un livre.'
- c. Der Peter schreibt.  
 Le Pierre écrit.  
 'Pierre écrit.'
- d. Schreiben macht müde.  
 Ecrire rend fatigué.  
 'Ecrire, ça fatigue.'

<sup>17</sup> Exemple tiré du journal taz Nr. 6716 du 4/4/2002, page 22.

<sup>18</sup> Un prédicat sémantique ne se réalise pas toujours à l'intérieur d'une phrase complète comprenant un verbe fléchi. Ainsi, si l'on prend comme exemple le sens de *geben* 'donner', son sens peut correspondre à un syntagme nominal :

- (i) die Büchergabe Peters an Maria  
 la livres\_donation de Pierre (génitif) à Marie  
 'la donation des livres à Marie par Pierre'

Il se peut également que le contraire se produise, c'est-à-dire que dans la réalisation de surface, on voit apparaître un nouvel élément syntaxique qui ne correspond à aucun actant sémantique. C'est le cas, par exemple, des sujets impersonnels, réalisé en allemand en tant que *es* 'il' explétif, comme dans les phrases (15) et (16). Le lexème 'weihnachten' 'être dans le temps de Noël' exprime une situation qui ne met en jeu aucun actant (bien que le lexème soit décomposable comme la plupart des lexèmes, voir Mel'čuk 1988 pour des détails), mais au niveau syntaxique, *es* 'il' est un argument syntaxique, le sujet, qui fait l'accord avec le verbe de manière habituelle.

(15) Es weihnachtet sehr.  
Il fait\_noël beaucoup.  
'Le temps de Noël s'approche vite.'

(16) Es ist soweit.  
Il est jusqu'à\_un\_certain\_point.  
'Ça y est.' / 'Le temps est venu.'

D'autres cas de non-congruence entre la structure sémantique et syntaxique incluent des auxiliaires de temps. En allemand (comme d'ailleurs en français), on peut exprimer le passé soit comme un seul morphème (le temps imparfait par exemple) soit comme un verbe auxiliaire (le temps passé composé par exemple)<sup>19</sup>, tel qu'il est illustré par les phrases en (17).

(17) a. der Peter gab der Maria ein Buch.  
le Pierre (nominatif) donnait le Marie (datif) un livre (accusatif)  
'Pierre donnait un livre à Marie.'

b. der Peter hat der Maria ein Buch gegeben.  
le Pierre (nominatif) a le Marie (datif) un livre (accusatif) donné  
'Pierre a donné un livre à Marie.'

Dans la phrase a., le sujet fait l'accord en nombre et en personne avec le verbe *gab* tandis que dans la phrase b., il le fait l'accord non avec le participe passé *gegeben* 'donné' mais l'auxiliaire *hat* 'a'. Il s'agit d'un bon critère pour dire que le sujet dépend non du participe passé *gegeben* qui correspondrait au sémantème 'donner', mais de l'auxiliaire, un élément qui a été introduit au niveau de la réalisation de surface pour exprimer le temps *passé* de la représentation sémantique. On considère donc que l'auxiliaire gouverne deux éléments syntaxiques : son sujet<sup>20</sup>, le syntagme nominal au nominatif et le participe passé. Le syntagme nominal au nominatif ne fait pas partie des arguments syntaxiques du participe passé *gegeben* qui ne gouverne plus que les syntagmes à l'accusatif et au datif, bien que au niveau sémantique, son sémantème correspondant avait bel et bien trois actants.

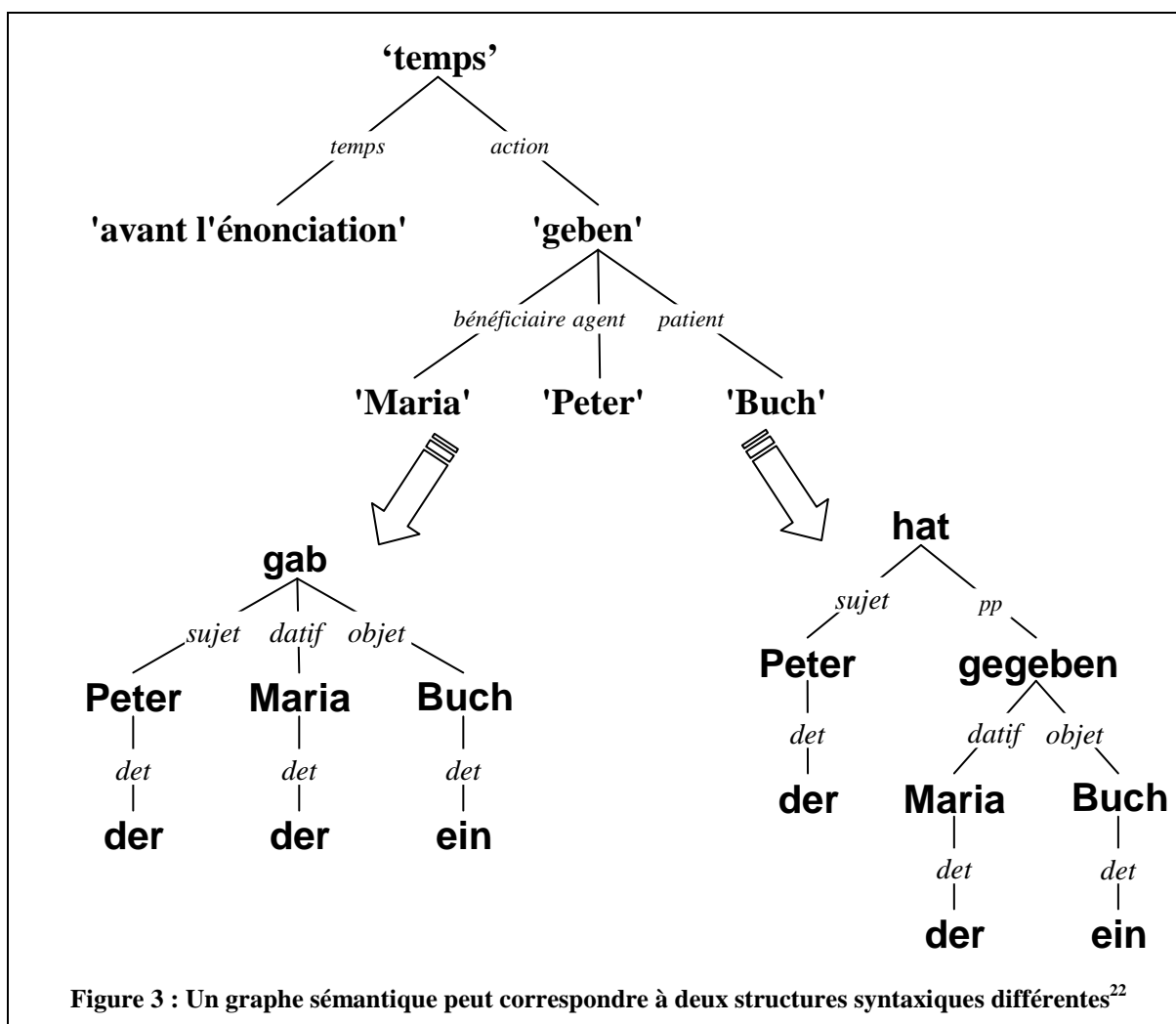
Dans la Figure 3, nous représentons les phrases sous (17), en montrant comment un même sens peut être traduit dans deux structures syntaxiques différentes.

Du point de vue de la représentation, il faut préciser que la structure de dépendance sémantique, partie essentielle de la représentation du sens d'une phrase, est représentée par un graphe. Il s'agit d'un graphe connexe ayant comme nœuds des sémantèmes (des unités de sens), qui sont liés entre eux par des *dépendances sémantiques*, exprimant par exemple le rôle d'actant ou de circonstant<sup>21</sup>. Une telle dépendance sémantique peut se réaliser en surface de différentes

<sup>19</sup> La distinction que fait le allemand entre l'imparfait et le passé composé est d'un caractère similaire mais beaucoup moins strict qu'en français. En allemand, il s'agit plutôt d'une question de style, l'imparfait relevant d'un langage plutôt soutenu (comme le passé simple du français), qui est encore compliquée par des différences régionales. Voir Duden Grammatik §262 pour des détails.

<sup>20</sup> On dit aussi que le verbe (ou l'auxiliaire) sous-catégorise le syntagme nominal.

<sup>21</sup> Pour une question de lisibilité, nous avons choisi de donner des noms aux liens qui existent entre les sémantèmes et non, comme le fait Mel'čuk 1988, des numéros.



manières. Par contre, au niveau syntaxique, les structures que nous retenons ne sont plus des graphes mais des arbres, c'est-à-dire que chaque élément à l'exception de la racine a un seul gouverneur syntaxique (sauf, éventuellement dans les cas de la coordination et des relatives, si on suit les analyses de Tesnière 1957, Hudson 1990 ou de Kahane 1997).

Les divergences entre actant sémantique et argument syntaxique ne s'arrêtent pas là. Il existe des exemples où un même lien sémantique est réalisé par différents cas grammaticaux. Ainsi dans la phrase (18), le verbe *geben* 'donner' apparaît dans une construction à contrôle de l'accusatif (*AcI*, accusativus cum infinitivo) tandis que dans la phrase (19), il apparaît dans une construction au passif.

- (18) Ich bitte Peter Maria ein Buch zu geben.  
 je (nominatif) demande Pierre (accusatif) Marie (datif) un livre (accusatif) à donner  
 'je demande à Pierre de donner un livre un Marie.'

<sup>22</sup> Nous laissons pour l'instant de côté l'apparition du déterminant ainsi que l'encodage de l'information qu'il contrôle dans la structure sémantique. Il est d'ailleurs pas très clair si au niveau syntaxique, le déterminant dépend du nom ou si inversement, le nom dépend du déterminant. Nous y reviendrons par la suite dans la section IV.6. Par ailleurs, nous précisons que nous notons directement les mots sous leurs formes fléchies pour une question de lisibilité (par exemple, *gab*), bien que, dans une tradition Sens-Texte, il eut été plus correct de noter le lexème accompagné de la liste de ses grammèmes (au lieu de *gab*, GABEN 3<sup>ème</sup> personne du singulier imparfait).

- (19) Ein Buch wurde Maria von Peter gegeben.  
 un livre (nominatif) devenait Marie (datif) par Pierre (datif dans SP) donné  
 'Un livre a été donné à Marie par Pierre.'

Dans la construction à contrôle de l'objet, deux actants de *geben* 'donner' sont réalisés en surface à l'accusatif; mais syntaxiquement, nous considérons qu'ils appartiennent à deux gouverneurs différents, puisque l'agent de *geben* est l'objet du verbe à contrôle tandis que son patient est réalisé comme objet de l'infinitif. Une telle construction ressemble au changement de diathèse (de voix) que nous avons illustré sous (19) : dans les deux cas, l'agent n'est pas réalisé comme argument syntaxique du verbe *geben* mais comme argument d'un verbe qui gouverne syntaxiquement ce verbe, et d'une certaine manière, nous pouvons dire que l'agent est « monté » vers un verbe supérieur. Dans les Figure 4 et Figure 5, nous reportons le graphe sémantique et l'arbre syntaxique des phrases (18) et (19).

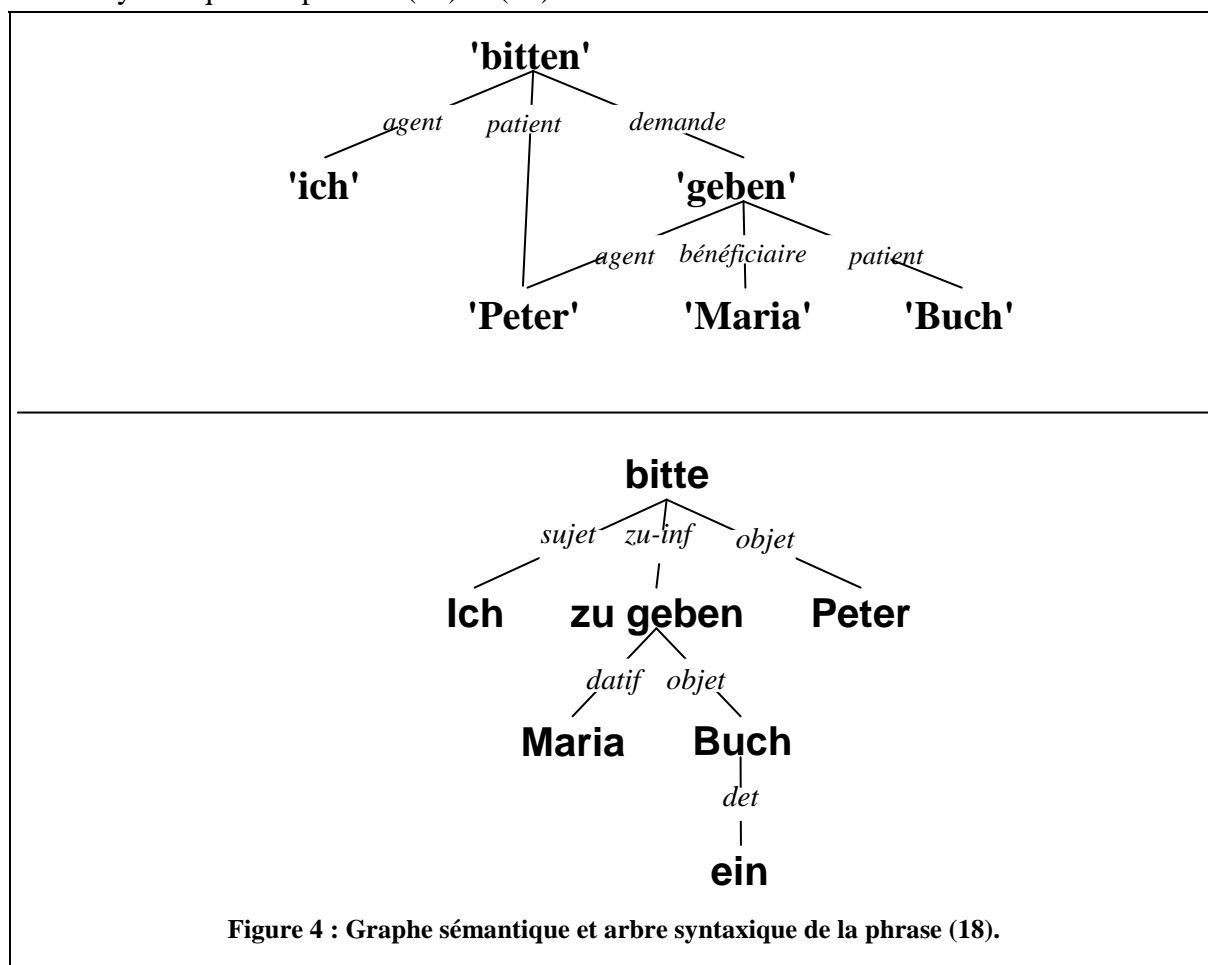
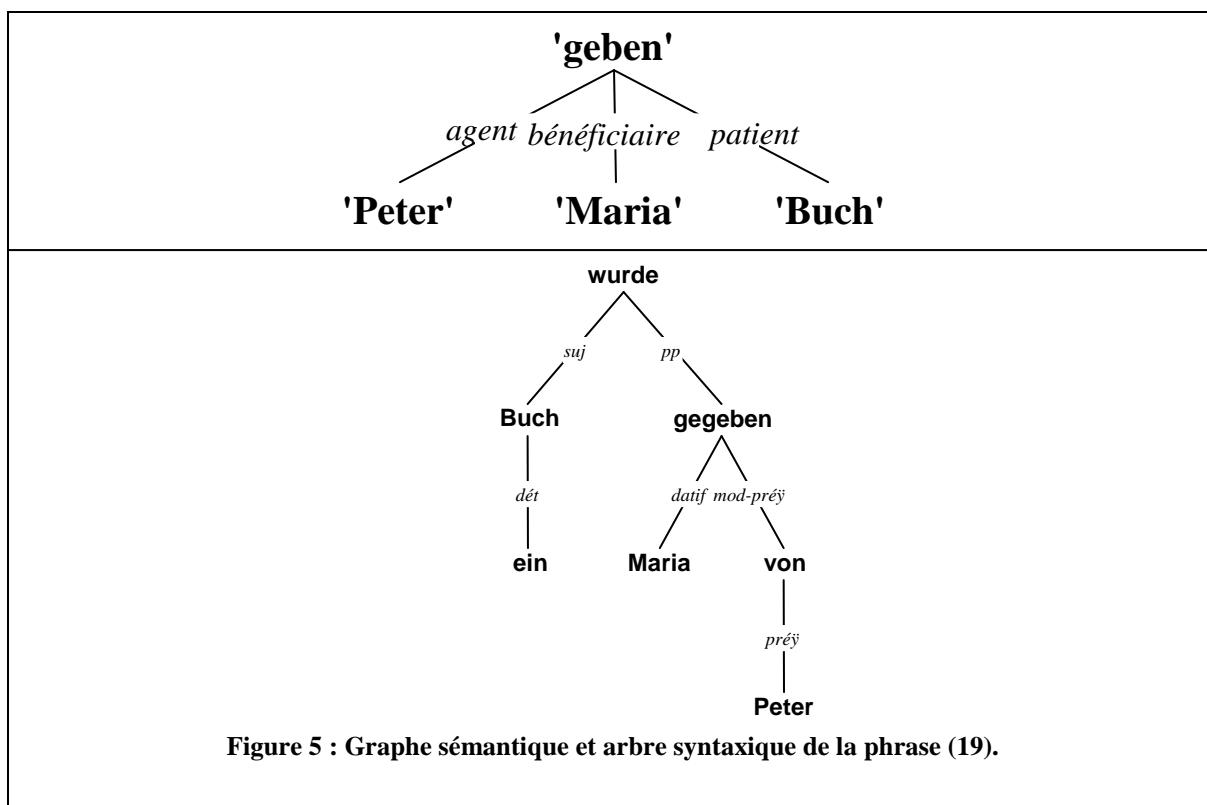


Figure 4 : Graphe sémantique et arbre syntaxique de la phrase (18).



Nous précisons déjà ici (mais nous y reviendrons en détail infra) que nous suivons les idées de Bech 1955 qui remarque que le gouverneur d'un argument verbal peut imposer trois formes de verbe non finies. Ainsi, le verbe *geben* 'donner' peut apparaître sous ces trois formes non finies :

- en tant qu'infinitif nu *geben* 'donner', que Bech appelle le premier statut,
- en tant qu'infinitif avec *zu*, *zu geben* 'à donner', qui correspond à son deuxième statut<sup>23</sup>,
- et le participe passé (pp) *gegeben* 'donné', le troisième statut selon Bech.

Par ailleurs, Bech n'inclut pas le participe présent (*gebend*) dans sa liste car aucun élément ne sous-catégorise un verbe au participe présent. Contrairement à l'anglais, les participes présents de l'allemand apparaissent exclusivement comme adjectif épithète à l'intérieur d'un syntagme nominal.

Nous voyons qu'il est possible de décrire le lien entre la structure argumentale au niveau sémantique (les actants sémantiques) et la structure argumentale au niveau syntaxique de surface (les arguments syntaxique). Dans la section suivante, nous discutons l'utilité pour la simplification des règles de transfert, d'introduire un niveau de description supplémentaire entre la sémantique et la syntaxe (de surface).

## 2.2 Introduction d'un niveau intermédiaire : la syntaxe profonde

Il est possible de construire des règles qui expriment la correspondance entre un actant sémantique et un argument syntaxique. Cependant, ces règles ne captent malheureusement pas des généralités importantes.

Supposons qu'au niveau sémantique, le sémantème *schreiben* 'écrire' ait comme actant « un destinataire » réalisé par exemple au datif, comme nous le montrons dans l'exemple (20)<sup>24</sup> :

<sup>23</sup> Notons que *zu* et *geben* forment une seule forme morphologique, bien qu'elle s'écrive en deux mots-formes à part. En effet, contrairement à l'anglais ou au français, aucun élément syntaxique ne peut s'insérer entre *zu* et l'infinitif et dans les cas des verbes avec particule (cf. section II.2), les conventions d'écriture veulent même qu'on l'écrive en un seul mot-forme, comme on peut le voir dans le cas du verbe *vorzuschlagen* de l'exemple (13)a.

- (20) Maria schreibt meinem Bruder  
 Marie écrit à\_mon frère(datif)  
 'Marie écrit à mon frère.'

Supposons encore que c'est le même lien sémantique « destinataire » qui relie le sémantème *anrufen* 'appeler' et *Ihrem Bruder* mais que cette fois-ci, le syntagme nominal soit à l'accusatif (cf. phrase (21)) :

- (21) Maria ruft meinen Bruder an  
 Marie appelle mon frère (accusatif) particule  
 'Marie appelle mon frère'

Etant donné la similitude qui existe entre ces deux exemples, il est difficile d'expliquer pourquoi la première phrase ne peut se réaliser sous une forme passive, contrairement à la deuxième phrase qui peut le faire :

- (22) a. \*Mein Bruder wird von Maria geschrieben.  
 mon frère devient par Marie écrit  
 mon frère est écrit par Marie  
 b. Mein Bruder wird von Maria angerufen.  
 mon frère devient par Marie appelé  
 'Mon frère est appelé par Marie.'

Il est assez clair que le changement de diathèse (de voix) est possible dans le cas de la phrase (21) parce que la réalisation du « destinataire » à l'actif se fait par le biais d'un cas accusatif. Nous sommes face à une généralité, qui relie le cas que porte un argument et la possibilité d'être au passif. Afin de capturer cette généralité, il convient d'introduire un niveau intermédiaire qui se situe entre la structure sémantique et la structure syntaxique : la structure de syntaxe profonde. Selon Mel'čuk 1988:60, cette structure « spécifie l'organisation syntaxique de la phrase en termes des mots qu'elle contient ainsi que des relations que ces mots entretiennent entre eux<sup>25</sup> ». Il s'agit en fait d'une structure arborescente : puisque un nœud ne peut plus avoir deux gouverneur et fonctionner ainsi en tant qu'argument pour deux prédicats, il est nécessaire de calculer la structure anaphorique à l'intérieur de la phrase pendant la traduction d'un graphe sémantique en un arbre de dépendance syntaxique profonde.

Dans les analyses habituelles dans la Théorie Sens-Texte, le choix de la diathèse se fait entre le niveau sémantique et le niveau syntaxique profonde : on ne considère donc jamais une structure dans laquelle on exprime la différence syntaxique, et non forcément sémantique, entre par exemple un verbe avec un argument « profond » à l'accusatif, permettant le passif, et un autre verbe sans un tel argument, où le passif n'est pas possible (cf. (22) a et b). Comme les niveaux profonds dans la théorie Sens-Texte sont considérés comme des niveaux intermédiaires et destinés à une simplification et une généralisation du lien entre deux niveaux principaux, nous préférons une analyse où, à ce niveau de syntaxe profonde, il n'est pas encore décidé avec quelle diathèse la phrase sera réalisée. Par contre, nous savons déjà sous quelle forme le sens sera exprimé, c'est-à-dire s'il sera traduit comme un nom ou un verbe. Par ailleurs, nous pouvons introduire des *relations syntaxiques profondes* qui expriment les liens qu'entretient un gouverneur de syntaxe profonde avec ses dépendants, que nous appelons « actants syntaxiques ». C'est ainsi que nous en venons à la distinction de trois notions, l'actant sémantique qui est

<sup>24</sup> Nous restons dans l'ordre de la supposition car les analyses sémantiques sont toujours discutables et dépendent de l'extension que l'on donne à un sémantème. Dans le cas de notre exemple, le sens de *schreiben* peut être défini sans destinataire.

<sup>25</sup> texte original : « It specifies the syntactic organization of the sentence in terms of its words and relationships between them ». (ma propre traduction)

présent au niveau sémantique, l'actant syntaxique qui est présent au niveau de la syntaxe profonde et l'argument syntaxique au niveau de la syntaxe de surface.

Dans la théorie Sens-Texte, les relations de la syntaxe profonde sont censées être universelles. Mel'čuk en dénombre neuf pour toutes les langues du monde. Les relations syntaxiques représentent une généralisation des actants de surface.. Dans ce sens, cette notion se rapproche du terme classique de «fonction grammaticale» ou encore de « sous-catégorisation initiale » utilisé par Candito 1999 dans sa description TAG du français et de l'italien.

Le diagramme suivant de la Figure 6 donne une représentation simplifiée d'une analyse Sens-Texte.

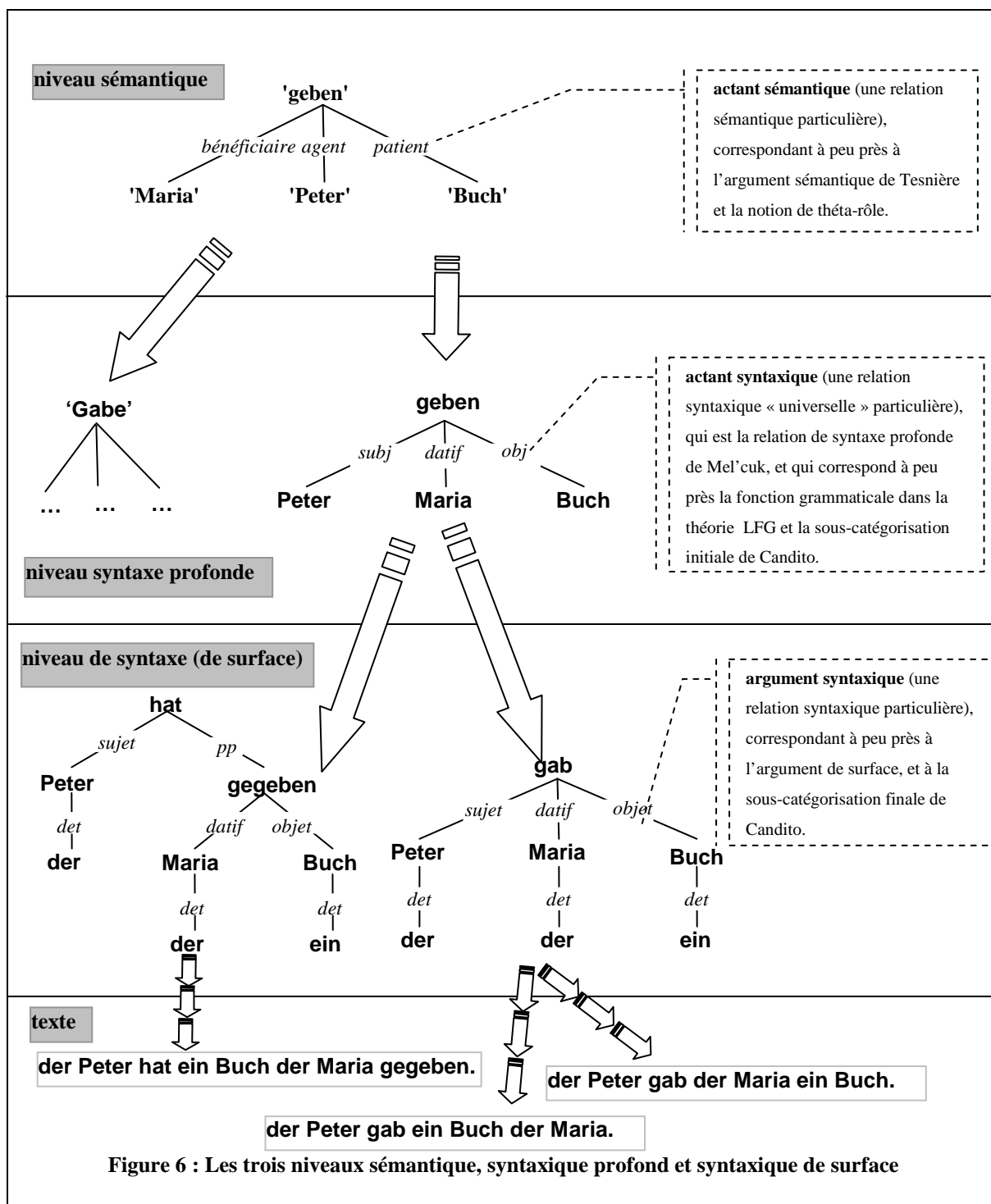


Figure 6 : Les trois niveaux sémantique, syntaxique profond et syntaxique de surface



Dans ce diagramme, nous avons utilisé les mêmes termes pour nommer la relation syntaxique entre deux éléments aussi bien au niveau de la syntaxe profonde qu'au niveau de la syntaxe de surface. C'est ainsi que nous utilisons de manière simplifiée le terme de *sujet* pour désigner à la fois l'élément qui se réalise comme l'actant syntaxique *sujet* et la relation syntaxique dans une diathèse donnée. En fait, nous considérons que dans le cadre de notre travail, il suffit de garder le nom de la relation qu'entretient un actant syntaxique avec son gouverneur étant donné que ce nom correspond à la réalisation la moins marquée de l'élément. Ainsi, pour les verbes, on utilisera la réalisation en tant que verbe fléchi au présent et à la voix active. De manière simplifiée, on dira donc qu'un objet (du niveau profond) est réalisé, au niveau de la syntaxe de surface, comme sujet dans la diathèse passive. Par ailleurs, nous notons les circonstanciels ainsi que les autres modificateurs *mod*, ce qui est une légère variation de la notation ATTR de Mel'čuk 1988. Nous proposons le critère suivant pour distinguer un actant syntaxique d'un modifieur :

*Critère de diathèse* : un élément syntaxique profond qui ne se réalise pas d'une seule manière au niveau syntaxique de surface mais qui peut être réalisé à l'aide de différentes relations syntaxiques, est obligatoirement un actant syntaxique.

Par contre, le contraire n'est pas vrai ; autrement dit, un élément qui intervient dans les changements de diathèse est clairement un actant syntaxique, mais ce n'est pas une condition nécessaire pour être un actant. Ainsi, il est possible que des verbes qui ne connaissent pas de changement de diathèse comme les verbes *kosten* 'coûter' ou *auffallen* 'se faire remarquer' portent quand même des actants syntaxiques.

- (23) a. Das Buch kostet einen Euro.  
ce livre coûte un euro (accusatif)
- b. \* Ein Euro wird vom Buch gekostet.  
un euro (nominatif) devient du livre coûté
- (24) a. Peter fällt mir auf.  
Pierre fait\_remarquer moi (datif) particule  
'Pierre se fait remarquer par moi.'
- b. \* Vom Peter wird mir aufgefallen.  
de Pierre devient moi (datif) fait\_remarqué\_passé

Dans la section suivante, nous nous attarderons plus précisément sur la question du passif et des différents changements de diathèse.

### 3 Le passif

Dans cette section nous allons voir les deux passifs « classiques » de l'allemand le *passif de processus* et le *passif de l'être*.

#### 3.1 Construction du passif

Les distinctions que nous avons établies dans la section précédente sur les notions et les différents niveaux de représentation sont primordiales pour expliquer le fonctionnement du passif. En français, on procède en deux temps pour exprimer le passif : le sujet étant un élément obligatoire d'une phrase en français, on réalise d'abord l'actant syntaxique d'objet en tant que sujet (ce que Comrie 1977 et aussi Candito 1999 appelle la promotion de l'objet), puis l'actant

syntactique de sujet se réalise comme un syntagme prépositionnel introduit par « par », ou il n'est pas réalisé (ce que Comrie et Candito appelle la *démotion* du sujet). Le but principal du passif est donc de promouvoir l'objet dans une position de sujet, qui comme on l'a vu dans la section précédente, entretient une relation privilégiée avec le verbe fini. Nous considérons qu'une telle promotion devient nécessaire pour des raisons plutôt communicatives.

Passons maintenant à l'allemand. Dans cette langue, le passif se construit également en plusieurs étapes. D'abord, le sujet est réalisé par une « démotion du sujet », c'est-à-dire qu'il est exprimé à l'aide d'une relation modificatrice avec le verbe ou il n'est pas exprimé du tout. Puis l'étape de la promotion de l'objet est accomplie chaque fois que l'actant syntaxique de l'objet est présent, mais cette étape n'est pas nécessaire dans le changement de diathèse. Contrairement au français, l'allemand n'exige pas d'arguments syntaxiques de sujet ou d'objet pour réaliser un passif. Dans les phrases suivantes, le verbe *tanzen* 'danser' de l'exemple (25) est un verbe absolu qui exige comme seul actant syntaxique un sujet. Le verbe *scheitern* 'échouer' (26) possède un sujet et un deuxième argument prépositionnel, qui n'est pas affecté par le changement de diathèse. Le verbe *helfen* 'aider' (27) est un verbe ayant un actant syntaxique sujet et un actant syntaxique d'objet indirect réalisé par un datif et enfin le verbe *singen* 'chanter' (28) est un verbe transitif standard avec deux actants syntaxiques, un sujet et un objet direct à l'accusatif. Chacun de ces verbes a une diathèse active (exemples sous a.) et une diathèse passive (sous b.).

(25) a. Wir tanzen die ganze Nacht.

nous (nominatif) dansions la toute nuit  
'Nous dansions toute la nuit'

b. die ganze Nacht wird (von uns) getanzt.

la toute nuit devient (par nous) dansé  
'il y avait de la danse toute la nuit' ou littéralement 'toute la nuit, a été dansé (par nous).'

(26) a. Das Theaterprojekt scheiterte nicht an Misswirtschaft oder fehlendem Publikum, sondern ....

le projet\_de\_théâtre échouait pas à mauvaise\_gestion ou manquant public, mais...  
'Le projet de théâtre n'a pas échoué à cause de mauvaise gestion ou parce qu'il manquait le public, mais ...

b. Gescheitert wurde dabei nicht an Misswirtschaft oder fehlendem Publikum, sondern weil Eigentumsverhältnisse plötzlich geklärt waren<sup>26</sup>.

échoué devenait ici pas à mauvaise\_gestion ou manquant public, mais parce\_que des\_conditions\_de\_propriété soudainement clarifié était souvent devient au Pierre (datif) (par nous) aidé  
'Ils (?) n'ont pas échoué à cause de mauvaise gestion ou parce qu'il manquait le public, mais parce qu'ils se sont soudainement retrouvés dans des conditions claires de propriété.

(27) a. Wir helfen dem Peter oft.

nous (nominatif) aidons au Pierre (datif) souvent.  
'Nous aidons souvent Pierre'

b. oft wird dem Peter (von uns) geholfen.

souvent devient au Pierre (datif) (par nous) aidé  
'Pierre souvent est aidé (par nous).'

---

<sup>26</sup> journal taz Berlin lokal N° 6594 du 7/11/2001, page 25. Notons aussi le processus d'état dans la deuxième moitié de la phrase, voir section suivante.

- (28) a. Wir singen ein schönes Lied.  
nous (nominatif) chantons un beau chanson (accusatif)  
'Nous chantons une belle chanson.'
- b. ein schönes Lied wird (von uns) gesungen.  
un beau chanson (nominatif) devient (par nous)chanté  
'Une belle chanson est chantée (par nous).'

### 3.2 Passif de processus et passif d'état

Morphologiquement, le passif de l'allemand ressemble à celui du français : le participe passé est construit à l'aide d'une structure qui ressemble à une phrase attributive comportant un auxiliaire (*werden* 'devenir' ou *sein* 'être'). L'auxiliaire fait l'accord avec le sujet s'il est présent, sinon il prend par défaut la forme de la troisième personne du singulier. La construction faite avec l'auxiliaire *werden* s'appelle un passif de processus tandis que celle faite avec *sein* un passif d'état. Nous illustrons les deux constructions possibles dans les exemples sous (29).

- (29) a. Die Frage ist ( ? vom Programmkomitee) entschieden.  
la question est (par comité\_du\_programme ?) décidé  
'La question a été décidée (par le comité de programme).'
- b. Die Frage wird (vom Programmkomitee) entschieden.  
la question devient (par comité\_du\_programme) décidé  
'La question est décidée (par le comité de programme).'

Ces deux auxiliaires sont tous les deux utilisés dans des structures attributives, et beaucoup de participes passés peuvent apparaître avec une fonction adjectivale. Ainsi, chaque structure au passif est potentiellement analysable avec la structure attributive, comme c'est le cas d'ailleurs du français (30) :

- (30) a. La question est discutée.  
b. La question est discutée par le comité de programme.  
c. La question est close.

En français, on a tendance à donner une interprétation différente aux phrases a. et b., et ce, en fonction de la présence de l'agent. Dans la phrase a., où l'agent est absent, on opte plutôt pour une interprétation du verbe en tant qu'action accomplie comme dans c.<sup>27</sup>. Par contre, dans la phrase b. où l'agent est présent, on a affaire à un « vrai » passif dans le sens qu'on l'interprète de la même manière que son homologue actif (phrase (31)), avec une idée de présent ou de futur.

- (31) Le comité de programme discute cette question.

Cette différence d'interprétation, qui en français, n'apparaît que comme une tendance, est complètement grammaticalisée en allemand par l'utilisation de deux passifs utilisant deux auxiliaires différents.

D'abord le passif d'état. Ce dernier (cf. Duden Grammatik 1995 §322) ressemble à une structure attributive au point que l'on pourrait les confondre : l'état final de l'action est attribué au sujet de la même manière qu'un adjectif. De plus, le verbe être peut être remplacé par d'autres copules en tant que « schèmes ». Nous donnons dans l'exemple (32) une construction avec la copule *scheinen* 'sembler'.

---

<sup>27</sup> Cette interprétation peut sembler plus claire avec des verbes comme « décider ».

- (32) Die Frage scheint entschieden  
la question semble décidé  
'La question semble être décidée.'

Dans le cas du passif de processus, l'analyse attributive est cachée mais reste encore accessible : l'utilisation de l'auxiliaire de futur *werden* implique que l'on peut avoir une interprétation où l'aboutissement de l'action est attribué au sujet de l'attribut. Parallèlement, le passif de processus est devenu une forme du verbe ayant sa propre morphologie : le participe passé de *werden* 'devenir' est *geworden* quand il est attribut (33), mais en tant qu'auxiliaire de passif, on utilise la forme *worden* (34).

- (33) Er ist alt geworden  
il est vieux devenu  
'Il est devenu vieux.'

- (34) Er ist geschlagen \_worden  
il est frappé devenu  
'Il a été frappé.'

En effet, dans une construction passive, le participe passé porte déjà le préfixe *ge*, et l'auxiliaire du passif ne le porte pas pour éviter de redoubler le même affixe.

Cette différence de forme peut servir de test pour distinguer les adjectifs des participes passés. Par exemple, le mot *verrückt* est ambigu car il peut renvoyer à l'adjectif « fou » ou au participe passé du verbe *verrücken* 'déplacer' (cf. (35) et (36)) :

- (35) Er ist verrückt geworden  
il est fou devenu  
'Il est devenu fou.'

- (36) Er ist verrückt \_worden  
il est déplacé devenu  
'Il a été déplacé.'

En même temps, il devient clair qu'il y a une ambiguïté avec tous les autres temps du passif entre la lecture au passif et du participe passif attributif avec *werden* (37).

- (37) Er wird geschlagen  
il devient frappé  
'Il est frappé.'

Dans le cas du passif d'état, l'événement n'est pas décrit comme un processus mais comme un état, le résultat d'une action terminée. L'acteur de ce processus n'a pas d'importance et par conséquent sa réalisation au niveau syntaxique de surface devient plus rare que si l'on avait affaire à un passif « normal ». Si l'on reprend l'exemple cité dans la grammaire de Duden §320 que nous reportons sous (38), on ne peut parler d'une dérivation faite à partir d'un passif normal tel (39) mais d'une dérivation faite directement à partir de sa correspondante active (40).

- (38) Die Bücher sind mit Staub bedeckt<sup>28</sup>.  
les livres sont avec poussière couvert  
Les livres sont couverts de poussière.

---

<sup>28</sup> Exemple pris de la Grammaire de Duden, §320.

(39) <sup>?</sup> Die Bücher werden mit Staub bedeckt.

les livres deviennent avec poussière couvert  
Les livres sont couverts de poussière.

(40) Staub bedeckt die Bücher.

poussière couvre les livres  
'La poussière couvre les livres'

La phrase (39) est bizarre parce que le seul contexte où une telle phrase pourrait apparaître serait dans une scène où quelqu'un se trouve avec une boîte de poussière devant une étagère de livres, en train de la « poussiérer ». Le mot *bedeckt* a donc deux sens différents : dans le premier sens, le sujet est inactif et désigne la substance couvrante tandis que dans le deuxième, le sujet désigne le « couvreur ». Par ailleurs, pour un verbe qui ne connaît qu'un sujet inactif, seul le passif d'état est disponible, ce qui est d'ailleurs un contre-exemple concernant l'affirmation de Müller 1999:291, selon laquelle le passif d'état n'est possible que quand ce verbe peut par ailleurs se construire avec un passif de processus. Nous donnons un exemple d'un tel verbe en (41) :

(41) In dieser Kanne<sup>29</sup> ist Tee enthalten (\*worden).

Dans ce pot est thé contenu (\*devenu [auxiliaire du passif de processus]).  
'Dans ce pot il y a du thé.'

Le cas inverse peut être illustré par un verbe transitif qui décrit une action qui n'a pas d'influence sur l'objet (42).

(42) Er (wird / \*ist) von allen bewundert.

Il (devient [aux du passif de processus] / \*est [aux du passif d'état]) par tous admiré.  
Il est admiré par tout le monde

On peut donc faire la remarque suivante. Bien que le sujet disparaisse dans les deux formes du passif, le passif de processus continue à présenter l'action du point de vue du sujet, et c'est pour cela qu'on trouve des restrictions sur le sens du sujet. Par contre, le passif d'état est avant tout concerné par l'effet de l'action sur l'objet, c'est-à-dire sur ce que l'objet a « vécu ». Il faut alors qu'une entité, qui est déjà présent avant le déclenchement du procès, soit affectée et mise dans un nouvel état par l'action même. Comme un tel objet est normalement réalisé par un accusatif, on a presque exclusivement affaire à des verbes transitifs, mais on trouve des exceptions, ce qui démontre qu'il ne s'agit pas d'une restriction syntaxique. Ainsi, l'actant qui profite de l'aide est (ou devrait être) affecté par le fait d' « aider » bien qu'il soit réalisée au datif, comme on peut le voir dans l'exemple (43).

(43) Ihm ist damit nicht geholfen.

à lui [datif] est avec cela ne pas aidé.  
Avec cela, il n'a pas été aidé (et maintenant il ne va pas mieux qu'avant)

On voit donc apparaître clairement des différences nettes de construction mais aussi d'interprétation entre les deux formes de passifs, le passif d'état et le passif de processus. Néanmoins, il est difficile de fournir une liste complète regroupant d'une part les verbes qui

---

<sup>29</sup> On ne peut d'ailleurs pas utiliser un SP avec *von* 'de' (*Von dieser Kanne...*), parce que le pot n'est pas déclencheur de cet état, contrairement à la phrase (38) où le syntagme prépositionnel peut aussi être introduit par *von* et ce sans changement du sens. La préposition *mit* 'avec' provient déjà du verbe à l'actif (*bedecken mit / couvrir de*) tandis que *von* 'de' provient de la passivation. Les deux propositions entrent ici en compétition.

permettent un passif d'état et ceux qui permettent un passif de processus (plus ceux qui ne construisent pas du tout de passif, et ceux qui permettent les deux types de passif). La raison principale qui empêche l'établissement d'une telle liste repose sur la difficulté de définir la notion d'ergativité (cf. la discussion dans Müller 1999:285 ou dans Grewendorf 1989:179). Sur cette remarque, nous terminons cette section générale sur le passif, mais il en sera question de nouveau dans les chapitres suivants, notamment avec le traitement du passif dans la grammaire TAG d'une part et dans la structure topologique d'autre part.

## 4 Prépositions et arguments syntaxiques

Tout comme le verbe, la préposition peut sous-catégoriser des éléments<sup>30</sup>. Dans cette section, nous discuterons de quelques questions intéressantes soulevées par la préposition comme gouverneur.

En allemand, certaines prépositions comme *mit* 'avec', *zu* 'à' et *von* 'de' sont toujours suivies d'un syntagme nominal au datif (cf. exemples (44) à (46)).

(44) zu dem Bade

à le bain (datif)  
'au bain'

(45) von einem Freund

de un ami (datif)  
'd'un ami'

(46) mit mir

avec moi (datif)  
'avec moi'

Mais une telle généralité sur le cas porté par un syntagme nominal qui est gouverné par une préposition ne se vérifie pas toujours et le fonctionnement de la plupart des prépositions de l'allemand est beaucoup plus complexe. En effet, le syntagme nominal qui suit une préposition autre que *mit*, *zu* et *von* peut être soit au datif, soit à l'accusatif, selon que la préposition a une interprétation locative ou directionnelle.

(47) a. Sie fliegt nicht gern in den USA.

elle vole pas volontiers au les E-U (datif)  
'Elle n'aime pas voler à l'intérieur des États-Unis.'

b. Sie fliegt nicht gern in die USA.

elle vole pas volontiers au les E-U (accusatif)  
'Elle n'aime pas voler vers les États-Unis.'

---

<sup>30</sup> Certains adjectifs peuvent aussi sous-catégoriser des arguments nominaux. En effet, des adjectifs comme *fremd* 'étrange', *bekannt* 'connu', *behilflich* 'serviable' et *zuwider* 'répugnant' prennent un argument au datif, très peu d'adjectifs comme *gewohnt* 'habitué' prennent l'accusatif, et quelques-uns comme *schuldig* '~dans l'obligation de ' peuvent prendre les deux. A ma connaissance, le seul exemple d'adjectif qui prenne le génitif est *sicher* 'sûr' (cet adjectif peut apparaître aussi avec un accusatif qui est réfléchi, Müller 1999:272). Dans le cadre de ce travail, nous nous limiterons à la description détaillée de la sous-catégorisation des verbes et des prépositions.

- c. Sie fliegt nicht gern.  
elle vole pas volontiers  
'Elle n'aime pas prendre l'avion.'
- (48) a. Sie joggt im Park.  
elle court au parc (datif).  
'Elle court dans le parc.'
- b. Sie joggt in den Park.  
elle court dans le parc (accusatif)  
'Elle court vers le parc.'
- c. Sie joggt.  
Elle court  
'Elle fait du footing.'
- (49) a. Sie springt wie wild auf der Matratze.  
Elle saute comme fou sur le matelas (datif)  
'Elle saute comme une folle sur le matelas.'
- b. Sie springt wie wild auf die Matratze.  
Elle saute comme fou sur le matelas (accusatif)  
'Elle se lance comme une folle vers le matelas.'
- c. Sie springt.  
Elle saute  
'Elle se lance.'
- (50) a. Sie singt wie wild auf der Matratze.  
Elle chante comme fou sur le matelas (datif)  
'Elle chante comme une folle sur le matelas.'
- b. \*Sie singt wie wild auf die Matratze.  
Elle chante comme fou sur le matelas (accusatif)  
'Elle chante comme une folle vers le matelas.'
- c. Sie singt.  
Elle saute  
'Elle se lance.'

Comme nous pouvons le voir dans les exemples (47) à (50), les prépositions *in* et *auf* peuvent prendre un syntagme nominal au datif mais aussi à l'accusatif. La question se pose donc d'abord à l'intérieur du syntagme prépositionnel : la préposition sous-catégorise-t-elle à la fois des datifs et des accusatifs comme actants syntaxiques, ou bien s'agit-il de deux formes homonymes qui ont chacune une valeur différente ? Un premier élément de réponse se trouve dans le rôle que joue un syntagme prépositionnel à l'intérieur de la phrase, et dans la décision de savoir s'il s'agit d'un modifieur ou d'un actant syntaxique. Or le syntagme prépositionnel ne répond pas au *critère de diathèse* que nous avons proposé comme critère de base permettant de distinguer un actant syntaxique d'un modifieur dans la sous-section 2.2. A cela s'ajoute le fait que les syntagmes prépositionnels des deux cas dans les exemples (47) à (50) et ne sont pas obligatoires comme on le voit dans les exemple *c*.

Egalement, des tests sur la proximité positionnelle entre verbe et argument supposé pour établir le statut d'argument, ne nous semblent pas conclusifs. Par exemple, le fait que (51)a soit meilleur

que (51)b (Abraham 1999, qui donne l'exemple, considère même (51)b comme agrammatical) n'implique pas que le syntagme prépositionnel (*in den Stall* 'dans l'étable') doit être un argument, juste parce qu'il se place plus près du verbe qu'un argument nominal régulier (*das Pferd*, 'le cheval'). Il est facile de donner d'autres exemples où l'objet direct se place sans problème particulier entre le syntagme prépositionnel et le verbe (cf. (51)).

- (51) a. (Als erstes wird er) das Pferd in den Stall hineinführen.  
D'abord va il le cheval dans l'étable (acc.) faire\_entrer.
- b. ? (Als erstes wird er) in den Stall das Pferd hineinführen.  
Il va dans l'étable (acc.) le cheval faire\_entrer.  
'D'abord, il fera entrer le cheval dans l'étable.'
- c. (Als erstes wird er) in den Stall ein wunderschönes Pferd hineinführen.  
Il va dans l'étable (acc.) un merveilleux cheval faire\_entrer.  
'D'abord, il fera entrer un merveilleux cheval dans l'étable.'

Nous considérons qu'en allemand, les restrictions sur le placement ne peuvent pas donner beaucoup d'indication sur le statut argumental des éléments d'une phrase. Nous discuterons ce problème dans la section 6.

En fait, seul le recours à la définition sémantique du verbe peut nous aider à résoudre le problème. Si on prend comme exemple le verbe *joggen* 'courir' sous (48), on peut considérer que l'endroit où cette action a lieu n'est pas nécessaire dans la définition sémantique de ce mot tandis que sa direction fait partie de la définition. Autrement dit, un sens de *joggen* implique dans sa définition un déplacement vers un endroit, mais pas nécessairement l'endroit même où cette action peut avoir lieu. Une explication parallèle peut être proposée pour l'ensemble des exemples donnés sous (47) à (50). Les syntagmes prépositionnels à l'accusatif indiquent une direction, et font partie intégrante de la définition du verbe de mouvement en tant qu'actant sémantique tandis que ceux au datif<sup>31</sup>, indiquent juste le lieu où l'action se déroule, ce qui n'est pas compris dans la définition. C'est d'ailleurs ce qui explique pourquoi la phrase (50)b. est agrammaticale : comme le verbe *singen* 'chanter' n'inclut dans sa définition ni la directionnalité ni le lieu. Le syntagme prépositionnel *auf die Matratze* 'sur le matelas' ne peut donc apparaître à l'accusatif avec ce verbe car l'idée de direction n'y est pas incluse. Par contre, il peut être au datif (phrase (50)a.) mais dans ce cas, il sera considéré comme un modifieur qui contient lui-même l'idée de localité, contrairement au syntagme prépositionnel à l'accusatif qui acquiert son sens de directionnalité grâce à la définition du verbe. On voit donc qu'il y a une distinction entre les syntagmes prépositionnels qui sont au datif et ceux qui sont à l'accusatif, les premiers étant des modifieurs et les deuxièmes étant régis par le verbe.

Cependant cela n'exclut pas qu'un verbe sous-catégorise un syntagme prépositionnel locatif (c'est-à-dire au datif). Par exemple le verbe *wohnen* 'habiter' a besoin d'un argument prépositionnel locatif, bien qu'il ne spécifie pas la préposition même puisque toute préposition locative peut remplir cette fonction, comme nous le voyons en (52).

- (52) a. \*Er wohnt.  
Il habite  
'\*il habite.'
- b. Er wohnt unter der Brücke.  
Il habite sous le pont (datif)  
'Il habite sous le pont.'

---

<sup>31</sup> La discussion sur le datif concerne plus précisément le datif locatif.



- c. Er wohnt neben der Uni.  
 Il habite à\_côté\_de le fac  
 'Il habite à côté de la fac.'

De plus, ces mêmes prépositions qui apparaissent avec le datif locatif peuvent aussi être directement sous-catégorisées par le verbe : dans ce cas, on parle de préposition régie.

- (53) Er glaubt an (\*auf ) die (\*den) indischen Gottheiten  
 il croit à(\*sur) les (accusatif/\*datif) indien divinité  
 'Il croit aux divinités indiennes.'
- (54) Sie steht auf heißen (#heißen) Pudding  
 Elle est\_debout sur chaud (accusatif/#datif) sur pudding  
 'Elle adore le pudding chaud' '#Elle se trouve sur le pudding chaud.'

Dans la phrase (53), *glaube* 'croit' prend comme argument *an* 'à' avec un accusatif, et il n'est possible de changer ni de préposition ni de cas. Dans le cas de la phrase (54), on retombe sur la lecture littérale d'une circonstancielle locative modifiant *steht* 'est debout', dès qu'on change le cas d'accusatif, imposé pour la préposition régie, en datif.

L'examen des différentes prépositions montre que le traitement de ces dernières ne peut pas être homogène. Tout syntagme prépositionnel peut être modifieur, et dans ce cas la préposition est introduite dans la phrase avec son sens propre. Si la préposition est régie, son sens ainsi que le cas que porte son syntagme nominal sont imposés par le verbe. Les verbes directionnels constituent un cas mixte car le sens de la préposition est influencé, mais pas complètement déterminé par le verbe : ainsi, le verbe sous catégorise une préposition à l'accusatif et le sens donné à cette préposition dépend de la définition du verbe. Par contre, si le syntagme prépositionnel est au datif, il s'agit là d'un modifieur qui garde son sens propre.

Ces différentes discussions sur les prépositions et leurs arguments seront décisives dans le choix des structures des arbres élémentaires de TAG, que nous décrivons dans la section III.4.3.

## 5 Le datif libre

Dans la section 03, nous avons présenté le passif de l'allemand, en discutant la raison d'être de deux formes du passif, le passif d'état qui se construit avec l'auxiliaire *sein* 'être' et le passif de processus qui se construit avec l'auxiliaire *werden* 'devenir'. Dans cette nouvelle section, nous en introduisons un troisième, le passif datif, qu'on appelle aussi passif de *bekommen* 'recevoir'.

### 5.1 Les phénomènes du passif du datif (construction avec bekommen)

Le passif datif partage avec les deux autres formes du passif le fait qu'il fait lui aussi une « démotion du sujet », ce dernier pouvant par la suite se réaliser optionnellement dans un syntagme prépositionnel introduit par la préposition *von* 'de'. Par contre, au lieu de faire la promotion de l'actant syntaxique objet comme les passifs d'état et de processus, c'est l'actant syntaxique datif qu'il choisit de promouvoir. Par ailleurs, ces deux étapes sont obligatoires dans la construction du le passif datif.

Par conséquent, cette nouvelle forme de passif n'est possible qu'avec un nombre limité de verbes, notamment ceux qui ont au niveau de leur représentation syntaxique profonde, un actant

syntactique sujet et un actant syntaxique datif. Le verbe *helfen* constitue un exemple du passif datif :

- (55) Peter bekommt (von uns) geholfen  
 Pierre reçoit (de nous) aidé  
 'Pierre est aidé (par nous)'

Le passif datif se construit normalement avec l'auxiliaire *bekommen* 'recevoir'. Dans la langue parlée, on préfère utiliser l'auxiliaire *kriegen* 'recevoir' qui a le même sens en tant que verbe plein ; on trouve également l'auxiliaire *erhalten* 'recevoir'. Le verbe est au participe passé, comme dans le cas des autres formes.

Pour savoir si un verbe peut construire un passif datif, il faut d'abord savoir si ce verbe en question a un datif comme actant syntaxique. Or le datif apparaît souvent dans les formes dites « libres », c'est-à-dire en tant que *dativus commodi*, *dativus incommodi*, *dativus iudicantis* ou datif éthique. Dans ces cas, il s'agit en fait d'introduire un participant supplémentaire, qui est indirectement influencé par l'action exprimé par le verbe (souvent, il s'agit d'une influence positive, d'où le nom de *dativus commodi*), dans la plupart des cas en tant que propriétaire ou récepteur de l'objet du verbe. Voyons quelques exemples :

- (56) Die Maria strich ihm die Wand.  
 la Marie peint lui le mur  
 'Marie lui a peint le mur'
- (57) Die Maria beschmierte ihm die Wand.  
 la Marie salit lui le mur  
 'Marie lui a sali le mur.'
- (58) Das ist mir zu blöd.  
 Cela est moi trop bête  
 'Je trouve ça trop bête'
- (59) Peter singt mir die Kantate zu schnell.  
 Pierre chante moi la cantate trop vite  
 'Pierre chante la cantate trop vite pour moi' / 'Je trouve que Pierre chante trop vite la cantate'
- (60) Sing mir bloß nicht während der Fahrt!  
 chante me surtout pas pendant le voyage  
 'Ne me chante surtout pas pendant le voyage'
- (61) Zerkratze mir nicht den Bildschirm!  
 gratte moi pas l'écran  
 'Ne m'abîme pas l'écran en le grattant.'
- (62) Der Peter schläft der Maria immer gleich ein.  
 le Pierre dort à Marie toujours tout\_de\_suite particule  
 'Au regret de Marie, Pierre s'endort toujours tout de suite.'

La phrase (56) illustre un *dativus commodi*, la phrase (57) un cas de *dativus incommodi*, la phrase (58) un cas de *dativus iudicantis*. Enfin la phrase (59) peut avoir deux interprétations, l'une de *dativus incommodi* et l'autre de *dativus iudicantis*, la deuxième étant préférée pour des raisons sémantiques. Les phrases (60) et (61) illustre le cas de datif éthique que l'on trouve souvent avec le verbe à l'impératif. Le classement de la phrase (62) reste incertain car il pourrait

aussi bien correspondre à un datif *dativus incommodi* qu'à un datif éthique. Syntactiquement, la distinction entre un *dativus commodi* (phrase (56)) et un *dativus incommodi* (phrase (57)) n'est pas pertinente comme l'a montré Wagner 1985, mais on continue à utiliser ces termes à cause de la distinction qui existe au niveau sémantique : l'événement est perçu comme positif dans le premier cas et comme négatif dans le deuxième.

## 5.2 Application du critère de diathèse

Comme nous l'avons posé dans la section 2.2, le *critère de diathèse* nous permet de distinguer un actant syntaxique d'un modifieur. En effet, si on pose qu'il est impossible qu'un modifieur devienne argument syntaxique dans un changement de diathèse<sup>32</sup>, on est obligé de considérer comme actants syntaxiques tous les éléments qui peuvent être réalisés comme sujet dans le passif datif.

L'application de ce critère donne comme résultat que tous les *dativus commodi* et *incommodi* sont toujours des arguments car ils peuvent toujours être passivés :

(63) Er bekam (von Maria) die Wand gestrichen.

il reçoit (par Marie) le mur peint  
'Il se voit peindre le mur (par Marie)'

(64) Er bekam (von Maria) die Wand beschmiert.

il reçoit (par Marie) le mur sali  
'Il se retrouve avec un mur sali (par Marie)'

Par contre, les *dativus iudicantis* ne peuvent pas être passivés :

(65) \*Ich bekomme das zu blöd geworden.

je reçois cela trop bête devenu

(66) # Ich bekomme (von Peter) die Kantate zu schnell gesungen.

il reçoit (par Pierre) la cantate trop vite chanté  
'#Pierre chante pour moi la cantate trop vite (comme je l'aime bien).'

Dans le cas de la phrase (66), seul le cas de l'interprétation de *dativus commodi* est possible et la lecture plus plausible à l'actif (où je juge Pierre) en tant que *dativus iudicantis* est exclue.

Enfin, le cas du datif éthique est similaire au *dativus iudicantis*. Dans les exemples suivants (67) et (68) on ne peut faire la passivation que si on a une lecture de *dativus commodi* ; si on veut garder une interprétation de datif éthique, la phrase passive est agrammaticale.

(67) ?? Ich bekomme (bloß)nicht während der Fahrt gesungen!

je reçois (surtout) pas pendant le voyage chanté  
'On ne me chante pas pendant le voyage.'

(68) Ich bekomme nicht den Bildschirm zerkratzt.

je reçois pas l'écran gratté  
'On ne m'abîme pas l'écran.'

L'interprétation de l'agrammaticalité de la phrase (69) est plus délicate. Si l'on considère qu'il s'agit ici exclusivement d'un datif éthique, la passivation est impossible comme pour les autres datifs de ce type. Par contre, si l'analyse sémantique nous amène à le considérer comme un

---

<sup>32</sup> Le contraire est possible : la démotivation du sujet, qui fait partie du passif, réalise l'actant syntaxique sujet en tant que modifieur syntaxique du verbe.

*dativus incommodi*, alors l'impossibilité de la passivation provient de l'information véhiculée par le verbe.

(69) \*Die Maria bekommt von Peter immer gleich eingeschlafen.

la Marie reçoit de Pierre toujours tout\_de\_suite endormi

Les datifs éthiques et les *dativus iudicantis* sont donc des modificateurs, et la plupart des *dativus commodi* et *incommodi* doivent être considérés comme actants syntaxiques. Engel 1977:166 considère un tel procédé comme désavantageux « parce que cela élargirait extrêmement la classe des verbes à datif et contiendrait en fin de compte pratiquement tous les verbes pleins qui peuvent exprimer une action intentionnelle »<sup>33</sup>. Nous suivons Eisenberg 1999:292 qui considère que l'apparence d'un grand nombre de verbes trivalents n'est pas gênant dans le cas de notre analyse de l'allemand.

Il est intéressant de noter que ce procédé ne crée pas beaucoup de « nouveaux » verbes bivalents nominatif-datif, bien que beaucoup de verbes intransitifs peuvent apparaître avec un datif éthique qui ne se distingue pas nettement du *datif incommodi* (qui peut être passivé et qui est donc considéré comme argument). Indépendamment d'une explication sémantique éventuelle de ce phénomène, il faudra disposer de l'information lexicale pour chaque verbe indiquant s'il peut ou non prendre un actant syntaxique au datif.

En résumé, la plupart des verbes transitifs (c'est-à-dire qui ont au niveau de la syntaxe profonde comme actants syntaxiques un sujet et un objet) acceptent aussi un datif et deviennent donc trivalents. Seuls les verbes transitifs dont le sujet n'est pas un agent sémantique et l'objet n'est pas un patient subissant un changement d'état ne prennent pas de datif comme troisième argument (70)

(70) Das Buch kostet (\*mir, datif) nur ein Euro

ce livre coûte (\*à moi) seulement un Euro

'Ce livre coûte seulement un euro' mais 'Ce livre m'a coûté un euro.'

(71) Die Teekanne enthält ( ?? mir) nur Kaffee

Cette théière contient ( ?? à moi) seulement café

1. 'Cette théière ne contient que du café' ou 2. 'il me semble que cette théière ne contient que du café.'

La deuxième interprétation en tant que *datif iudicantis* est possible mais elle reste toujours très difficile.

Cependant, même si on inclut le passif datif parmi les voix, le *critère de diathèse* ne suffit pas pour recenser tous les actants syntaxiques au datif. Ainsi la phrase (72) ne construit pas de passif au datif ((72)b.) et on a besoin de faire appel à la définition sémantique du verbe pour donner au nom réalisé au datif le statut d'actant.

(72) a. Peter ist (mir) aufgefallen

Pierre est (à moi) fait\_remarquer

'Pierre se fait remarquer (à moi).'

b.\*Ich bekam von Peter aufgefallen

je reçois par Pierre fait\_remarquer (passivation impossible)

---

<sup>33</sup> Texte original : „weil dann die Klasse der Dativverben extrem ausgeweitet würde, die ein absichtliches Tun ausdrücken können“, traduit par mes propres soins.

### 5.3 Cas structurel et cas lexical

Dans beaucoup de travaux (notamment Müller 1999, Haider 1985:71, Abraham 1995 mais aussi dans les livres de grammaire), on trouve un classement des cas en *cas structurel* et *cas lexical*. En effet, on considère généralement qu'un cas est structurel quand il apparaît en alternance avec un autre cas dans des diathèses différentes ; par contre il est lexical, s'il est assigné de manière obligatoire pour la réalisation d'un verbe donné et ne peut changer de relation syntaxique selon la diathèse choisie. En termes de la Théorie Sens-Texte cela peut être compris comme suit : une relation de syntaxe profonde qui n'a qu'un seul correspondant au niveau syntaxique de surface est dite lexicale, car on considère que le cas (i.e. la relation de surface) est déterminé lexicalement ; quand la relation profonde correspond à plusieurs relations de surface, et cela en fonction de la voix choisie, la relation profonde est dite structurelle, car dépendant de la structure choisie.

Abraham 1995 comprend la dichotomie entre *cas structurel* et *cas lexical* dans des termes de la théorie du Gouvernement et du Liage : un cas structurel est « généré » dans le syntagme verbal (donc il fait partie du syntagme verbal dans la structure syntagmatique « profonde ») et un cas lexical est généré à l'extérieur du syntagme verbal. Il donne trois critères qui établissent la différence de comportement entre un syntagme à l'accusatif et un syntagme au datif qui indiquent, dans cette perspective, que le datif ne fait pas partie du syntagme verbal, et qu'il est donc un cas structurel<sup>34</sup>. Son analyse est étroitement liée au cadre théorique et ne s'applique donc pas à notre problématique.

Dans l'analyse traditionnelle du passif « standard », on considère que le nominatif et l'accusatif sont généralement des cas structurels et que le datif est un cas lexical. Le génitif est généralement lui aussi un cas lexical, sauf quand il est modifieur, dans ce cas il est considéré comme structurel. Des justifications de cette dichotomie peuvent être trouvées notamment dans l'existence des constructions à contrôle de l'accusatif (*AcI*, accusativus cum infinitivo, cf. section 2.1), où le sujet du verbe enchâssé est réalisé à l'accusatif, et jamais au génitif ou au datif.

Cependant, la situation se complique si on accepte de classer le passif datif comme un changement de diathèse au même niveau que le passif de l'accusatif. Dans ce cas, le datif devient lui aussi un cas structurel, et seuls les rares noms au génitif sous-catégorisés par un verbe comme dans l'exemple 32 restent des cas structurels.

---

<sup>34</sup> Il est difficile à vérifier ces données. Une raison est que nous ne partageons pas les mêmes jugements sur l'acceptabilité des phrases. Abraham 1985 donne par exemple le couple (i) et (ii) (phrases (9) a et b, page 5, la première est considérée comme mauvaise, la deuxième comme bonne) pour établir une plus grande proximité entre verbe et accusatif qu'entre verbe et datif. Nous considérons les deux phrases comme complètement exclues, et même si on crée une situation syntaxique dans lequel nous acceptons le groupement du verbe avec un de ses arguments, nous ne constatons pas de différence remarquable. (Par contre, comme il s'agit d'un exemple d'un verbe support, nous aurons quelques différences de structures dans le complexe verbal, voir section II.3)

- (i) \**[Armen geben] glaubte Frieda nicht, daß Herbert Almosen würde.*  
*[aux\_pauvres donner] pense Frieda pas, que Herbert l'aumône voudrait*
- (ii) \**[ Almosen geben] glaubte Frieda nicht, daß Herbert Armen würde.*  
*[l'aumône donner] pense Frieda pas, que Herbert aux\_pauvres voudrait*  
*sens intentionnel : Frieda ne pense pas que Herbert ferait l'aumône aux pauvres.*
- (iii) *[die Almosen gegeben] hat Herbert den Armen.*  
*[l'aumône donnée] a Herbert aux pauvres.*
- (iv) *[den Armen gegeben] hat Herbert die Almosen.*  
*[aux pauvres donnée] a Herbert l'aumône.*  
*Herbert a donné l'aumône aux pauvres.*

- (73) Ich gedenke seiner Pfannkuchen  
 je songe à\_ses\_ crêpes (génitif)  
 'Je me rappelle de ses crêpes.'

Etant donné cette nouvelle dichotomie, la question qui se pose est de savoir si une telle distinction est vraiment nécessaire, car elle crée seulement une distinction de certains noms au génitif des autres noms. La différenciation structurel/lexical pose plus de problème qu'elle résout. Par exemple, Müller, qui fait cette distinction entre cas lexical et cas structurel en se basant sur le passif d'accusatif (Müller 1999:270), défend que le passif datif fonctionne exactement comme un vrai passif (p.297), ce qui rend son propos contradictoire puisque le cas accusatif est vu comme un cas structurel et le cas datif comme un cas lexical, bien que les deux aient un fonctionnement similaire.

Nous considérons donc qu'une telle distinction est sans grande utilité, surtout si notre but est de lexicaliser la grammaire. En effet, dans une telle grammaire, toutes les possibilités de réaliser les actants syntaxiques d'un verbe doivent être prises en compte par un arbre élémentaire distinct. Ainsi, s'il n'y a qu'une seule réalisation possible de l'actant syntaxique, alors le nombre d'arbres sera plus réduit. On ne fera pas la distinction entre le cas d'unicité de la réalisation et le cas de plusieurs diathèses possibles. Nous reviendrons sur ce point dans le chapitre III.4.4.

## 6 La notion d'ordre libre des mots

Aux différents phénomènes complexes de l'allemand que nous avons décrits dans les sections précédentes, s'ajoute le facteur de l'ordre des mots. En effet, on décrit souvent l'allemand comme une langue à ordre des mots assez libre, rendu possible grâce à son système casuel développé. Mais à quoi renvoie l'idée d'un ordre libre ?

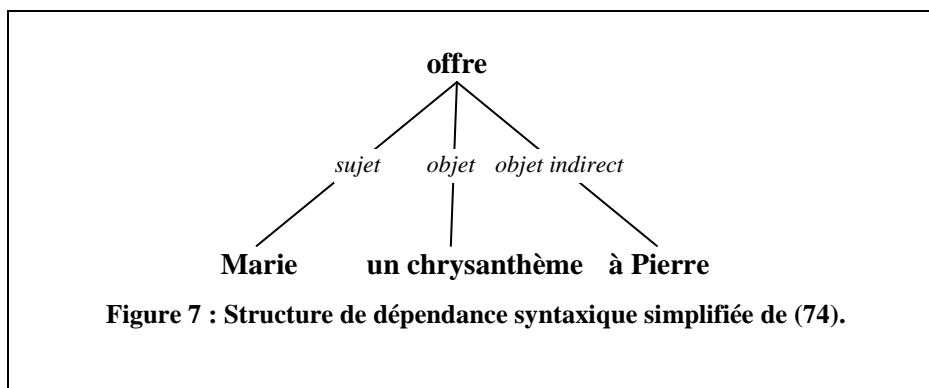
### 6.1 Un ordre libre des mots

D'une manière générale, on considère que l'ordre entre deux éléments syntaxiques est libre si la phrase reste grammaticale même après l'échange de place de ces éléments. Il va de soit qu'il ne s'agit pas d'un ordre libre des mots l'un par rapport à l'autre, mais d'une liberté de placement entre deux éléments dont la structure interne peut ne pas être libre.

L'ordre des mots d'une phrase est donc étroitement liée à l'organisation syntaxique de cette phrase. Plus précisément, le placement d'un élément (ainsi que la question de savoir s'il se place librement ou pas) se pose par rapport à un autre élément : on dit qu'un élément se place librement par rapport à un autre élément si pour tous les ordres de mots possibles d'une analyse syntaxique donnée, la relation syntaxique de ces deux éléments reste inchangée.

Ainsi, l'ordre entre le syntagme nominal « un chrysanthème » et le syntagme prépositionnel « à Pierre » est libre dans les phrases sous (74) parce qu'on peut librement les échanger dans la phrase sans changer de sens, ni de relation syntaxique entre eux et leur gouverneur :

- (74) a. Marie offre un chrysanthème à Pierre.  
b. Marie offre à Pierre un chrysanthème.



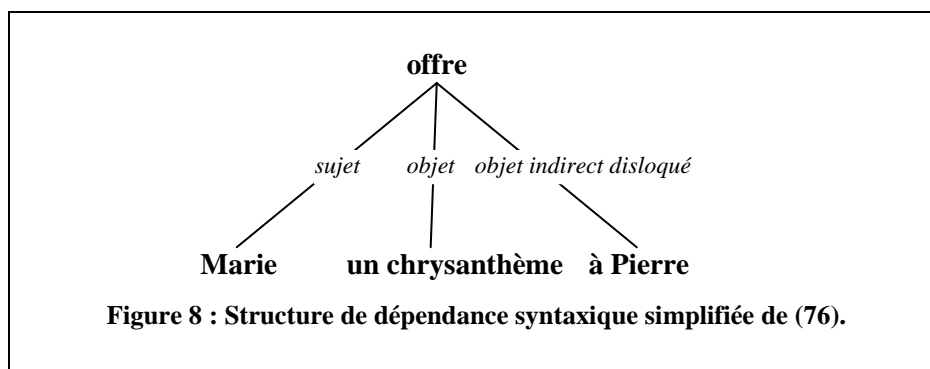
Dans les deux phrases, la même analyse syntaxique peut être appliquée : les actants syntaxiques restent les mêmes et toutes les relations syntaxiques de la phrase restent inchangées. Ces entités syntaxiques de la phrase peuvent être représentés dans une structure de dépendance syntaxique, un arbre non ordonné dont les branches marquent les différents actants syntaxiques. Nous donnons dans la Figure 7 une représentation simplifiée de la structure de dépendance syntaxique des deux phrases sous (74) où on laisse de côté pour l'instant les relations syntaxiques qui existent entre les noms et les déterminants ou leurs prépositions.

Par contre, l'ordre des mots de la phrase (74)a. n'est pas libre de s'ordonner comme dans la phrase (75) car l'analyse syntaxique n'est pas la même, dans le cas de la deuxième phrase, le syntagme nominal « un chrysanthème » étant le sujet et « Marie » l'objet (créant d'ailleurs une phrase sémantiquement bizarre).

- (75) <sup>2#</sup> un chrysanthème offre Marie à Pierre.  
(76) A Pierre, Marie offre un chrysanthème.  
(77) <sup>??</sup> Un chrysanthème Marie offre à Pierre.

Le cas de la phrase (76) est plus complexe car deux analyses syntaxiques sont possibles : dans la première analyse, nous conservons la structure que nous présentons dans la Figure 7 et nous considérons donc que le syntagme prépositionnel « à Pierre » a changé de place<sup>35</sup>. Une deuxième possibilité d'analyse est de dire que l'actant sémantique « à Pierre » n'est plus un objet indirect « ordinaire » mais de considérer qu'il s'agit d'une structure syntaxiquement différente appelée *objet indirect disloqué*. Cela pourrait être justifié justement par la distribution linéaire différente d'objet indirect disloqué et par le fait que un objet direct, par exemple, se comporte différemment (phrase (77)). Nous représentons sa structure de dépendance syntaxique simplifiée dans la Figure 8.

<sup>35</sup> Nous laissons pour l'instant indéterminé le terme de place que nous serons amenés à définir par la suite sous le nom de « champ » dans le cadre du modèle topologique.



Avec cette nouvelle analyse syntaxique, la phrase (76) n'est plus considérée comme ayant un ordre libre puisque elle ne suit plus la définition de l'ordre libre qui se base sur une préservation des relations syntaxiques. En effet, les relations syntaxiques ont changé puisqu'on n'a plus affaire à un objet indirect « ordinaire » mais à un objet indirect « disloqué ». Pourtant, les deux phrases en (74) et la phrase (76) gardent la même structure sémantique (et se distinguent dans la structure communicative, cf. section suivante). Il s'agit là d'un choix linguistique de structure syntaxique qui dans un cas permet de parler d'ordre libre, tandis que dans l'autre cas, elle ne le permet pas.

L'introduction de l'idée qu'un élément syntaxique détient une place « à lui » dans la phrase nous permet d'élargir notre définition de l'ordre libre de manière à pouvoir l'appliquer à la phrase (76). Nous proposons la définition suivante pour l'ordre libre :

*Ordre libre* : une phrase a un ordre libre si son analyse syntaxique permet plusieurs linéarisations.

Autrement dit, on peut dire qu'un élément d'une structure syntaxique se place librement soit quand il apparaît dans deux positions différentes de la phrase, soit si dans une position donnée de la phrase, il existe plusieurs éléments avec qui il peut échanger sa place.

Ainsi, le syntagme prépositionnel « à Pierre » de la structure syntaxique que nous avons présenté dans la Figure 7 se place librement puisqu'il peut prendre la position de dislocation gauche ou la position post-verbale dans les différentes linéarisations possibles. De plus, ce même syntagme se place librement car à l'intérieur de la position post-verbale, il peut échanger sa place avec celle du syntagme nominal « un chrysanthème ».

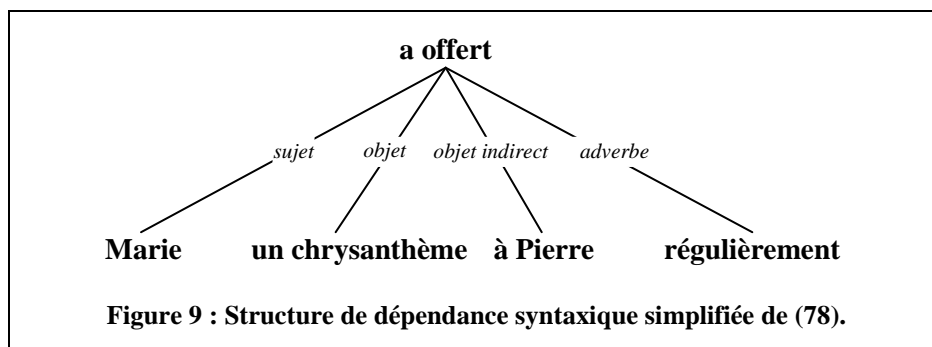
Cette définition de l'ordre libre peut être appliquée aux phrases présentées sous (78) :

- (78) a. Régulièrement, Marie a offert un chrysanthème à Pierre.  
 b. Marie a régulièrement offert un chrysanthème à Pierre.  
 c. Marie a offert régulièrement un chrysanthème à Pierre.  
 d. Marie a offert un chrysanthème régulièrement à Pierre.  
 e. Marie a offert un chrysanthème à Pierre régulièrement.  
 f. ? Régulièrement, à Pierre, Marie a offert un chrysanthème.  
 g. ?? Régulièrement, Marie a à Pierre offert un chrysanthème.

Si l'on faisait la distinction entre adverbes de phrase et adverbes de syntagme, les phrases (78)a. et (78)b. auraient une analyse syntaxique différente. Cependant, nous considérons qu'une telle distinction n'est pas nécessaire et nous proposons que toutes les phrases sous (78) peuvent être analysées dans une dépendance syntaxique, que nous représentons dans la Figure 9<sup>36</sup>.

<sup>36</sup> Il s'agit encore une fois d'une représentation syntaxique, où nous n'avons pas pris en compte l'auxiliaire. Par ailleurs, si l'on voulait représenter les deux analyses possibles selon qu'on prend en compte qu'il s'agit d'un adverbe de phrase ou de syntagme, il suffirait de reprendre l'arbre et de





Dans cette structure, l'adverbe « régulièrement » et le syntagme prépositionnel « à Pierre » peuvent prendre tous les deux soit la place de dislocation gauche soit la place post-verbale. Dans la position post-verbale, leur place est considérée comme libre puisque non seulement ils peuvent mutuellement échanger leur place, mais aussi se placent-ils librement par rapport au syntagme nominal objet « un chrysanthème ». De plus, l'adverbe a la possibilité de se placer entre l'auxiliaire et le participe passé. D'ailleurs, dans un langage parlé, le syntagme prépositionnel peut aussi occuper cette place. Nous voyons donc que la structure de dépendance syntaxique permet plusieurs linéarisations de la phrase, et on peut considérer que l'ordre de ses mots est libre. On peut alors être amené à dire que le français est aussi une langue à ordre de mots libres, ce qui va à l'encontre des caractéristiques habituelles que l'on trouve pour une description du français.

Cependant, si l'on regarde en détail les phrases sous (78), on s'aperçoit que bien que les relations sémantiques des différentes linéarisations restent identiques, la situation dans laquelle elles peuvent être énoncées n'est plus du tout la même. Chaque phrase correspond à un contexte différent. Ainsi, chacune des phrases de (78) se distingue soit par sa structure communicative (Mel'čuk 2001<sup>37</sup>), soit par la portée de l'adverbe. Or la définition que nous avons donnée pour l'ordre libre ne prend pas en compte les distinctions communicatives ou sémantiques, et ne repose que sur un critère syntaxique de préservations des actants et des relations. En effet, si on prenait en compte le critère communicatif et sémantique, il serait impossible de parler d'ordre libre car chaque linéarisation d'une dépendance syntaxique donnée correspond à un contexte communicatif spécifique. C'est dans ce sens qu'on peut être amené de dire qu'une langue comme le français n'a pas d'ordre libre. Finalement, dire que le français a un ordre de mots libre ou non dépend strictement de la définition que l'on choisit pour rendre compte de ce phénomène.

## 6.2 Et l'allemand ?

Revenons à l'allemand. L'ordre libre des mots est une caractéristique que l'on met en avant dans toute description de cette langue. Et en effet, l'analyse est parallèle à celle du français puisqu'on trouve plusieurs linéarisations possibles pour une structure de dépendance donnée, où les relations syntaxiques sont préservées.

Pourtant, lorsqu'on compare les deux langues, on entend souvent que l'ordre des mots est plus libre en allemand qu'en français. Mais qu'implique un tel propos ?

La phrase (79) qui est un exemple parallèle à celui du français que nous avons donné sous (74) a elle aussi un ordre libre pour les mêmes raisons que pour sa correspondante française : « dem Peter » et « Blumen » peuvent échanger leur place tout en préservant le même sens et les mêmes relations syntaxiques, comme on peut le voir dans (79)b. La même structure syntaxique (Figure 10) peut donc être proposée pour les deux phrases sous (79).

---

remplacer la relation *adverbe* qui existe entre le verbe et « régulièrement » par *adverbe de phrase* d'une part et *adverbe de syntagme* d'autre part.

<sup>37</sup> On trouve plus fréquemment le terme de structure informationnelle, Lambrecht 1994. Nous discutons de cette notion dans la sous-section suivante 6.3.

(79) a. die Maria hat dem Peter Blumen geschenkt<sup>38</sup>.

la Marie (nominatif) a au Pierre (datif) fleurs (accusatif) offert

b. die Maria hat Blumen dem Peter geschenkt.

la Marie (nominatif) a fleurs (accusatif) au Pierre (datif) offert

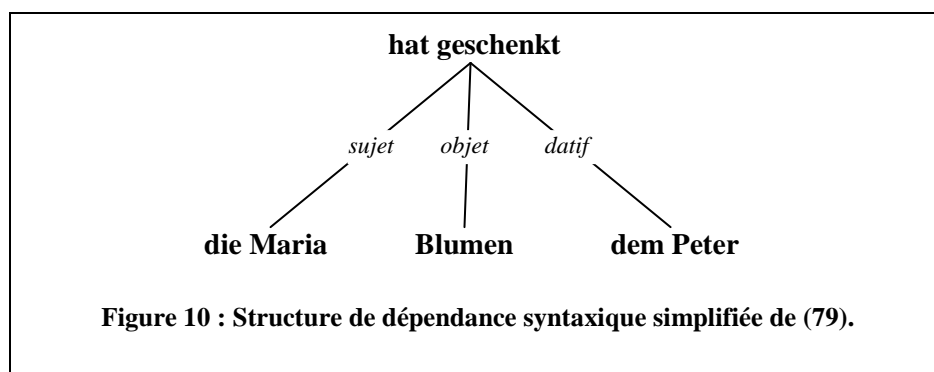
'Marie a offert des fleurs à Pierre.'

Par contre, contrairement au français, l'allemand n'a pas de place de dislocation gauche. « dem Peter » ne peut donc pas être déplacé dans une position préverbale, et la phrase est considérée comme agrammaticale.

(80) \*dem Peter die Maria hat Blumen geschenkt.

au Pierre (datif) la Marie (nominatif) a fleurs (accusatif) offert

S'il en est ainsi, on devrait considérer le français comme une langue à ordre des mots plus libre



que l'allemand. Néanmoins, une telle affirmation serait trop rapide car la situation est plus complexe : en effet, des ordres qui sont interdits en français (un exemple est donné en (82), se trouvent être tout à fait possibles en allemand, comme on peut le voir dans les phrases en (81) .

(81) a. dem Peter hat die Maria Blumen geschenkt.

au Pierre (datif) a la Marie (nominatif) fleurs (accusatif) offert

b. dem Peter hat Blumen die Maria geschenkt.

au Pierre (datif) a fleurs (accusatif) la Marie (nominatif) offert

c. Blumen hat dem Peter die Maria geschenkt.

fleurs (accusatif) a au Pierre (datif) la Marie (nominatif) offert

d. Blumen hat die Maria dem Peter geschenkt.

fleurs (accusatif) a la Marie (nominatif) au Pierre (datif) offert

(82) \* à Pierre a Marie offert les fleurs

L'ordre (81)a. est possible en français comme on peut le voir dans l'exemple (83), mais dans ce cas, les relations syntaxiques ne sont pas préservées : dans (83), Pierre est le sujet et Marie l'objet indirect.

(83) Pierre a offert à Marie les fleurs

<sup>38</sup> Les déterminants sur les prénoms sont optionnels. Ils sont souvent utilisés dans un langage familier.

On voit donc apparaître des restrictions. En français, la place devant le verbe ne peut être occupée que par le sujet sauf quand il y a inversion (ce qui est lexicalement et stylistiquement très restreint, et qui est de toute façon impossible dans l'exemple ci-dessus). Par contre en allemand, c'est le marquage casuel des éléments nominaux, qui permet de montrer les relations syntaxiques qui existent entre les différents éléments et le verbe qui permet d'avoir des linéarisations diversifiées.

Parler d'ordre libre, que ce soit à l'intérieur d'une langue ou en comparaison de deux langues, a du sens seulement si on explicite par rapport à quelle invariante la langue est libre. Et cette invariante, qui peut être, comme nous le proposons, la dépendance syntaxique, est un objet dépendant de l'analyse linguistique de la phrase. C'est ainsi que différentes analyses linguistiques peuvent nous amener à différentes affirmations sur la liberté de l'ordre de mots d'une langue.

### **6.3 Les structures communicatives**

En fait, s'il existe une telle liberté d'ordre des mots pour une même structure syntaxique donnée (comme nous l'avons décrit dans la section précédente pour le français et l'allemand) il apparaît clairement que ce n'est pas pour exprimer juste la relation syntaxique qu'entretiennent les mots entre eux. Déjà pour des raisons d'économie d'expression, il est impossible qu'un outil comme l'ordre des mots, qui consiste à donner une forme linéaire à une pensée non linéaire, apparaisse d'une manière arbitraire et que deux linéarisations qui ont la même dépendance syntaxique de surface renvoient à exactement la même « pensée ». Reprenons les exemples (74) et (76) que nous redonnons sous (84) :

- (84) a. Marie offre un chrysanthème à Pierre.  
 b. A Pierre, Marie offre un chrysanthème.

Ces deux phrases ont aussi bien la même structure syntaxique que la même structure sémantique. Et pourtant, tout locuteur natif de cette langue sera d'accord de dire qu'il existe une différence entre ces deux phrases : elles se distinguent au fait par leur structures communicatives (Polguère 1990 et Mel'čuk 2001) ou encore informationnelles (Lambrecht 1994).

Analyser la structure communicative d'une phrase, revient à voir comment le contenu de la phrase, l'information, est « communiquée », c'est-à-dire transmise. En effet, la prise de la parole consiste à dire quelque chose à propos de quelque chose. Autrement dit, toute structure communicative de la phrase peut être analysée comme une partie thématique (ce dont on parle) et une partie rhématique (ce qu'on en dit). Nous suivons Mel'čuk 2001 et nous supposons que la distinction entre rhème (ou encore « focus ») et thème (ou encore « topique ») est la première étape essentielle de la structure communicative. Une phrase ne contient pas obligatoirement une partie thématique, mais doit avoir toujours un rhème : autrement dit, la prise de parole implique qu'on a quelque chose à transmettre, et il peut s'agir d'une information complètement nouvelle. Dans la Théorie Sens-Texte, la structure communicative est présente à chacun des niveaux de représentations linguistiques et fait partie intégrante de la représentation complète de chaque niveau.

La structure communicative d'une phrase peut être très complexe. Mel'čuk 2001 considère qu'il existe au delà de la thématique, encore sept autres oppositions communicatives qui permettent de décrire cette structure : l'opposition donné nouveau, le paramètre de focalisation, de perspective, d'émphasisation, de présupposition, d'unitarité et de locutionnalité. Chacune des oppositions peut porter une valeur négative, positive, et pour certaines d'entre elles, la valeur neutre. Ces oppositions peuvent se combiner pour donner la structure communicative complète d'une phrase, mais toutes les phrases ne sont jamais concernées par toutes les huit oppositions.

Dans le cadre de ce travail, nous nous contenterons de la distinction principale de thémativité, combinée au paramètre de focalisation. Plus précisément, nous nous rangeons aux côtés de Vallduvì 1992 et nous nous contentons d'une division tripartite de la phrase, avec une partie focus, une partie topique et une partie « tail ». En fait, Vallduvì pose la division focus/ground comme la division communicative fondamentale, le *focus*<sup>39</sup> renvoyant à la partie rhématique et le *ground* à la partie thématique de la phrase. Par ailleurs, cette partie thématique peut être subdivisée en deux parties, le topique (Vallduvì emploie le terme de *link* 'gauche'), une partie thématique proéminente, qui renvoie à un thème focalisé dans la terminologie de Mel'čuk 2001, et le *tail* (terme introduit par Vallduvì, correspondant à 'queue'), une partie thématique non proéminente (un thème non focalisé selon Mel'čuk 2001) qui suit généralement un élément rhématique focalisé, le focus. Nous considérons que cette division tripartite permet d'expliquer certains phénomènes d'ordre de l'allemand, sans avoir à rentrer dans les détails de la structure communicative complète<sup>40</sup>.

Il est notable de constater que les constituants communicatifs (dans notre analyse, focus topique et tail) partagent des caractéristiques de syntaxiques, sémantiques et prosodiques que l'on retrouve à travers diverses langues. Nous proposons les quelques critères suivants permettant de distinguer un élément topique d'une phrase déclarative du français et de l'allemand :

- les topiques se trouvent toujours au début d'un énoncé ;
- un topique ne constitue pas un élément obligatoire de la phrase et une phrase peut comporter plusieurs topiques : on peut donc avoir zéro, un ou plusieurs topiques successifs à l'intérieur d'une même phrase ;
- chaque topique d'une phrase est marquée par un contour prosodique particulier : la courbe intonative qui l'accompagne est montante et un topique (ou un groupe de topiques) est suivie souvent d'une pause ;
- un topique T peut être repris dans une paraphrase de type « à propos de T... ». Aussi, si T est un topique de la phrase, la phrase peut être une réponse à une question du type « Et sur T, (qu'est-ce que tu veux me dire) ? »
- en français, on constate que les éléments disloqués à gauche sans ou avec reprise pronominale sont toujours des topiques. Par contre, le contraire n'est pas vrai : tous les éléments topiques ne sont pas toujours disloqués à gauche.

Ainsi, dans la phrase (84)b., *à Pierre* est un topique. La phrase peut être paraphrasé par la phrase (85) et elle peut être une réponse à la phrase (86) :

(85) A propos de Pierre Marie lui offre un chrysanthème.

(86) Et à Pierre ?

De la même manière, dans la phrase (84)a. *Marie* peut être vu comme un topique si on retrouve les caractères prosodiques du topiques, comme il apparaît clairement dans la phrase (87).

(87) Marie, elle offre un chrysanthème à Pierre.

En allemand, la situation est similaire que celle du français. L'élément qui se trouve devant le verbe fini, par exemple le constituant *dem Peter* peut être avoir l'intonation topique, c'est-à-dire

---

<sup>39</sup> La notion de *focus* qui a été introduit originellement par Jackendoff 1972 est, à la suite des travaux de Lambrecht (notamment son livre de 1994) la notion qui est le plus couramment utilisée dans la littérature récente. Elle renvoie à l'idée du cœur de l'information, mis en relation avec le contexte.

<sup>40</sup> Nous renvoyons le lecteur à Choi 1999 pour une étude de l'ordre des mots en allemand et en coréen dans le cadre de la théorie de l'optimalité basée sur le formalisme LFG. Choi considère un jeu de deux traits binaires pour le traitement de l'ordre des mots dans les deux langues : la distinction thème/focus et la distinction emphatique/non emphatique. Notre approche se rapproche de l'analyse de Choi, sauf que nous considérons que toute partie rhématique de la phrase est aussi focalisée.

être accompagné d'une courbe mélodique montante ; par ailleurs, il peut constituer une réponse à la question sous (89).

(88) dem Peter hat die Maria Blumen geschenkt.  
au Pierre (datif) a la Marie (nominatif) fleurs (accusatif) offert  
'A Pierre, Marie a offert des fleurs.'

(89) Und Peter ? Was kannst du mir über den sagen?  
et Pierre ? que peux tu me sur lui dire ?  
'Et sur Pierre, que peux-tu me dire sur lui ?'

En allemand, la dislocation gauche n'est pas possible. Par conséquent, le deuxième élément topical peut se retrouver après le verbe fini. La phrase (88) peut donc aussi avoir une structure communicative différente, avec deux topiques, *dem Peter* et *die Maria*, de part et d'autre de l'auxiliaire *hat* qui lui n'est communicativement pas marqué.

Passons maintenant à ce que nous considérons être focus d'une phrase. Un élément focal est défini comme suit :

- une phrase déclarative a exactement un seul focus ;
- Le focus suit les topiques de la phrase ;
- On peut ajouter l'adverbe « probablement » devant l'élément focal, sans que cela implique un changement communicatif dans l'énonciation ;
- l'élément focal F d'une phrase représente une question contenant un pronom interrogatif, F étant l'élément qui correspond à ce pronom. De la même manière F est un élément focal si F remplace un élément C d'une manière contrastive ;
- En français, les éléments clivés sont très souvent des focus.
- Le contour intonatif d'un focus est descendant ; ce qui suit un focus est toujours très bas et faible en intensité.

Ainsi, Les phrases sous (90) illustrent l'application de ces différents critères pour définir *un chrysanthème* comme le focus de ces phrases. Dans tous les cas, le contour intonatif qui accompagne l'élément *un chrysanthème* est descendant, tout ce qui étant d'un contour plat et bas.

- (90) a. C'est un chrysanthème que Marie offre à Pierre.  
b. Marie offre probablement un chrysanthème à Pierre.  
c. Qu'offre Marie à Pierre ?  
d. Marie offre un chrysanthème à Pierre, et non un tournesol.

Ces différents critères permettant de distinguer le focus s'appliquent également pour l'allemand, moyennant quelques complications concernant les verbes.

Ainsi, dans la phrase (91), on peut considérer que *Blumen* 'des fleurs' constitue le focus de la phrase si la courbe intonative qui l'accompagne est descendante, et que la phrase, munie d'une prosodie correcte, est une réponse acceptable aux questions (92) et (93).

(91) dem Peter hat die Maria Blumen geschenkt.  
au Pierre (datif) a la Marie (nominatif) fleurs (accusatif) offert  
'A Pierre, Marie a offert des fleurs.'

(92) Was hat die Maria dem Peter geschenkt.  
que a la Marie (nominatif) au Pierre (datif) offert  
'Qu'a offert Marie à Pierre ?'

(93) dem Peter hat die Maria Blumen geschenkt oder dem Peter hat die Marie Gummibärchen geschenkt. ?

au Pierre (datif) a la Marie (nominatif) fleurs (accusatif) offert ou au Pierre (datif) a la Marie (nominatif) oursions-Haribo (accusatif) offert  
 'Marie a offert des fleurs à Pierre ou des oursions Haribo ?'

Par contre, la situation se complique quand le focus consiste d'un verbe non fini et d'un de ses arguments non verbaux qui est placé juste avant lui : dans ce cas, le contour prosodique est différent : le focus n'est pas marqué sur le dernier élément qui est le verbe mais déjà sur l'argument non verbal qui se trouve devant lui.

Examinons les exemples suivants.

(94) Maria will im Netz surfen.

Maria veut dans\_le\_ouèbe surfer  
 'Maria veut surfer sur le ouèbe.'

(95) Wo will Maria surfen? Im Netz oder im Meer?

où veut Maria surfer ? sur\_le ouèbe ou dans\_la mer  
 'Où veut surfer Marie ? Sur le ouèbe ou dans la mer ?'

(96) Was will Maria (tun) ?

que veut Marie (faire) ?  
 'Que veut faire Marie ?'

Si on suppose que la phrase (94) a un contour descendant sur *Netz* 'ouèbe', on peut considérer qu'il s'agit d'une réponse à la question (95), où seul *im Netz* est l'élément focal, mais aussi qu'il s'agit d'une réponse à la question (96). Le contour descendant ne peut pas apparaître sur le verbe *surfen* 'surfer' s'il s'agit d'une réponse à (95) ou (96) ; par contre, un tel contour est possible si l'on veut répondre à la question sous (97) :

(97) Was will Maria im Netz (tun)?

que veut Marie sur\_le ouèbe (faire) ?  
 'Que veut faire Maria sur le net ?'

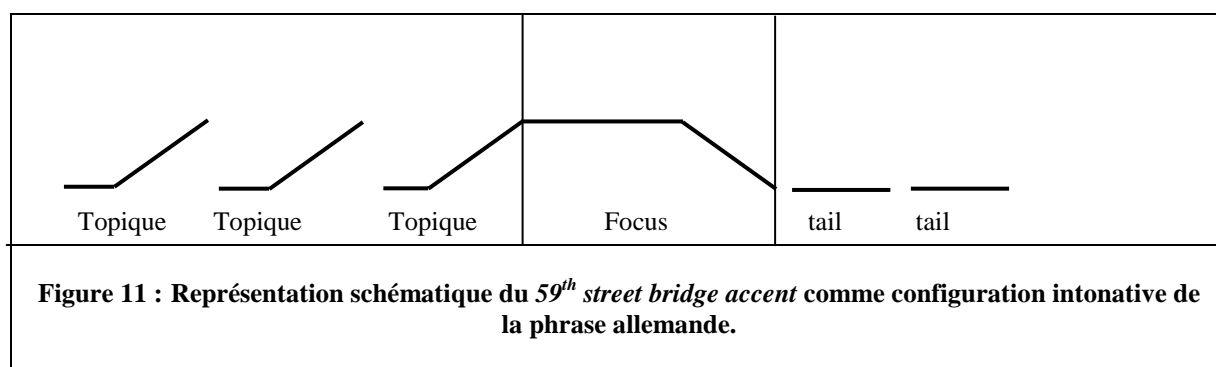
Nous reviendrons sur cette question dans la section IV.2.

Le dernier élément communicatif considéré est le thème non proéminent, qu'on appelle *tail*. Nous considérons que les critères suivants permettent de définir le *tail* :

- un tail suit toujours un focus.
- un tail constitue une information thématique dans le sens que son contenu est déjà présent dans le contexte et l'absence d'un tail ne gêne pas à la compréhension de la phrase.
- le tail t peut apparaître en tant que tel dans une question, et la réponse t ne constitue pas une nouvelle information.
- un tail est marqué par un contour intonatif bas et plat, souvent avec peu d'intensité.

Ainsi, dans les exemples que nous avons considérés ci-dessus, le *tail* est constitué par l'élément qui suit la partie focale de la phrase.

Il apparaît donc que le découpage en constituants communicatifs repose en grande partie sur l'organisation prosodique de la phrase. En générale, une séquence de topiques (avec une contour intonatif montant) est suivie d'un focus (contour intonatif descendant) qui est suivie d'une séquence de tails (contour intonatif bas et plat). Ainsi, pour la phrase allemande, nous obtenons la configuration intonative en forme de chapeau, que Büring 1995 appelle le *59<sup>th</sup> street bridge accent* que nous reproduisons schématiquement dans la Figure 11:



Comme nous l'avons déjà remarqué, le topique n'est pas obligatoire dans une phrase. Une phrase peut aussi bien commencer par un focus. Ainsi, pour une phrase telle que (94), *Maria* peut aussi bien être un topique et un focus : les deux lectures seront bien distinguées par un contour prosodique différent. Nous voyons donc que l'élément qui se situe avant le verbe fini peut avoir deux valeurs communicatives, puisqu'il peut être soit topique, soit focus. Par contre, cet élément ne peut jamais être un tail puisque une phrase nécessite exactement un focus et qu'un tail suit toujours le focus.

Cette distinction tripartite de la structure communicative, certes incomplète, nous sera utile tout au long de ce travail dans les discussions sur les différentes possibilités d'ordres des mots. Cette sous-section ne vise pas à discuter les termes utilisés ou les éléments communicatifs que nous retenons mais de poser clairement les outils qu'on utilise pour bien définir ce dont on parle et éviter toute possible confusion (notamment de termes).

#### 6.4 Ordre des mots standard

Il n'existe probablement aucune langue pour laquelle l'ordre des mots a été étudié d'une manière aussi intensive et variée que l'allemand. Et pourtant très peu d'approches prennent comme point de départ l'arbre de dépendance syntaxique bien que le statut de ce dernier semble moins problématique qu'un arbre syntagmatique X-barre, qui contient entre autres, des informations sur l'ordre et la dépendance des éléments (encodée à l'aide de la notion de tête). En effet, dans la plupart des approches X-barre, la discussion porte presque exclusivement sur l'ordre des éléments (ordre SVO, SOV etc.) et la dépendance sous-jacente est prise comme acquise. Le travail de Rambow 1994 est, à ma connaissance, le seul travail sur l'allemand qui essaie de proposer une liste presque exhaustive des différentes possibilités d'ordres des mots d'une phrase donnée. Rambow 1994 procède de la manière suivante : à partir de l'ensemble des mots d'une phrase donnée, il teste les différents ordres que ces mots peuvent prendre, sans essayer de faire un lien direct avec la dépendance sous-jacente.

Beaucoup d'approches prennent le chemin inverse de celui que nous avons choisi : ainsi, elles proposent une analyse d'un phénomène d'ordre donné (par exemple, l'extraposition) en montrant que beaucoup d'exemples, souvent pris dans différentes langues, vérifient l'analyse donnée, pour ensuite poser la question de comment distinguer ce phénomène d'ordre, d'autres phénomènes voisins (par exemple, l'intraposition). Nous ne critiquons pas cette approche et nous pensons qu'elle peut être utile dans le traitement de certains phénomènes (nous utilisons cette approche nous même dans cette thèse). Cependant, nous considérons qu'il est intéressant et complémentaire de proposer des analyses parallèles basées sur la variabilité d'ordre que nous obtenons à partir d'un arbre de dépendance syntaxique donné.

Dans la suite de la section, nous discuterons le terme d'*ordre standard*, puis nous reprendrons certains phénomènes, souvent discutés à part, qui mettent en jeu des structures de verbes non

finis, tels que la dislocation interne au syntagme, le brouillage d'argument, l'extraction lointaine, l'antéposition et l'intrapolation.

## 6.5 L'ordre standard : standard par rapport à quoi ?

Nous définissons l'ordre dit *standard* ou encore *non marqué* de la manière suivante :

Pour une dépendance syntaxique de surface D, l'ordre standard est, parmi les différentes *linéarisations*<sup>41</sup> de D, celle qui permet le plus grand nombre d'interprétations communicatives possibles ainsi que le plus grand nombre de lectures possibles, ces lectures renvoyant à différentes structures prosodiques qui peuvent être réalisées pour la même phrase.

La définition que nous fournissons diffère légèrement de celle proposée par entre autres, Lennerz 1977, Wabelhuth 1985, Abraham 1986 et Choi 1999. Par contre, notre définition étant placée dans le cadre de la Théorie Sens-Texte, rend explicite l'invariant des différents ordres : la structure de dépendance syntaxique de surface.

En fait cette invariante est souvent implicite chez ces différents auteurs. Ainsi, Choi 1999:29 affirme que les différents ordres qu'elle considère « ont plus ou moins le même sens vériconditionnel »<sup>42</sup>. De la même manière, Lennerz 1977:27 donne la définition suivante : « Quand deux parties de la phrase A et B peuvent aussi bien apparaître dans l'ordre AB que dans l'ordre BA, et que BA ne peut apparaître que sous des conditions véritables contraintes par AB, alors AB est 'l'ordre non marqué' est AB 'l'ordre marqué' »<sup>43</sup>. Or cette définition, si elle prise sans la condition implicite que le sens (ou la structure syntaxique) de la phrase en question reste inchangée, ne permet pas de décider laquelle des deux phrases (par exemple, entre les phrases a. et b. sous (98)) est considérée comme l'ordre standard.

(98) a. Peter himmelt Maria an

Pierre idolâtre Marie particule

'Pierre idolâtre Marie' ou, avec une prosodie très marquée 'Marie idolâtre Pierre'.

b. Maria himmelt Peter an

Marie (nom ou acc) idolâtre Pierre (nom ou acc) particule

'Marie idolâtre Pierre' ou, avec une prosodie très marquée 'Pierre idolâtre Marie'.

Cependant, même si la définition de Lennerz était comprise avec la condition sous-jacente qu'un sens vériconditionnel reste invariant, un problème persiste car on se retrouve avec d'autres constructions, comme par exemple des phrases avec différentes diathèses mais aussi des nominalisations ayant un sens recherché. Il faut donc comprendre cette définition en considérant l'idée suivante comme sous-jacente : l'invariance ne porte pas seulement sur le sens vériconditionnel<sup>44</sup>, mais aussi sur l'ensemble des mots fléchis qui sont utilisés pour exprimer ce sens.

A la différence de ces auteurs, la définition de l'ordre standard que nous proposons ne met en jeu ni le sens de la phrase ni la forme des mots qui sont utilisés pour réaliser ce sens, mais exclusivement une structure de dépendance syntaxique de surface.

<sup>41</sup> Par *linéarisation*, nous définissons simplement l'ordre linéaire qui existe entre les différents mots en jeu. La linéarisation que nous proposerons dans la suite de ce travail comprendra aussi une structure de constituants qu'on appelle *structure topologique*.

<sup>42</sup> Texte original : «... have more a less the same truth-conditional meaning », ma propre traduction.

<sup>43</sup> Texte original : « Wenn zwei Satzglieder A und B sowohl in der Abfolge AB wie in der Abfolge BA auftreten können, und wenn BA nur unter bestimmten, testbaren Bedingungen auftreten kann, denen AB nicht unterliegt, dann ist AB die 'unmarkierte Abfolge' und AB die 'markierte Abfolge'. »

<sup>44</sup> Ce sens vériconditionnel est interprété dans la Théorie Sens-Texte par un graphe sémantique où s'exprime la structure prédicat-argument.

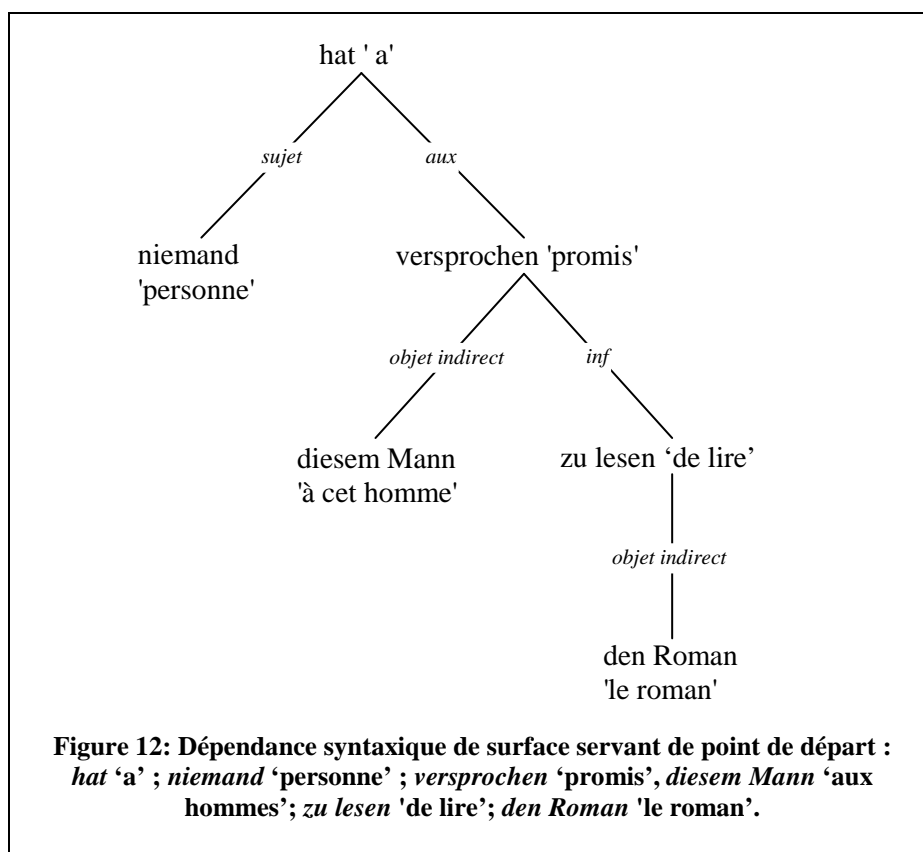


Deux remarques s'imposent encore sur la notion d'*ordre standard* :

- il est très difficile de proposer des généralisations sur l'ordre standard et on trouve une diversité de caractéristiques chez les auteurs qui traitent de cette question. Par exemple, Uszkoreit 1987:117 inclut dans sa définition d'ordre standard le fait que les arguments au datif doivent précéder les arguments à l'accusatif. Par contre, Müller 1999:11.4. montre que l'ordre entre les arguments au datif et à l'accusatif est un phénomène plus complexe qui prend en compte des restrictions lexicales mais aussi purement sémantiques (d'où la différence entre (99) et (101)). Par ailleurs, Hoberg 1991 a montré à partir d'une étude de corpus, qu'un argument animé est presque toujours placé devant un argument non animé (d'où la différence entre (99) et (100)).

- (99) a. ich gebe die Blumen dem Mann. (ordre marqué)  
je donne les fleurs à homme
- b. ich gebe dem Mann die Blumen. (ordre non marqué)  
je donne à homme les fleurs  
'je donne les fleurs à l'homme.'
- (100) a. ich gebe dem Mann das Kind. (ordre marqué / ? non marqué)  
je donne à homme l'enfant
- b. ich gebe das Kind dem Mann. (ordre marqué / ? non marqué)  
je donne l'enfant à homme  
'je donne l'enfant à l'homme.'
- (101) a. ich überlasse dem Mann das Kind. (ordre marqué)  
je laisse à homme l'enfant
- b. ich überlasse das Kind dem Mann. (non marqué)  
je laisse l'enfant à homme  
'je laisse l'enfant à l'homme.'

Wegener 1985 introduit l'idée d'un degré « d'implication » (de l'allemand « Involviertheit ») de l'argument nominal comme base de l'ordre des mots. Nous considérons qu'on peut même trouver une interaction entre les arguments de différents verbes : on peut donc définir un *ordre standard* pour une structure syntaxique de surface donnée, mais pas pour un prédicat hors contexte.



- Le fait qu'on puisse définir un ordre standard pour une dépendance donnée, n'implique dans aucun cas que ce dernier est plus « basique » dans un sens transformationnel. Nous considérons que l'ordre des éléments dépend des restrictions syntaxiques et communicatives qui, à partir de la structure syntaxique de surface, définissent ensemble un ordre précis. Nous spécifierons ce propos dans la section IV.5. Nous pouvons même considérer que l'ordre standard est par définition l'ordre qui est le plus communicativement ambigu, puisqu'il s'agit de la linéarisation qui offre le plus d'interprétations communicatives possibles. Il est donc difficile de considérer que ce dernier constitue d'une manière ou d'une autre une étape intermédiaire dans la réalisation langagière d'une idée. L'existence de cette ambiguïté est pourtant essentielle pour un système information de génération de textes qui ne dispose pas d'un module de calcul de la structure communicative. En effet, nous montrerons dans le Chapitre II comment une grammaire d'Arbres Adjoints génère au moins l'ordre dit standard pour chaque sens donné.

## 6.6 Les différents phénomènes d'ordre des mots

Après avoir discuté la notion d'*ordre standard*, nous passons maintenant aux différents phénomènes qu'on trouve comme caractérisant un problème d'ordre des mots.

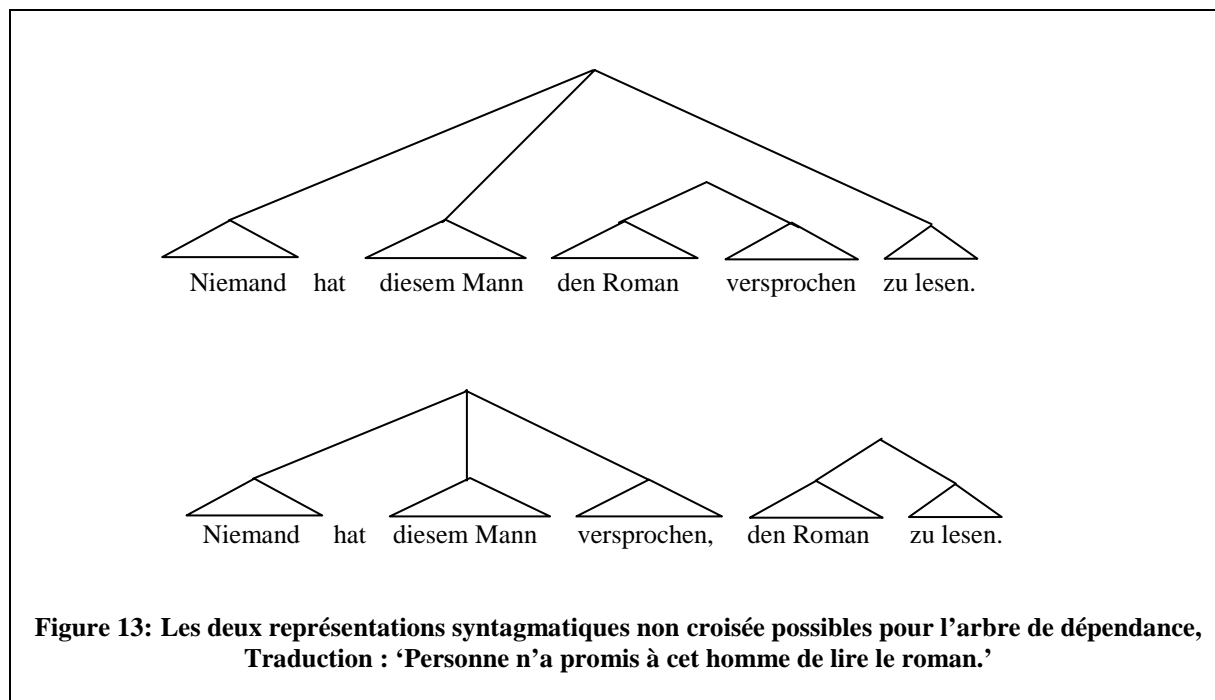
Comme nous l'avons affirmé dans la sous-section précédente, l'arbre de dépendance syntaxique de surface constitue le point de départ de notre analyse. Considérons donc l'arbre de dépendance<sup>45</sup> de la Figure 12.

Dans les analyses proposées par les grammaires basées sur une structure syntagmatique on considère souvent une coïncidence entre l'ordre standard et l'ordre qui donne une représentation

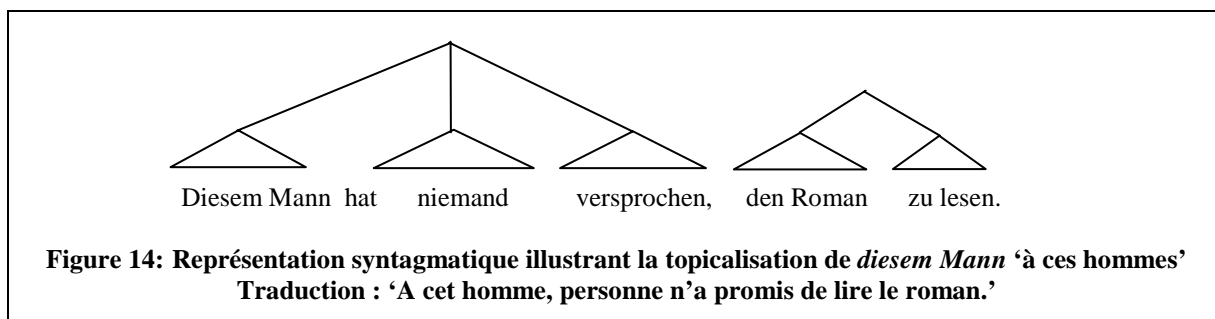
<sup>45</sup> Nous ne discutons pas ici la structure interne du syntagme nominal et nous renvoyons le lecteur à la section IV.6 pour l'analyse des syntagmes nominaux.

syntagmatique sans branches croisées (ce qui n'est pas toujours vrai comme on le verra dans la section III.2.4).

Dans les figures suivantes, nous donnons une structure syntagmatique très simplifiée pour quelques phénomènes d'ordre des mots. D'une manière générale, on trouve des analyses avec des constituants intermédiaires, des regroupements comme par exemple celui du syntagme verbal. Sans prendre partie pour une analyse particulière, nous considérons la structure la plus simple, donc la structure plate pour nos exemple. Les résultats seront identiques si on choisissait une représentation avec un enchâssement quelconque. Dans les figures, nous donnons une structure syntagmatique qui contient le verbe et les éléments nominaux qui correspondent aux théta-rôles du verbe. Dans le cas de l'arbre de dépendance de la Figure 12, deux choix se présentent pour l'ordre standard, et il n'est pas clair laquelle des deux ordres que nous présentons dans la Figure 13 est la moins marquée :



En fait, il y a une différence de groupement d'information dans ces deux ordres des mots possibles : dans le premier, l'action *zu lesen versprochen* 'promettre de lire' est considéré comme une unité tandis que le deuxième, c'est juste *den Roman* 'le roman' qui forme une unité avec le verbe à l'infinitif *zu lesen* 'de lire'. En accord avec les termes habituels, nous référons définissons le premier cas comme *ordre standard* et le deuxième cas *extraposition*.



Il est possible d'avoir un ordre interne du syntagme différent. La représentation suivante (Figure 14) illustre le phénomène qu'on trouve sous le nom de *topicalisation*<sup>46</sup> (Ross 1967) ou

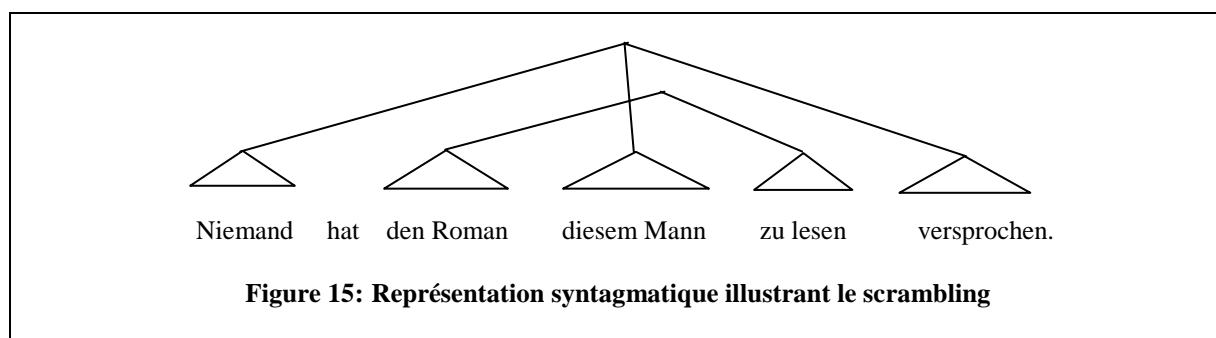
<sup>46</sup> Précisons que le terme de *topicalisation* renvoie à un phénomène syntaxique et n'est pas lié l'unité communicative de *topic* comme nous l'entendons.

d'*inversion du sujet* où le syntagme *diesem Man* 'à cet homme' a échangé sa place avec le syntagme nominal *niemand* 'personne'.

La structure de dépendance syntaxique de surface donne lieu à d'autres linéarisations que nous répertorions dans la liste suivante :

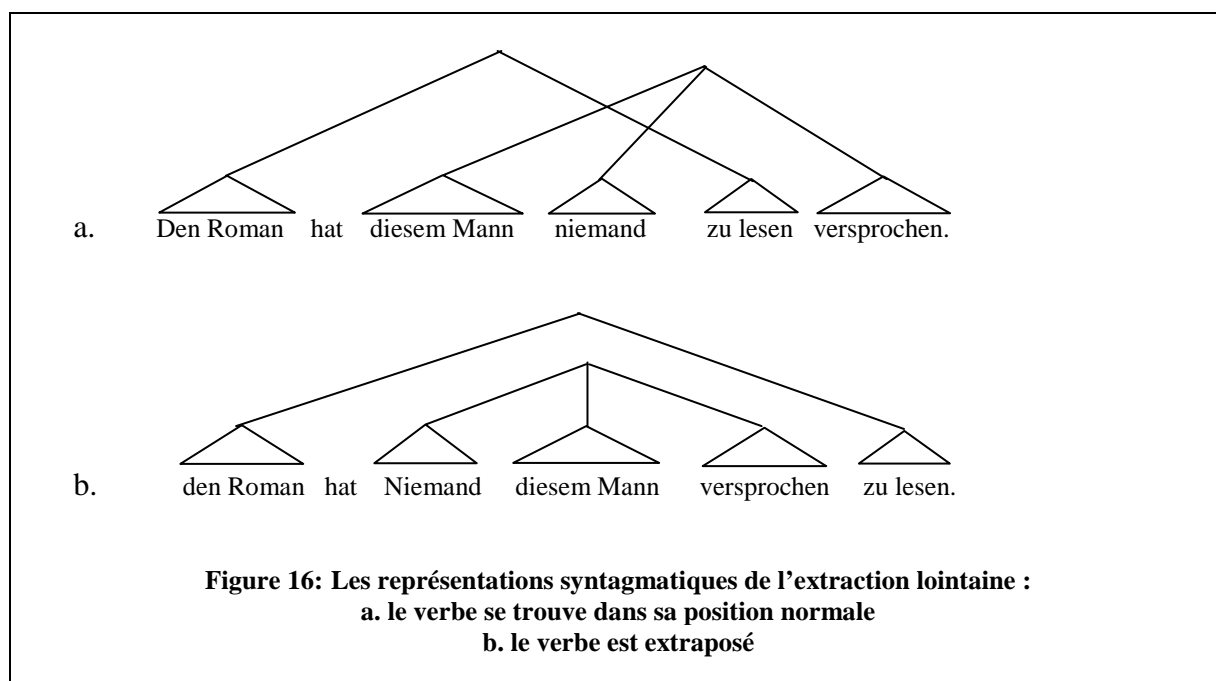
- brouillage d'arguments (*scrambling*)

Le terme malheureux de *scrambling* 'brouillage d'arguments' a été introduit par Ross 1967 pour décrire les phénomènes où les structures de constituants de différents verbes « croisent leurs branches » afin de joindre leurs arguments qui se trouvent normalement entre le verbe fini et les autres verbes non finis (Figure 15).



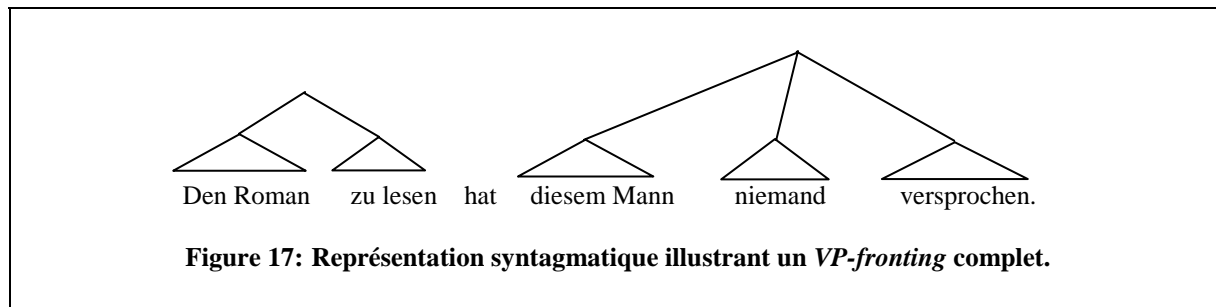
- extraction lointaine

Un argument d'un verbe enchâssé (par exemple l'objet) peut être placé en début de phrase, aussi bien quand le verbe se trouve dans sa position normale (Figure 16a.) que quand il est extraposé (Figure 16b.) :



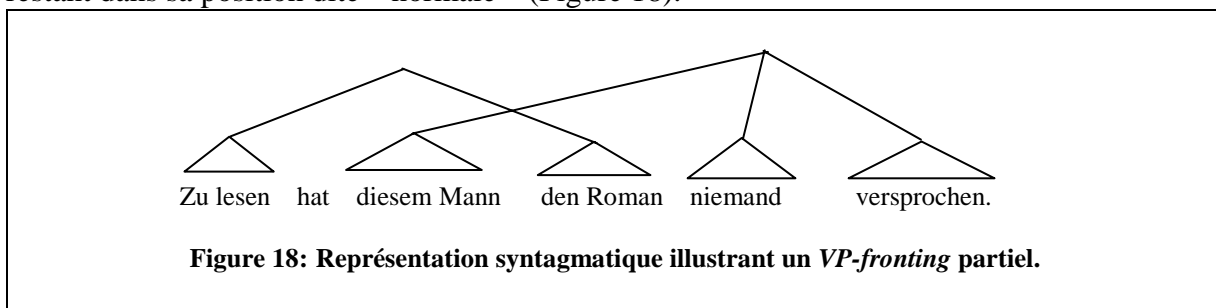
- *VP-fronting* complet (antéposition complète du syntagme verbal)

On parle de *VP-fronting* lorsque le verbe enchâssé est placé en début de phrase, juste après ses arguments : dans le cas présent, l'antéposition est dite complète car elle concerne tout le syntagme verbal (le verbe et son argument) (Figure 17).



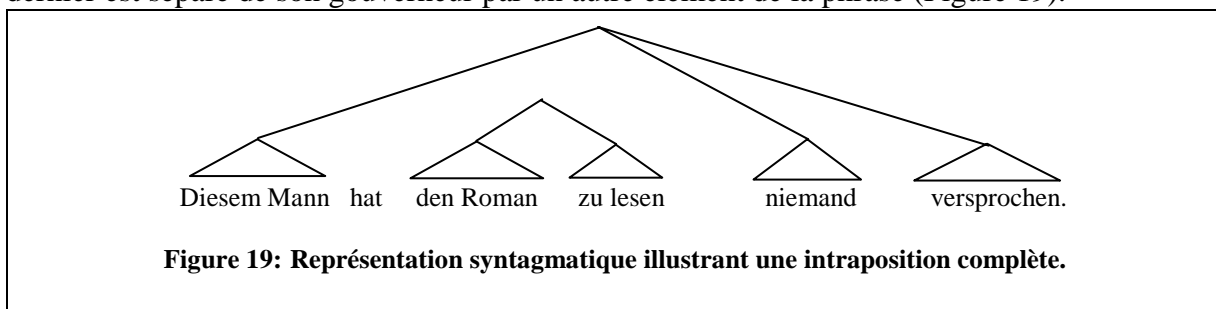
- *VP-fronting* partiel (antéposition partielle du syntagme verbal)

L'antéposition est dite partielle car seul le verbe enchâssé sort en début de phrase, son argument restant dans sa position dite « normale » (Figure 18).



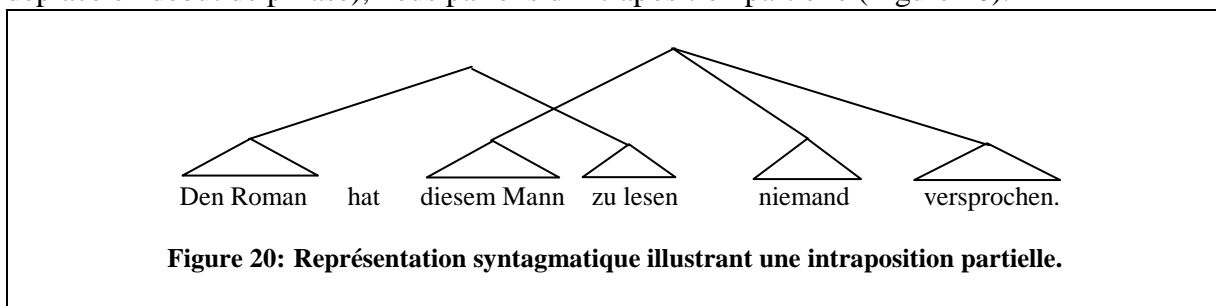
- Intraposition complète

Par *intraposition complète*, on réfère au cas où le verbe enchâssé et ses arguments se trouvent placés entre son gouverneur verbal et le verbe fini. Mais contrairement à l'ordre standard, ce dernier est séparé de son gouverneur par un autre élément de la phrase (Figure 19).

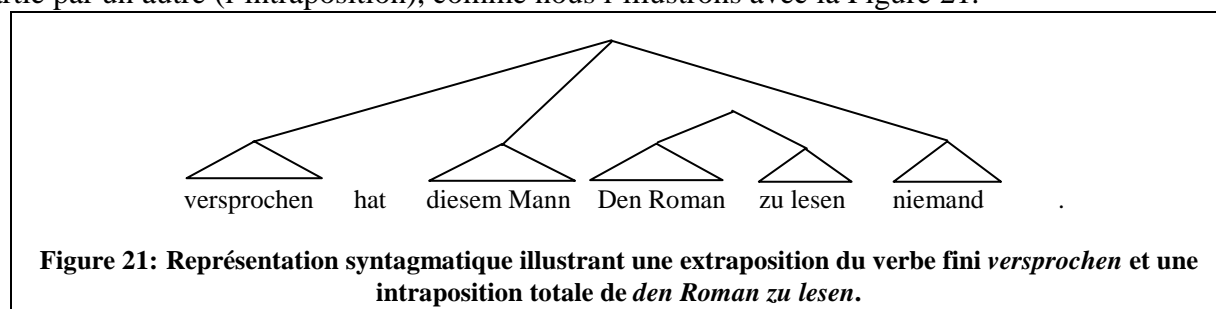


- Intraposition partielle

Quand le verbe enchâssé se trouve intraposé sans l'un de ses arguments qui a été par exemple déplacé en début de phrase), nous parlons d'intraposition partielle (Figure 20).



Pour compléter l'image (effrayante) de ces quelques ordres des mots de l'allemand très souvent discutés par les syntacticiens de cette langue, il convient de préciser que ces phénomènes peuvent apparaître dans une seule même phrase et ce de manière croisée, c'est-à-dire où une partie de la phrase est concernée par un phénomène (l'extraposition par exemple) et une autre partie par un autre (l'intrajonction), comme nous l'illustrons avec la Figure 21.



Les quatre-vingt dix-neuf résultats de ces croisements sont indiqués dans un ordre alphabétique dans la Figure Figure 22 suivante<sup>47</sup>. Nous précisons que nous n'avons pas inclus dans cette liste les cas d'extraposition nominale car un tel positionnement n'est possible que pour les syntagmes nominaux dits « lourds », c'est-à-dire des syntagmes comportant un nombre important de syllabes (hormis les relatives, cf. section III.2.4) et de plus, ne sont pas à l'accusatif. En fait, une structure de dépendance syntaxique de surface comportant des syntagmes nominaux lourds donnerait d'avantage d'ordres. Par ailleurs, le tableau ne montre que les structures de la phrase déclarative, les phrases interrogatives étant réalisées par l'intermédiaire d'autres arbres. Notons également que bien que le chiffre 99 puisse sembler impressionnant, nous sommes loin des 720 (= 6!) ordres qui sont théoriquement possibles pour un arbre contenant six éléments. Les 621 autres possibilités étant absolument impossibles, on peut considérer que le chiffre 99 représente un tri assez important des différents ordres des mots.

<sup>47</sup> Notons que la liste que nous présentons n'est pas directement comparable à celle donnée par Rambow 1994 car quelques unes des phrases qu'il propose sont analysées sans leur structure syntaxique sous-jacente.

Pour notre exemple de base, nous avons choisi le prédicat *lesen* 'lire' au lieu de *reparieren* 'réparer' (le verbe qu'utilise Rambow 1994), parce que *lesen* 'lire' accepte plus difficilement un datif commode, ce qui rend la structure de dépendance syntaxique de la phrase moins ambiguë. En effet, le contraste qu'on constate entre (i) et (ii) est de nature sémantique : on ne voit pas en quoi Karl peut aider ou déranger mon père en lisant le roman (la lecture à haute voix étant lexicalisée différemment en allemand : *vorlesen* 'lire à haute voix à quelqu'un')

- (i) Meinem Vater den Kühlschrank zu reparieren hat Karl versprochen.  
à mon père (dat.) le frigo (acc.) à réparer a Charles promis.  
Réparer le frigo pour mon père, Charles l'a promis.
- (ii) ?? Meinem Vater den Roman zu lesen hat Karl versprochen.  
à mon père (dat.) le roman (acc.) à lire a Charles promis.  
Lire le roman pour mon père, Charles l'a promis.

Il en suit que beaucoup d'ordre des mots de la phrase (i) sont ambigus en ce qui concerne le gouverneur syntaxique du syntagme nominal au datif : soit c'est *versprochen* 'promis', soit c'est *zu reparieren* 'à réparer'. Les choix lexicaux de la phrase (ii) rendent cette différence plus claire.

1. den Roman diesem Mann versprochen zu lesen hat niemand
2. den Roman diesem Mann zu lesen versprochen hat niemand
3. den Roman hat diesem Mann niemand versprochen zu lesen
4. den Roman hat diesem Mann niemand zu lesen versprochen
5. den Roman hat diesem Mann zu lesen niemand versprochen
6. den Roman hat niemand diesem Mann versprochen zu lesen
7. den Roman hat niemand diesem Mann zu lesen versprochen
8. den Roman hat niemand zu lesen diesem Mann versprochen
9. den Roman hat zu lesen diesem Mann niemand versprochen
10. den Roman hat zu lesen niemand diesem Mann versprochen
11. den Roman versprochen zu lesen hat diesem Mann niemand
12. den Roman versprochen zu lesen hat niemand diesem Mann
13. den Roman zu lesen diesem Mann versprochen hat niemand
14. den Roman zu lesen hat diesem Mann niemand versprochen
15. den Roman zu lesen hat niemand diesem Mann versprochen
16. den Roman zu lesen versprochen hat diesem Mann niemand
17. den Roman zu lesen versprochen hat niemand diesem Mann
18. diesem Mann den Roman versprochen zu lesen hat niemand
19. diesem Mann den Roman zu lesen versprochen hat niemand
20. diesem Mann hat den Roman niemand versprochen zu lesen
21. diesem Mann hat den Roman niemand zu lesen versprochen
22. diesem Mann hat den Roman zu lesen niemand versprochen
23. diesem Mann hat niemand den Roman versprochen zu lesen
24. diesem Mann hat niemand den Roman zu lesen versprochen
25. diesem Mann hat niemand versprochen den Roman zu lesen
26. diesem Mann hat niemand zu lesen den Roman versprochen
27. diesem Mann hat zu lesen den Roman niemand versprochen
28. diesem Mann hat zu lesen niemand den Roman versprochen
29. diesem Mann versprochen den Roman zu lesen hat niemand
30. diesem Mann versprochen hat den Roman niemand zu lesen
31. diesem Mann versprochen hat zu lesen niemand den Roman
32. diesem Mann versprochen hat niemand den Roman zu lesen
33. diesem Mann versprochen hat niemand zu lesen den Roman
34. diesem Mann versprochen hat zu lesen den Roman niemand
35. diesem Mann versprochen hat zu lesen niemand den Roman
36. diesem Mann versprochen zu lesen hat den Roman niemand
37. diesem Mann versprochen zu lesen hat niemand den Roman
38. diesem Mann zu lesen den Roman versprochen hat niemand
39. diesem Mann zu lesen versprochen hat den Roman niemand
40. diesem Mann zu lesen versprochen hat niemand den Roman
41. niemand hat den Roman diesem Mann versprochen zu lesen
42. niemand hat den Roman diesem Mann zu lesen versprochen
43. niemand hat den Roman zu lesen diesem Mann versprochen
44. niemand hat diesem Mann den Roman versprochen zu lesen
45. niemand hat diesem Mann den Roman zu lesen versprochen
46. niemand hat diesem Mann versprochen den Roman zu lesen
47. niemand hat diesem Mann zu lesen den Roman versprochen
48. niemand hat zu lesen den Roman diesem Mann versprochen
49. niemand hat zu lesen diesem Mann den Roman versprochen
50. versprochen den Roman zu lesen hat diesem Mann niemand
51. versprochen den Roman zu lesen hat niemand diesem Mann
52. versprochen hat den Roman diesem Mann niemand zu lesen
53. versprochen hat den Roman diesem Mann zu lesen niemand
54. versprochen hat den Roman niemand diesem Mann zu lesen
55. versprochen hat den Roman niemand zu lesen diesem Mann
56. versprochen hat den Roman zu lesen diesem Mann niemand
57. versprochen hat den Roman zu lesen niemand diesem Mann
58. versprochen hat diesem Mann den Roman niemand zu lesen
59. versprochen hat diesem Mann den Roman zu lesen niemand
60. versprochen hat diesem Mann niemand den Roman zu lesen
61. versprochen hat diesem Mann niemand zu lesen den Roman
62. versprochen hat diesem Mann zu lesen den Roman niemand
63. versprochen hat diesem Mann zu lesen niemand den Roman
64. versprochen hat niemand den Roman diesem Mann zu lesen
65. versprochen hat niemand den Roman zu lesen diesem Mann
66. versprochen hat niemand diesem Mann den Roman zu lesen
67. versprochen hat niemand diesem Mann zu lesen den Roman
68. versprochen hat niemand zu lesen den Roman diesem Mann
69. versprochen hat niemand zu lesen diesem Mann den Roman
70. versprochen hat zu lesen den Roman diesem Mann niemand
71. versprochen hat zu lesen den Roman niemand diesem Mann
72. versprochen hat zu lesen diesem Mann den Roman niemand
73. versprochen hat zu lesen diesem Mann niemand den Roman
74. versprochen hat zu lesen niemand den Roman diesem Mann
75. versprochen hat zu lesen niemand diesem Mann den Roman
76. versprochen zu lesen hat den Roman diesem Mann niemand
77. versprochen zu lesen hat den Roman niemand diesem Mann
78. versprochen zu lesen hat diesem Mann den Roman niemand
79. versprochen zu lesen hat diesem Mann niemand den Roman
80. versprochen zu lesen hat niemand den Roman diesem Mann
81. versprochen zu lesen hat niemand diesem Mann den Roman
82. zu lesen den Roman diesem Mann versprochen hat niemand
83. zu lesen den Roman versprochen hat diesem Mann niemand
84. zu lesen den Roman versprochen hat niemand diesem Mann
85. zu lesen diesem Mann den Roman versprochen hat niemand
86. zu lesen diesem Mann versprochen hat den Roman niemand
87. zu lesen diesem Mann versprochen hat niemand den Roman
88. zu lesen hat den Roman diesem Mann niemand versprochen
89. zu lesen hat den Roman niemand diesem Mann versprochen
90. zu lesen hat diesem Mann den Roman niemand versprochen
91. zu lesen hat diesem Mann niemand den Roman versprochen
92. zu lesen hat niemand den Roman diesem Mann versprochen
93. zu lesen hat niemand diesem Mann den Roman versprochen
94. zu lesen versprochen hat den Roman diesem Mann niemand
95. zu lesen versprochen hat den Roman niemand diesem Mann
96. zu lesen versprochen hat diesem Mann den Roman niemand
97. zu lesen versprochen hat diesem Mann niemand den Roman
98. zu lesen versprochen hat niemand den Roman diesem Mann
99. zu lesen versprochen hat niemand diesem Mann den Roman

**Figure 22: Les 99 possibilités de linéariser l'arbre de dépendance syntaxique de la Figure 12**

Enfin, comme tout phénomène d'ordre libre, il est clair que les différents que nous obtenons correspondent à différents regroupements de l'information et à différentes valeurs que portent ces groupes, et ce en fonction de la structure communicative en jeu qui peut être parfois très complexe. Par exemple, le phénomène de topicalisation correspond à un regroupement des éléments topicalisés, qui peuvent porter soit la valeur communicative de thème soit de rhème ; deux valeurs clairement distinctes d'un point de vue prosodique. Par ailleurs, le contexte d'une intraposition est déjà beaucoup plus complexe puisqu'on a affaire à une double thématisation, celle d'un élément qui est placé devant le verbe fini et celle d'un élément qui est placé après le verbe fini. Nous discuterons ces possibilités en détail dans la section IV.5.

Parmi les phrases de cette liste, certaines d'entre elles nécessitent un contexte communicatif très spécifique et quelques approches préfèrent les déclarer syntaxiquement fausses. Par exemple, les phrases 28 30 et 33 de la liste que nous reportons sous (102) sont souvent considérés comme agrammaticales car le contexte communicatif qu'elles requièrent est difficilement reconstituable.

- (102) ? Diesem Mann hat zu lesen niemand den Roman versprochen. (phr. 28 de la liste)  
(103) ? Diesem Mann versprochen hat den Roman niemand zu lesen. (phr. 30 de la liste)  
(104) ? Diesem Mann versprochen hat niemand zu lesen den Roman. (phr. 33 de la liste)

Le survol de ces différents phénomènes qui illustrent l'ordre des mots en allemand montre à quel point cette question semble diversifiée et difficile à capter sous un même angle et qu'elle peut être rendue plus complexe si on regarde le groupement informationnel. Néanmoins, nous considérons qu'il s'agit juste d'une question d'approche : l'approche que nous adoptons est d'abord syntaxique, dans le sens qu'on regarde la combinaison de différents phénomènes d'ordres puis on ajoute le traitement de la structure communicative. Cette information communicative permet de générer **la** bonne phrase qui correspond à la structure syntaxique munie de l'information communicative. Dans une application de génération, il est très rare qu'on soit amené à générer les phrases sous (102) car il n'arrive presque jamais qu'on soit confronté à des structures communicatives aussi complexes où la génération de ces phrases est nécessaire. Cependant, cela ne veut pas dire qu'il est impossible de produire de telles phrases. Notre approche cherche à considérer tous les ordres possibles et seulement les ordres possibles, pour une structure de dépendance syntaxique de surface.

Comment donc arrivons nous à ces 99 phrases ? Quelles restrictions sont à la base de ces possibilités ? Qu'est-ce qui les distinguent des 621 phrases non possibles ? Peut-on fournir une analyse simple pour l'ensemble de ces 99 phrases ?

Nous tenterons de répondre à ces différentes questions dans la section suivante.

Un des buts principaux du chapitre suivant sera d'analyser de plus près tous les possibilités d'ordres des mots en allemand et de montrer comment une structure non ordonnée de la dépendance syntaxique de surface peut être linéarisé à l'aide d'une structure topologique, une structure qui formalise l'idée d'une hiérarchie de positions fixes.



## Chapitre II

# Topologie

## Le modèle topologique de la phrase allemande

# Topologie

## Le modèle topologique de la phrase allemande

*Dans cet espace de jouissance, prendre quelque chose de borné, fermé, c'est un bien, et en parler c'est une topologie<sup>48</sup>.*

Il est très difficile de parler de l'ordre des mots en allemand sans utiliser la notion de « place » car, « la position des mots dans la phrase allemande n'est pas **relative**, comme on l'enseigne trop souvent avec la théorie de l'**inversion**, ce qui brouille tout, mais bien **absolue**, puisque le verbe ne se place pas **après** tel ou tel autre mot, mais bien, selon la nature de la phrase, à une place fixe et déterminée en elle-même » (Tesnière 1959:130)<sup>49</sup>. Le but de ce chapitre est de présenter la formalisation de l'ordre des mots en allemand, basée justement sur cette idée de « place ». Cette formalisation résulte de travaux menés en collaboration avec Sylvain Kahane. Dans cette section, nous reprenons en grande partie différentes publications, notamment Gerdes & Kahane 2002.

Dans la Théorie Sens-Texte, l'ordre des mots est introduit dans le passage du niveau syntaxique de surface au niveau morphologique profond. Bien que Mel'čuk considère l'étude de ce passage comme une « tâche fascinante<sup>50</sup> », très peu de recherches dans la cadre de la Théorie Sens-Texte portent sur la les possibilités d'une linéarisation non projective. Le travail le plus poussé dans cette direction est probablement Mel'čuk 1965 sur l'ordre des mots en russe<sup>51</sup>. Dans Mel'čuk 1988, la structure complète de morphologie profonde n'apparaît qu'une seule fois, notamment dans la figure 2-7 page 71 de son livre. Selon Mel'čuk 1988, le niveau morphologique profond « spécifie la forme d'une phrase particulière en termes de mots-formes et des syntagmes qui la constituent, ainsi que leur ordre linéaire, sans considérer l'organisation interne du mot-forme lui-même »<sup>52</sup>. Et pourtant, la notion de « syntagme » ou encore de « constituant morphologique » est le parent pauvre de la Théorie Sens-Texte, bien que dans la définition de la représentation morphologique profonde, il est considéré qu'elle consiste en deux structures :

- une structure morphologique, soit une chaîne linéaire ordonnée de représentations morphologiques profondes de tous les mots-formes qui constituent la phrase, arrangée dans leur ordre linéaire que tiennent ces mots-formes dans la phrase ;
- une structure prosodique, c'est-à-dire un découpage de cette chaîne par un marquage prosodique, composé de « pauses, contours intonatifs, etc. » (*ibid.*:71)<sup>53</sup>.

Nous considérons qu'il est nécessaire de spécifier cette structure morphologique, et nous proposons de la remplacer par une nouvelle structure syntagmatique basée sur l'idée de « places ».

Il s'agit d'une structure hiérarchique et ordonnée, qu'on appelle *structure topologique*, qui entretient une correspondance directe avec la dépendance syntaxique de surface : la première

---

<sup>48</sup> Jacques Lacan, *Le Séminaire de Jacques Lacan. Livre XX*, Encore, 1972-1973.

<sup>49</sup> caractères en gras dans le texte original.

<sup>50</sup> Mel'čuk 1988:38 : « The study of the conditions under which non-projectivity is allowed is a fascinating task (which has, by the way, serious typological implications). »

<sup>51</sup> Voir aussi le travail de sur la linéarisation de l'anglais : Mel'čuk & Pertsov 1987.

<sup>52</sup> Mel'čuk 1988:69 : « A deep-morphological representation specifies the form of a particular sentence in terms of the word forms and the phrases that constitute it and their linear order, without regard to the internal organization of the word forms themselves. », ma propre traduction.

<sup>53</sup> Mel'čuk 1988:71 : « pauses, intonation contours, and the like », ma propre traduction.

contient l'information sur l'ordre des mots tandis que la deuxième exprime l'information de la sous-catégorisation sans indiquer l'ordre des éléments. Ainsi, à partir d'une dépendance syntaxique de surface, il s'agit de voir quelles sont les différents ordres, ou encore linéarisations, qui peuvent lui correspondre.

L'idée de *structure topologique* s'inspire largement de la description de l'ordre des mots de l'allemand appelée le *modèle topologique* (l'historique de ce terme sera discuté dans la section 5). Dans la suite de cette section, nous commençons par donner un aperçu des travaux qui se basent sur cette notion de topologique, des travaux traditionnels tels que Drach 1937 et Bech 1955, mais aussi à des travaux plus récents (notamment Kathol 1995 et Debusmann & Duchier 2001). Puis nous donnons les principes de linéarisations d'un arbre de dépendance syntaxique pour l'allemand et nous analysons différents ordres obtenus, tels que les ordres présentés dans la section 6 du premier chapitre, mais aussi l'ordre des verbes dans un complexe verbal et le  *pied piping*. Enfin, nous terminerons cette section en comparant notre analyse à une analyse basée sur le mouvement et nous discuterons la notion de structure topologique par rapport à d'autres structures syntagmatiques, notamment celle utilisée dans la théorie X-barre.

## 1 Les règles de base de la topologie

L'idée de base d'un modèle topologique est de considérer toute phrase comme un patron contenant différentes places, qu'on appelle champs, chacune d'entre elles pouvant accueillir différents types de constituants. Dans notre approche, nous considérons qu'il s'agit d'une structure topologique de constituants : par conséquent, placer un élément dans un ordre linéaire revient à créer des constituants topologiques, chacun d'entre eux étant organisé comme une séquence de champs qui peut de nouveau accueillir des constituants topologiques et ainsi de suite. Soulignons le fait qu'un champ n'est qu'une place pour accueillir un constituant ; dans aucun cas, on considère qu'un champ constitue lui-même une unité quelconque.

Les constituants minimaux sont lexicaux et sont désignés par le nom de la partie du discours du mot auquel ils renvoient. Autour de ces constituants minimaux, les mots peuvent ensuite créer des constituants topologiques plus larges qui peuvent accueillir certaines des dépendants de ce mot. Le domaine est le plus haut constituant que les mots peuvent créer (mais il n'est pas nécessaire qu'un mot en crée un). Les constituants non lexicaux sont constitués d'une séquence de champs. Ainsi, le *domaine principal* qui est ouvert par un verbe principal fini représente le gabarit sous-jacent de la phrase déclarative allemande. Ce dernier comporte cinq champs, le *Vorfeld*, la *parenthèse gauche*, le *Mittelfeld*, la *parenthèse droite* et le *Nachfeld*. On peut donc considérer qu'un domaine ressemble à une boîte qui contient des compartiments alignés qu'on appelle des champs, et qui peuvent à leur tour accueillir des boîtes. Dans notre formalisme, nous avons trois types de règles :

## Règles de base

- *règles de création de constituants* : ce sont des règles qui donnent le type de constituants qu'un mot peut créer et spécifient quel champ du nouvel constituant ce mot occupera.
- *règles de description des constituants* : ce sont des règles qui décrivent la liste ordonnée des champs contenus dans le constituant, en indiquant de quel type de champ il s'agit, c'est-à-dire le nombre de constituants que le champ peut ou doit contenir.
- *règles de placement des constituants* : ce sont des règles qui indiquent dans quel champ un mot peut être placé, et ceci en fonction de son gouverneur, notamment la place et le constituant de ce dernier.

Ces trois règles constituent les principes de base de notre module syntaxique qui réalise l'interface syntaxe-topologie. Nous postulons que le module syntaxique de n'importe quelle langue peut être décrit dans nos termes<sup>54</sup>. La structure des constituants que nous construisons peut être considéré comme une hiérarchie de boîtes. Dans le cas de l'anglais par exemple, la structure topologique ressemblerait à la structure syntagmatique traditionnelle, incluant un syntagme verbal composé d'un verbe et d'un objet<sup>55</sup>. Nous verrons par la suite que l'allemand présente des différences considérables entre les constituants syntaxiques et les syntagmes topologiques.

Passons maintenant aux règles proprement dites des phrases déclaratives de l'allemand. Ces règles ne prennent pas en compte la structure interne des noms. Les règles que nous présentons s'appliquent pour la linéarisation des verbes et de leurs dépendants.

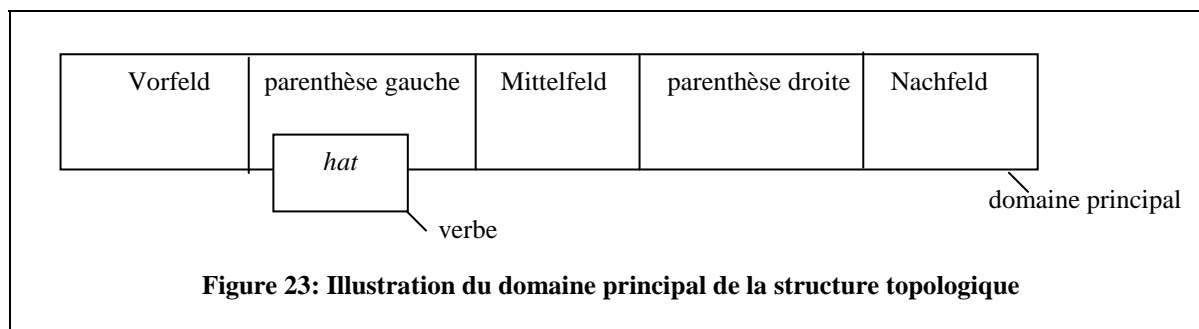
## Règles d'ordre pour l'allemand

- Tout nœud de l'arbre de dépendance syntaxique crée un constituant lexical qui est nommé à partir du nom de la partie de discours du mot auquel il renvoie.
- Les *règles de description de constituants* sont comme suit :
  - o le Vorfeld et la parenthèse gauche de la structure topologique finale doivent contenir exactement un constituant ;
  - o la parenthèse droite peut accueillir au plus un constituant
  - o Le nombre de constituants contenus dans le Mittelfeld et le Nachfeld n'est pas contraint.
- le verbe principal fini (appelé encore V2), tête de la dépendance syntaxique, crée un domaine principal qui comporte cinq champs : [*Vorfeld, parenthèse gauche, Mittelfeld, parenthèse droite, Nachfeld*]. Il se place dans la deuxième position du domaine principal, c'est-à-dire dans la parenthèse gauche (Figure 23).

---

<sup>54</sup> Cf. section IV.7 pour l'application du modèle à d'autres langues. Notons qu'il est possible que la description en termes de topologie soit possible sans pour autant être élégante ou « parlante ».

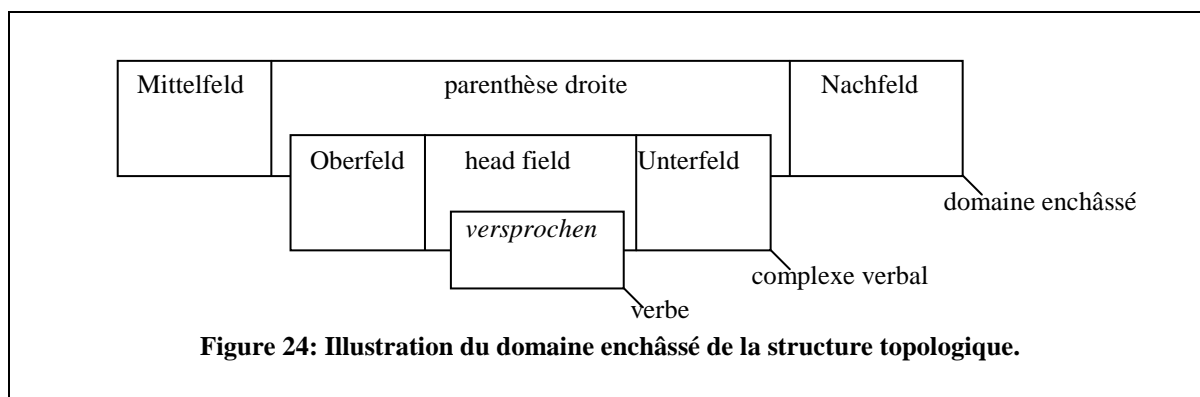
<sup>55</sup> Un autre débat que nous écartons ici concerne la légitimité du syntagme verbal dans les représentations syntaxiques de l'anglais et des autres langues (Cf. Abeillé 1996-1997 pour le français). Nous pensons qu'il s'agit plutôt d'une question interne au formalisme chomskyen qui n'a aucun appui théorique dans les grammaires de dépendance. Par ailleurs, nous pensons que l'hypothèse du syntagme verbal a largement été imprégnée par le rôle prédominant de l'anglais dans les recherches linguistiques.



- Tout dépendant verbal ouvre un constituant qu'on appelle complexe verbal. Ce constituant a trois champs, le *Oberfeld* qui constitue la place pour le constituant lexical, et le *Unterfeld*. Le champ *Oberfeld* et, sous certaines conditions le *Unterfeld*, sont une place possible pour les dépendants verbaux non finis.
- Le complexe verbal occupe toujours la parenthèse droite d'un domaine. Ensuite, deux possibilités se présentent :
  - o Le complexe verbal (qui est ouvert par le dépendant verbal du verbe principal) prend la place dans la parenthèse droite du domaine principal, prévue pour lui par son gouverneur syntaxique. Suivant Bech 1955, on appelle cette possibilité une *construction cohérente*.
  - o soit le dépend verbal ouvre à son tour un domaine et le complexe verbal se place dans la parenthèse droite de ce sous-domaine. Bech 1955 parle dans ce cas d'une *construction incohérente*.

Si un troisième verbe subséquent se joint au verbe qui se trouve déjà dans la parenthèse droite, ce dernier va à nouveau ouvrir une boîte avec une position à sa gauche. Ce procédé peut se répéter d'une manière récursive.

- Le domaine ouvert par un verbe non fini s'appelle *domaine enchâssé* (Figure 24). Il contient trois champs, le *Mittelfeld*, la *parenthèse droite* et le *Nachfeld*. Le complexe verbal se place dans la parenthèse droite.



- Les dépendants d'un verbe principal peuvent créer un sous-domaine. Le sous-domaine créé doit obligatoirement aller dans l'un des *champs majeurs* (le *Vorfeld*, le *Mittelfeld* ou le *Nachfeld*). Contrairement aux autres dépendants, un dépendant verbal peut choisir de ne pas ouvrir de sous-domaine. Le choix du champ principal où le sous-domaine est ouvert n'est pas libre : il obéit à des restrictions d'ordre syntaxique et communicatif. Ainsi :
  - o Les participes passés et les infinitifs nus créent le domaine le plus contraint, puisque ce dernier doit occuper le *Vorfeld* du domaine principal.
  - o Le cas des domaines créés par les verbes à l'infinitif *zu* est plus complexe : Le gouverneur qui régit un verbe à l'infinitif *zu* peut restreindre la possibilité de placement du domaine enchâssé de son dépendant. Il existe des verbes qui ne permet

ni l'extraposition ni l'intraposition de leur dépendant, donc le domaine enchâssé se comporte comme un domaine enchâssé de l'infinitif nu. On appelle de tels verbes matriciels des *verbes à construction cohérente obligatoire*. Les exemples incluent les verbes *brauchen* 'avoir besoin' et *pflügen* 'avoir l'habitude de'.

Il existe aussi des verbes, qui, dans le sens inverse, préfèrent des constructions incohérentes, mais comme le dit Bech 1955:84,85 lui-même, « il ne s'agit pourtant pas d'une loi absolue, mais d'une tendance, i.e. il s'agit d'une distribution statistique clairement en faveur de la construction incohérente.<sup>56</sup>

Pourtant, mis à part ces exceptions, on peut affirmer qu'un infinitif *zu* peut participer aux constructions cohérentes et incohérentes et au cas où il crée un domaine enchâssé, ce domaine peut aller dans le champ majeur de son choix.

- Le domaine des compléments phrastiques peut être ouvert soit dans le Vorfeld, soit dans le Nachfeld.
- Les domaines nominaux et prépositionnels peuvent aller dans le champ majeur de leur choix, incluant le Nachfeld où s'appliquent toutefois des contraintes de « lourdeur »<sup>57</sup>.
- Les particules (il s'agit de préfixes verbaux séparables, comme *an* du verbe *anfangen* 'commencer') se comportent exactement comme les verbes, et se placent donc dans la parenthèse droite (quand le gouverneur suit directement la particule, il est d'usage d'écrire ces deux éléments en un seul mot). Cette particule ouvre une position pour son dépendant verbal.
- Certains dépendants non verbaux comme les adjectifs prédicatifs et les noms gouvernés par une copule ou un verbe support peuvent aussi aller dans la parenthèse droite. Par contre, contrairement aux verbes et aux particules, ces éléments n'ouvrent généralement pas de nouvelle position pour leurs dépendants qui doivent trouver une place ailleurs.
- Un sous-domaine peut aller dans un champ majeur ou tout domaine occupé déjà par son le mot qui est gouverneur du domaine. Nous appelons *émancipation* le cas où ce sous-domaine ne se place pas dans le champ qui contient son gouverneur mais ailleurs. Ainsi, dans les phrases (105) *das Buch* 'le livre' est émancipé du domaine ouvert par son gouverneur *zu lesen* 'à lire' et se place dans un des champs majeurs du domaine ouvert par l'auxiliaire *hat* 'a'.

- (105) a. *Zu lesen hat das Buch diesem Mann niemand versprochen*  
 b. *Das Buch hat niemand diesem Mann versprochen zu lesen*  
 c. *Das Buch hat zu lesen niemand diesem Mann versprochen*

Dans la section 3 infra, nous étendrons cette grammaire à la structure interne des complexes verbaux, aux compléments phrastiques et aux propositions relatives.

---

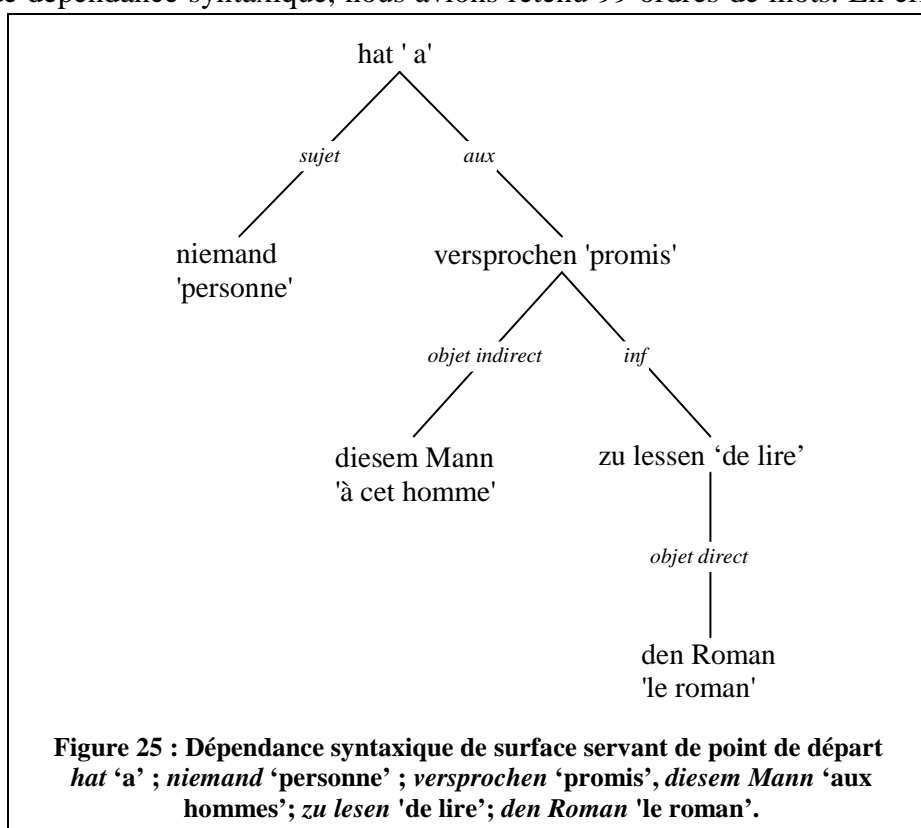
<sup>56</sup> texte original (des minuscules ont été laissées telles que Bech les utilisent) : « Es handelt sich aber nicht um ein absolutes gesetz, sondern um eine tendenz, d.h. um ein beträchtliches statistisches überwiegen der inkohärenzkonstruktion. »

<sup>57</sup> Il est peut-être utile de mentionner qu'on y rencontre aussi une restriction de type syntaxique, car les objets directs ont le plus du mal à se retrouver dans le Nachfeld. Notons que notre formalisme permet sans problème d'exclure de telles phrases sur la base de leur lien syntaxique, au cas où on voudrait décrire un ensemble plus restreint de phrases.

## 2 Différentes structures verbales de la topologie

Après avoir posé les bases de la structure topologique dans la section précédente, nous présentons ici les différentes structures topologiques qu'on peut avoir pour une structure de dépendance syntaxique donnée. Nous reprenons l'exemple présenté dans la section I.6 (nous redonnons la structure de dépendance syntaxique de surface dans la Figure 25) et nous analysons les différents phénomènes que nous avons présenté à l'aide de la structure topologique.

L'arbre de dépendance syntaxique de surface ci-dessus a permis d'obtenir 144 structures de constituants topologiques<sup>58</sup>. Ce chiffre monte à 188 si on inclut dans les phrases le pronom *es* (du Vorfeld) et à 666 si on permet l'extraposition nominale dans le Nachfeld. Notons que pour cet arbre de dépendance syntaxique, nous avons retenu 99 ordres de mots. En effet, certains ordres



des mots peuvent avoir plusieurs analyses topologiques, comme nous le verrons par la suite. Les structures de constituants topologiques que nous obtenons peuvent être classifiés selon les domaines qui les composent. Indépendamment de la position linéaire finale de ces constituants, plusieurs cas de figure se présentent :

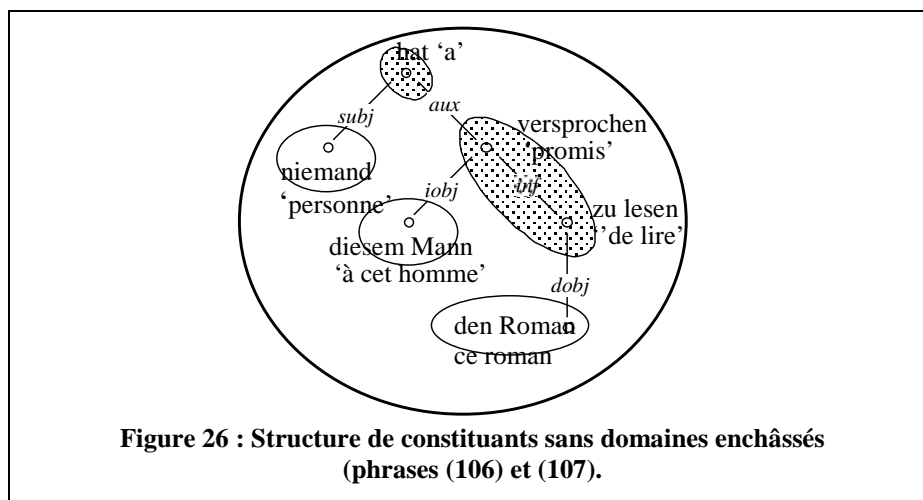
- des structures qui n'ont pas de domaines enchâssés
- des structures où certains verbes créent un domaine enchâssé
- des structures où les domaines enchâssés ne contiennent aucun des dépendants du verbe qui est le gouverneur de ce domaine.

Nous présentons ces cas de figure un par un, en commençant par le plus simple, les structures dites « plates ».

<sup>58</sup> Ces structures ont été obtenues grâce à un calcul effectué par DepLin (*dependency linearizer*), un logiciel que nous avons développé pour calculer les différentes linéarisations pour une structure de dépendance syntaxique. Nous présentons le logiciel dans la section IV.1.3. Voir également la page web [talana.linguist.jussieu.fr/~kim/deplin](http://talana.linguist.jussieu.fr/~kim/deplin).

## 2.1 Les structures plates

La structure la moins marquée des structures topologiques de l'allemand est celle qui ne comporte pas de sous-domaine. Dans une telle structure, les dépendants verbaux *versprochen* 'promis' et *zu lesen* 'de lire' se placent tous les deux dans le domaine principal du verbe gouverneur *hat* 'a'. Comme aucun sous-domaine n'est créé, tous les dépendants non verbaux des trois verbes se partagent les champs principaux : le placement de ces dépendants non verbaux se réalise indépendamment de la hiérarchie qui existe entre les verbes mais selon des contraintes d'ordre communicatif. Nous pouvons représenter une projection des domaines sur l'arbre de dépendance syntaxique comme dans la Figure 26. Dans cette figure, nous avons superposé l'arbre de dépendance aux groupes topologiques. Les constituants qui occupent les parenthèses gauche et droite sont représentés par des ovales ombrés.



Nous obtenons donc des linéarisations des phrases (106) et (107). Nous donnons l'analyse topologique de ces deux phrases dans la Figure 27.

(106) Niemand hat diesem Mann den Roman zu lesen versprochen.

(107) Diesem Man hat den Roman niemand zu lesen versprochen.



	<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>
(106)	Niemand	hat	diesem Mann den Roman	zu lesen	versprochen
	personne (nom.)	a	à cet homme (dat.) le roman (acc.)	de lire	promis

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>
(107)	diesem Mann	hat	den Roman Niemand	zu lesen	versprochen
	à cet homme (dat.)	a	le roman (acc.) personne (nom.)	de lire	promis

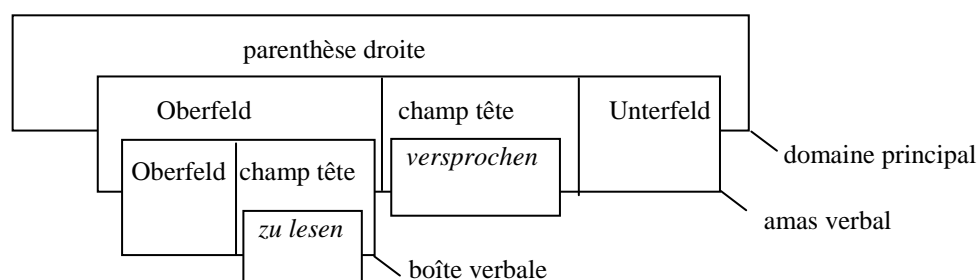
Figure 27 : Structures topologiques sans domaines enchâssés (respectivement phrases (106) et (107))<sup>59</sup>.

Dans une structure topologique plate, les trois syntagmes non verbaux *niemand* ‘personne’, *diesem Mann* ‘à cet homme’ et *den Roman* ‘ce roman’ se trouvent au même niveau : l’un des trois syntagmes doit aller dans le *Vorfeld*, les deux autres dans le *Mittelfeld*. Six ordres de mots sont possibles pour cette structure topologique, dont les phrases (106) et (107). Notons toutefois que nous ne prenons pas en considération certaines restrictions générales sur l’ordre interne du *Mittelfeld* (voir par exemple Lennerz 1977 et Uszkoreit 1987). En fait, l’ordre des constituants dans le *Mittelfeld* dépend très peu de leur position hiérarchique dans la structure syntaxique. Il est plutôt conditionné par des phénomènes comme la pronominalisation, la focalisation, le poids des constituants etc. Le néerlandais, par exemple, a la même structure topologique de base que l’allemand mais son champ *Mittelfeld* est plutôt conditionné syntaxiquement. Nous pensons que cela est lié au fait que le néerlandais a perdu beaucoup de ses marques casuelles (à l’exception des pronoms) : par conséquent, sur chaque complément qu’on le place dans le *Mittelfeld*, on doit indiquer sa hauteur hiérarchique dans l’arbre de dépendance syntaxique, et faire en sorte que les constituants soient placés de manière descendante.

## 2.2 Analyse du « brouillage d’arguments »

Le fait qu’une projection verbale (par exemple le verbe et tous ses dépendants) ne constitue pas un syntagme continu est connu sous le terme de brouillage d’arguments (voir section I.6). Cependant, la notion de « brouillage d’argument » est basée sur une conception syntaxique qui suppose que l’ordre des mots est guidé par la hiérarchie syntaxique et que toute analyse qui dévie de cette hiérarchie doit être considérée comme problématique. Ceci présuppose qu’il existe un ordre standard qui reflète la hiérarchie syntaxique. En fait, nous pensons que former un sous-domaine pour le verbe et chacun de ses constituants n’a pas de sens. Au contraire, nous considérons que tous les verbes sont placés dans le même domaine avec leurs dépendants. Dans

<sup>59</sup> La structure de la parenthèse droite que nous ne détaillons pas dans la figure s’organise de la manière suivante :



notre approche, il n'y a aucun avantage descriptif à utiliser le terme de *brouillage* puisqu'on aboutit à différentes structures, mais rien n'est « brouillé ». Nous renvoyons le lecteur à la discussion du chapitre I en ce qui concerne la notion d'ordre standard.

L'exemple de brouillage d'arguments que nous avons donné dans la section 6 du premier chapitre (et que nous redonnons sous 0) a la structure topologique que nous représentons dans la Figure 29. Nous pouvons voir que la base de la structure topologique est toujours la même et que la notion de brouillage d'arguments (scrambling) renvoie juste à un placement de l'argument *den Roman* 'le roman' dans un autre champ que celui de son gouverneur verbal *zu lesen* 'de lire'.

(108) Niemand hat den Roman diesem Mann zu lesen versprochen.

	Vorfeld	(	Mittelfeld	)	Nachfeld
0	Niemand	hat	den Roman diesem Mann	zu lesen versprochen	
	Personne (nom.)	a	le roman (acc.) à cet homme (dat.)	de lire promis	

Figure 28 : Structure topologique de la phrase 0 illustrant le brouillage d'argument.

### 2.3 L'enchâssement

Comme nous l'avons indiqué dans les règles de base, un verbe peut ouvrir un domaine enchâssé dans l'un des champs majeurs. Par exemple, le verbe fini *versprochen* 'promis' ou l'infinitif en « zu » *zu lesen* 'de lire' (ou les deux) peut ouvrir un domaine enchâssé. C'est le cas illustré dans la Figure 29. Les domaines sont représentés par un ovale avec une bordure en gras.

L'auxiliaire *hat* 'a' et le verbe fini *versprochen* 'promis' occupent respectivement la parenthèse gauche et droite du domaine principal. On trouve trois constituants au même niveau : *niemand* 'personne', *diesem Mann* 'à cet homme' et le constituant enchâssé *den Roman zu lesen* 'de lire ce roman'. L'ordre des ces trois constituants est libre, la seule restriction étant que l'un des trois doit prendre la place du Vorfeld (selon les règles de bonnes formations de la structure topologique). Le domaine enchâssé peut aller dans le Vorfeld ((109)b.), le Nachfeld ((109)c.) ou le Mittelfeld (((109)a. et d.).

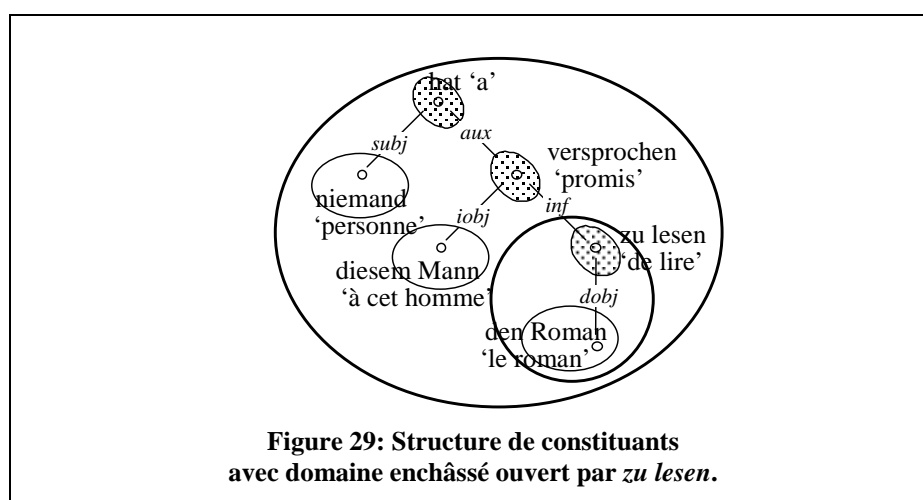


Figure 29: Structure de constituants avec domaine enchâssé ouvert par *zu lesen*.

- (109) a. Niemand hat diesem Mann den Roman zu lesen versprochen.
- b. Den Roman zu lesen hat diesem Mann niemand versprochen.
- c. Diesem Mann hat niemand versprochen, den Roman zu lesen.
- d. Diesem Mann hat, den Roman zu lesen, niemand versprochen.

Dans la Figure 30, nous donnons les structures topologiques des phrases en (109). Ces phrases illustrent les phénomènes d'extraposition complète ((109)c.), de VP fronting complet ((109)b.) et d'intraposition complète ((109) a. et d.) que nous avons discutés dans le chapitre précédent.

Les locuteurs natifs de l'allemand considèrent les phrases (107) et (109)c. (le premier sans enchâssement, le deuxième avec enchâssement) comment ayant la même complexité syntaxique. Le fait qu'il y ait un domaine enchâssé ou non dépend de la structure communicative. Une entité communicative (comme le thème, le rhème etc.) entraîne le verbe à ouvrir un domaine et créer un constituant séparé qui reçoit le contour prosodique correspondant à l'entité communicative (voir section I.6.3).

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>				)	<i>Nachfeld</i>
(109)a	niemand	hat	diesem Mann	<i>MF</i>	)	<i>NF</i>	versprochen	
	personne (nom.)	a	cet homme (dat.)	den Roman	zu lesen			
				le roman (acc.)	de lire		promis	

	<i>Vorfeld</i>		(	<i>Mittelfeld</i>			)	<i>Nachfeld</i>
(109)b	<i>MF</i>	)	<i>NF</i>	hat	diesem Mann niemand		versprochen.	
	Den Roman	zu lesen						
	le roman (acc.)	de lire		a	à cet homme (dat.) personne (nom.)		promis	

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>			)	<i>Nachfeld</i>	
(109)c	Diesem Mann	hat	niemand	versprochen,		<i>MF</i>	)	<i>NF</i>
	à cet homme (dat.)	a	personne (nom.)	promis,		den Roman	zu lesen	
						le roman (acc.)	de lire	

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>				)	<i>Nachfeld</i>
(109)d	diesem Mann	hat	<i>MF</i>	)	<i>NF</i>	niemand	versprochen	
	Personne (nom.)	a	den Roman	zu lesen				
			le roman (acc.)	de lire		personne (nom.)	promis	

Figure 30: Structure topologique des phrases (109) illustrant le brouillage d'argument.

Notons que la création d'un domaine enchâssé dans le *Mittelfeld* (par exemple l'intraposition dans (109)d.), est un phénomène considéré comme plus complexe que la création d'un domaine enchâssé dans le *Vorfeld* ou le *Nachfeld*. La raison est encore une fois communicative : le *Vorfeld* et le *Nachfeld* sont communicativement plus marqués que le *Mittelfeld*, en ce sens que le *Vorfeld* est généralement occupé par le thème ou un rhème proéminent et le *Nachfeld* par un thème non proéminent, tandis que le *Mittelfeld* n'a pas de vraie caractérisation communicative.

En d'autres termes, il y a peu de raisons communicatives d'ouvrir un domaine enchâssé pour le placer dans le Mittelfeld.

Si on reprend le phénomène d'intraposiion, ce dernier a lieu quand il y a une certaine coïncidence entre différents impératifs communicatifs. Ainsi, si un élément est marqué comme un topique et que la place du Vorfeld ne lui est pas accessible, alors cet élément a de fortes chances d'être intraposé. Une intraposiion peut donc se produire dans deux circonstances. D'abord, la topicalisation peut avoir lieu à l'intérieur d'une complétive : comme le domaine où se trouve la complète ne possède pas de Vorfeld et le Vorfeld du domaine principal n'est pas accessible pour un élément faisant partie de la complétive (seul les mots qu- peuvent sortir de ce domaine), une intraposiion à lieu. Nous illustrons ce cas avec un exemple repris de Kathol 1995:44 :

(110) Ich glaube, dass [dieses Buch zu lesen] der Professor den Studenten empfohlen hat.

Je pense, que [ ce livre de lire] le professeur au étudiants conseillé a  
'De lire ce livre, c'est ce que je pense que le professeur à conseillé aux étudiants.'

Ensuite, l'intraposiion peut avoir lieu quand le Vorfeld est déjà occupé par un thème plus proéminent que l'élément intraposé. Ce thème « second » est donc placé dans la première place du Mittelfeld comme dans notre exemple (109)d., qu'on peut traduire de la manière suivante : A cet homme, de lire ce livre, PERSONNE ne le lui a promis<sup>60</sup>.

Dans la figure suivante, nous donnons la représentation complète de la structure topologique de la phrase (109)b.

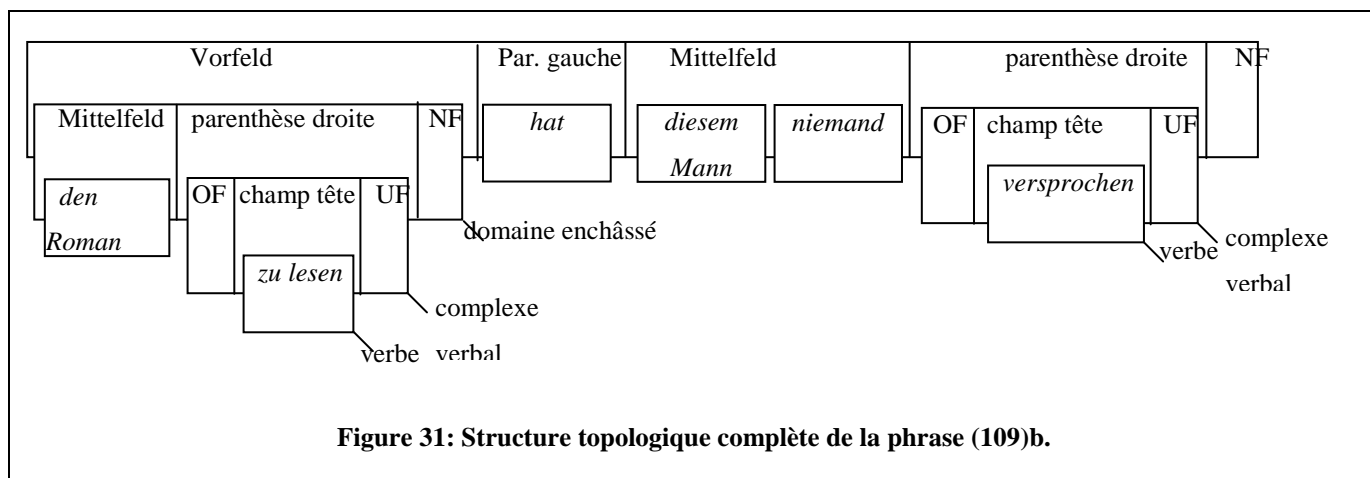


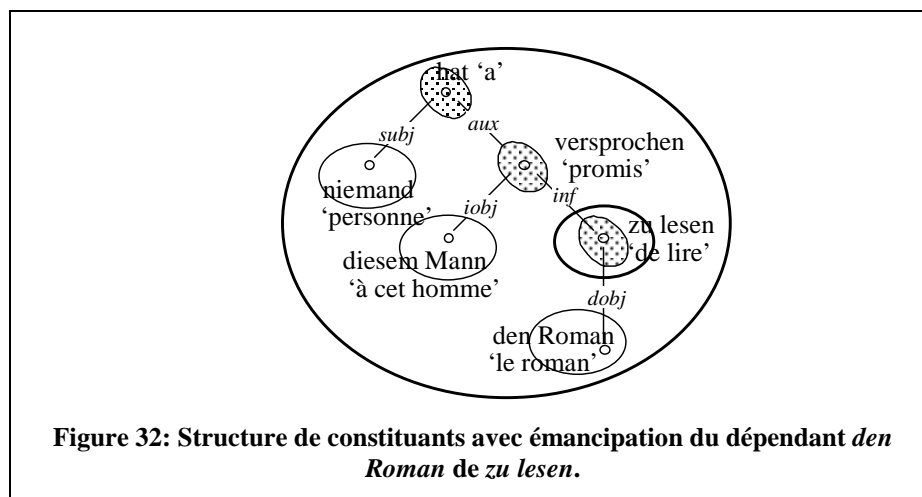
Figure 31: Structure topologique complète de la phrase (109)b.

<sup>60</sup> Les majuscules indiquent un marquage prosodique de focus.

## 2.4 L'émancipation

Comme nous l'avons évoqué lors de la description des règles de linéarisation, les dépendants d'un verbe ne sont pas placés obligatoirement dans le domaine de leur gouverneur, et peuvent être émancipés dans un domaine supérieur au domaine de leur gouverneur<sup>61</sup>. La Figure 32 représente un exemple de dépendance syntaxique avec émancipation.

Dans cette figure, le verbe *zu lesen* 'de lire' crée un domaine enchâssé duquel son dépendant *den Roman* 'le roman' a été émancipé. Par conséquent, quatre compléments doivent être placés dans



le domaine supérieur, permettant ainsi trente ordres de mots possibles. Les phrases (111) et (112) représentent deux exemples parmi ces trente possibilités.

(111) Zu lesen hat den Roman diesem Mann niemand versprochen.

(112) Den Roman hat niemand diesem Mann versprochen zu lesen.

Parmi les trente ordres de mots obtenus, seuls ceux qui ont *den Roman* ou *zu lesen* dans la position du Vorfeld sont facilement acceptables ; autrement dit, seules les phrases où il est facile de trouver un contexte communicatif simple semblent correctes quand elles sont présentées sans un contexte particulier. un tel contexte communicatif simple peut être que *den Roman* ou *zu lesen* sont topicalisés ou focalisés.

D'un point de vue computationnel (quel que soit le type de calcul considéré, qu'elle soit faite par une machine ou le cerveau humain), l'émancipation rend plus complexe la correspondance entre l'arbre de dépendance syntaxique de surface et la structure topologique. C'est la principale raison pour laquelle une émancipation doit être communicativement bien motivée. Dans notre exemple de référence, *den Roman* 'le livre' peut être émancipé du domaine ouvert par son gouverneur *zu lesen* 'de lire' seulement si ce gouverneur peut former une entité communicative sans son dépendant. Ainsi, la phrase (111) est considérée comme appropriée si *zu lesen* est mis en contraste avec un autre verbe, comme nous le montrons dans l'exemple (113) :

(113) Zu lesen hat den Roman diesem Mann niemand versprochen, wohl aber zu übersetzen.

de lire    a    ce roman    à cet homme personne promis,    bien mais de traduire

<sup>61</sup> Dans le cas où le verbe enchâssé n'ouvre pas de domaine mais construit un complexe verbal dans la parenthèse droite comme c'est le cas des structures plates que nous avons décrites infra, ses dépendants nominaux doivent être placés dans un domaine qui n'a pas été ouvert par leur gouverneur. Néanmoins, nous considérons qu'il ne s'agit pas ici de cas d'émancipation. On opte plutôt pour une analyse où on considère qu'il y a eu un accord entre les verbes de partager un même domaine, et donc on considère que les dépendants sont placés dans le domaine de leur gouverneur.

‘Personne n’a promis de LIRE ce roman à cet homme, mais juste de le TRADUIRE.’

Dans la Figure 33, nous présentons les structures de constituants topologiques des phrases (111) et (112) : par rapport à la discussion présentée dans le chapitre précédent, la phrase (111) illustre un cas de VP fronting partiel tandis que la phrase (112) est un cas d’extraction lointaine (ou encore d’émancipation) de l’objet :

<i>Vorfeld</i>				<i>Mittelfeld</i>				<i>Nachfeld</i>
<i>MF</i>	)	<i>NF</i>						
	zu lesen de lire		hat	den Roman diesem Mann niemand				versprochen.
(111)			a	le roman (acc.) à cet homme (dat.) personne (nom.)				promis

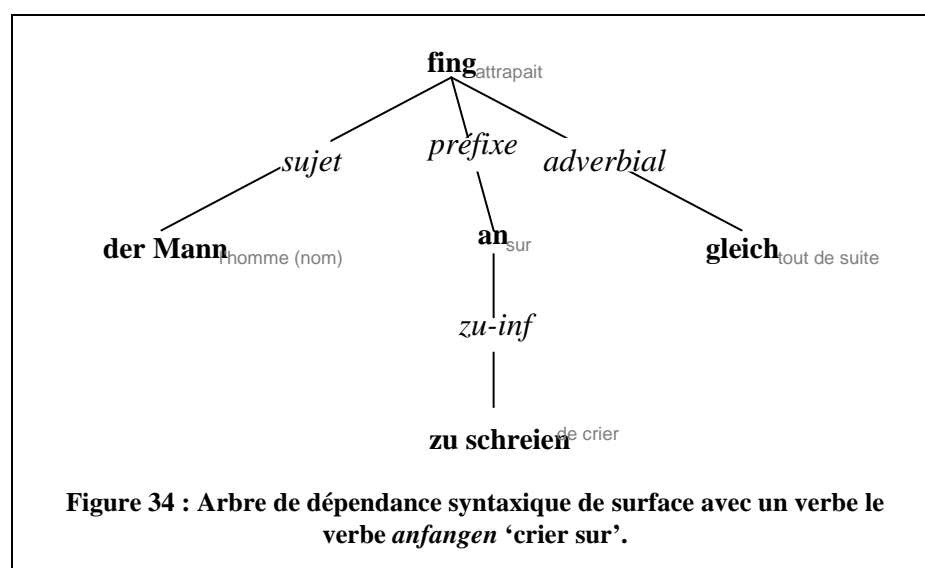
<i>Vorfeld</i>				<i>Mittelfeld</i>			<i>Nachfeld</i>	
	)							
	hat			niemand diesem Mann				versprochen,
(112)			a	personne (nom.) à cet homme (dat.)				promis,
								zu lesen de lire

**Figure 33: Structures topologiques avec émancipation (la phrase (111) correspond à un cas de VP fronting partiel et la phrase (112) à un cas d’émancipation de l’objet).**

## 2.5 Le traitement des préfixes

Nous finissons cette section avec la structure topologique des préfixes qui se comportent exactement de la même manière que les dépendants des verbes et prennent place dans les parenthèses. Si un verbe dépend d’un verbe contenant un préfixe comme le verbe *anfangen*, alors ce dépendant verbal se comporte comme si son gouverneur était le préfixe.

Prenons l’arbre de dépendance syntaxique de surface que nous reproduisons dans la Figure 34 :



Le verbe *anfangen* consiste en une partie verbale *fangen* (‘attraper’) et en une partie prépositionnelle *an* (‘sur’).

Dans le cas où la partie verbale prend la parenthèse gauche, la partie prépositionnelle prend la parenthèse droite. Les conventions d'écriture font que dans le cas où la partie verbale prend la parenthèse droite et la partie prépositionnelle se place comme un dépendant verbal devant son gouverneur dans la parenthèse droite, on écrit les deux parties en un seul mot, même si le verbe est préfixé par la forme du participe passé *ge* ou de l'infinitif *zu* : on obtient ainsi des formes du type *angefangen* et *anzufangen*. Dans la Figure 35, nous montrons les différentes structures de constituants topologiques que l'on peut obtenir pour une préposition et ses dépendants. Comme dans le cas du verbe, la préposition peut ouvrir un domaine enchâssé dans l'un des champs majeurs, le *Vorfeld*, le *Nachfeld* ou le *Mittelfeld*.

### linéarisation plate

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>
(a)	der Mann l'homme (nom.)	fing attrapait	gleich tout de suite	zu schreien an de crier sur	

### Domaine enchâssé dans le *Nachfeld*

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>	
					<i>MF</i>	<i>NF</i>
(b)	der Mann l'homme (nom.)	fing attrapait	gleich tout de suite	an, sur,	zu schreien de crier	

### Domaine enchâssé dans le *Vorfeld*

	<i>Vorfeld</i>			(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>
	<i>MF</i>	)	<i>NF</i>				
(c)	zu schreien de crier			Fing Attrapait	der Mann gleich l'homme (nom.) tout de suite	an sur	

### Domaine enchâssé dans le *Mittelfeld*

	<i>Vorfeld</i>	(	<i>Mittelfeld</i>			)	<i>Nachfeld</i>
			<i>MF</i>	)	<i>NF</i>		
(d)	der Mann l'homme (nom.)	fing attrapait	zu schreien de crier			gleich tout de suite	an sur

Figure 35 : Structures topologiques indiquant le placement du préfixe *an*.

## 3 Structure du complexe verbal

Dans la section précédente, nous avons revisité les phénomènes d'ordre de mots présentés dans la section 6 du premier chapitre et nous avons essayé de montrer comment la structure topologique et les règles de linéarisation de l'allemand permettaient d'obtenir une analyse simple et élégante des différents ordres de mots. Dans cette présente section, nous approfondissons la grammaire de base de l'allemand en étudiant plus spécifiquement la structure du complexe verbal.

### 3.1 Description des possibilités d'ordre dans un complexe verbal

Dans un complexe verbal, le dépendant se place généralement à gauche de son gouverneur, formant ainsi des ordres des verbes tels que  $V_1$ ,  $V_2V_1$ ,  $V_3V_2V_1$ ,  $V_4V_3V_2V_1$  etc<sup>62</sup>. Cependant,

<sup>62</sup> Les indices renvoient à la position qu'ont les verbes dans une chaîne subordonative, donc dans une chaîne de verbe où un verbe d'index  $i$  prend régit le verbe d'index  $i+1$ .

l'ordre inverse est aussi possible. Ainsi, les auxiliaires de temps HABEN 'avoir' pour le passé et WERDEN 'devenir' pour le futur permet à ses dépendants de se placer à droite du complexe verbal. Ce phénomène, que nous illustrons dans les exemples en (114), est connu sous le nom de *permutation d'auxiliaire* ou encore *Oberfeldumstellung* (Bech 1955)<sup>63</sup>.

- (114) a. Er wird das Buch haben<sub>1</sub> lesen<sub>3</sub> können<sub>2</sub>  
 il va ce livre avoir<sub>1</sub> lire<sub>3</sub>pouvoir<sub>2</sub>  
 'il aura pu lire le livre.'
- b. ... weil er das Auto hätte<sub>1</sub> zu fahren<sub>4</sub> versuchen<sub>3</sub> können<sub>2</sub>.  
 ... car il le voiture a<sub>1</sub> de conduire<sub>4</sub> essayer<sub>3</sub> pouvoir<sub>2</sub>  
 '... parce qu'il aurait pu essayer de conduire la voiture.'
- c. ... weil er das Auto hätte<sub>1</sub> kaputt<sub>4</sub>fahren<sub>3</sub> können<sub>2</sub>.<sup>64</sup>  
 ... car il le voiture a<sub>1</sub> casser<sub>4</sub>conduire<sub>3</sub> pouvoir<sub>2</sub>  
 '... parce qu'il aurait pu complètement bousiller la voiture.'
- d. Ich glaube, dass er das Buch wird<sub>1</sub> haben<sub>2</sub> lesen<sub>4</sub> können<sub>3</sub>.  
 Je pense, que il ce livre va<sub>1</sub> avoir<sub>2</sub> lire<sub>4</sub> pouvoir<sub>3</sub>  
 'Je pense qu'il aura pu lire le livre.'

Dans les phrases (114) a, b et c, le dépendant V<sub>3</sub> du verbe V<sub>2</sub>, placé à droite de l'auxiliaire V<sub>1</sub>, se place ) gauche de son gouverneur V<sub>2</sub>, en suivant l'ordre standard. On obtient donc les ordres V<sub>1</sub>V<sub>2</sub>, V<sub>1</sub>V<sub>3</sub>V<sub>2</sub>, V<sub>1</sub>V<sub>4</sub>V<sub>3</sub>V<sub>2</sub> etc.<sup>65</sup>. Dans le cas de l'exemple (114) d., c'est une particule et non un verbe plein qui joue le rôle du quatrième verbe de la chaîne subordonnée. Les possibilités d'ordres qui en résultent sont identiques à celles obtenues pour les compléments verbaux.

Les verbes V<sub>2</sub> gouvernés qui peuvent accepter l'ordre inverse peuvent être regroupés dans une classe fermée de verbes qui inclut les modaux, les verbes de perception et quelques autres verbes, comme HELFEN 'aider', le verbe causatif LASSEN 'laisser' etc. Cette classe comporte également l'auxiliaire HABEN 'avoir', ce qui explique le cas de double permutation d'auxiliaire comme dans la phrase (114)c, permettant par conséquent les ordres V<sub>1</sub>V<sub>2</sub>V<sub>3</sub>, V<sub>1</sub>V<sub>2</sub>V<sub>4</sub>V<sub>3</sub> etc.

Le dépendant V<sub>3</sub> d'un V<sub>2</sub> peut également se placer à gauche de l'auxiliaire V<sub>1</sub>, et on obtient ainsi les ordres V<sub>3</sub>V<sub>2</sub>V<sub>1</sub>, V<sub>4</sub>V<sub>3</sub>V<sub>1</sub>V<sub>2</sub>, etc. Il s'agit là d'une variante de la permutation d'auxiliaire, appelée *Zwischenstellung* (allemand, pour « position immédiate ») ou encore *scission du complexe verbal* (voir entre autres Meurers 1999). Nous illustrons ce cas avec les exemples en (115), qui ne sont acceptés que par certains locuteurs de l'allemand. Notons que le *Zwischenstellung* peut également concerner au lieu des V<sub>3</sub>, les V<sub>4</sub> qui sont placés à gauche de l'auxiliaire du complexe verbal, permettant ainsi l'ordre V<sub>4</sub>V<sub>1</sub>V<sub>3</sub>V<sub>2</sub> comme dans (115)c. Le

<sup>63</sup> On trouve également les appellations *Verb Raising* et *Double-Infinitive Construction*.

<sup>64</sup> Nous empruntons cet exemple à Ferret 2002.

<sup>65</sup> On considère souvent que le complexe verbal V<sub>1</sub>V<sub>2</sub> est impossible et l'argumentation repose sur le fait que la permutation d'auxiliaire ne peut se faire que s'il y a au moins trois verbes. En effet, la permutation d'auxiliaire n'est possible que si le dépendant verbal V<sub>2</sub> appartient à la classe des verbes modaux qui gouvernent un verbe à l'infinitif V<sub>3</sub>. Cependant, V<sub>3</sub> ne fait pas nécessairement partie du complexe verbal et une phrase comme (i) n'est pas problématique.

(i) Das Buch lesen<sub>3</sub> wird<sub>0</sub> er bis morgen wohl haben<sub>1</sub> können<sub>2</sub>  
 le livre lire<sub>3</sub> car<sub>0</sub> il jusqu'à demain bien avoir<sub>1</sub> pouvoir<sub>2</sub>  
 'Il aura pu lire le livre jusqu'à demain.'

La construction elliptique sans dépendant apparent de *können*, offre une autre possibilité, comme on peut le voir dans l'exemple proposé par Hinrichs and Nakazawa 1994 :

(ii) ... weil er nicht anders hat<sub>1</sub> können<sub>2</sub>  
 ... car il pas différemment a<sub>1</sub> pouvoir<sub>2</sub>  
 '... parce qu'il ne pouvait (le faire) différemment.'



*Zwischenstellung* est également compatible avec la double permutation, où le verbe  $V_4$  peut aller à gauche du premier ou du deuxième auxiliaire, permettant les ordres  $V_4V_1V_2V_3$  et  $V_1V_4V_2V_3$  comme dans les exemples (115)d. et (115)e.

- (115) a. ?Ich glaube, dass er das Buch lesen<sub>3</sub> wird<sub>1</sub> können<sub>2</sub>  
 je pense, que il le livre lire<sub>3</sub> va<sub>1</sub> pouvoir<sub>2</sub>  
 'Je pense qu'il pourra lire le livre.'
- b. ?... weil ich sonst drei Stunden warten<sub>3</sub> hätte<sub>1</sub> müssen<sub>2</sub>.<sup>66</sup>  
 ... car je sinon trois heures attendre<sub>3</sub> eu<sub>1</sub> devoir<sub>2</sub>  
 '... parce que sinon, j'aurais dû attendre trois heures.'
- c. ?... weil er das Auto kaputt<sub>4</sub> hätte<sub>1</sub> fahren<sub>3</sub> können<sub>2</sub>  
 ... car il le voiture casser<sub>4</sub> eu<sub>1</sub> conduire<sub>3</sub> pouvoir<sub>2</sub>  
 '... parce qu'il aurait pu complètement bousiller la voiture.'
- d. ?Ich glaube, dass er das Buch lesen<sub>4</sub> wird<sub>1</sub> haben<sub>2</sub> können<sub>3</sub>  
 je pense, que il le livre lire<sub>4</sub> va<sub>1</sub> avoir<sub>2</sub> pouvoir<sub>3</sub>  
 'Je pense qu'il aura pu lire le livre.'
- e. ?Ich glaube, dass er das Buch wird<sub>1</sub> lesen<sub>4</sub> haben<sub>2</sub> können<sub>3</sub>  
 je pense, que il le livre va<sub>1</sub> lire<sub>4</sub> avoir<sub>2</sub> pouvoir<sub>3</sub>  
 'Je pense qu'il aura pu lire le livre.'

Notons que le dépendant de *haben* 'avoir' doit être dans sa forme d'infinitif nu et non dans celle de participe passé lorsque la permutation d'auxiliaire a lieu. Cette forme, qu'on appelle *Ersatzinfinitiv*, est également possible et même préférable pour certains verbes lorsque l'auxiliaire se trouve dans la position V2. Nous illustrons ce cas avec l'exemple (116).

- (116) Er hat das Buch lesen können / ?gekonnt  
 il a le livre lire pouvoir/ ?pu  
 'Il a pu lire le livre.'

### 3.2 Modélisation du complexe verbal

Nous pouvons modéliser l'ordre des mots du complexe verbal en suivant les mêmes principes que pour la modélisation de la phrase. Nous considérons que le complexe verbal constitue un constituant topologique avec trois champs : l'*Oberfeld*, le *champ tête* et l'*Untersfeld*. Le champ tête est occupé par le gouverneur du complexe verbal (par exemple le verbe qui se place dans la parenthèse droite et crée ce complexe verbal, ou encore qui se place dans le *Untersfeld* dans le cas d'une permutation d'auxiliaire). Tout dépendant verbal du gouverneur du complexe verbal peut aller dans le *Oberfeld* : dans ce cas, il ouvre un constituant topologique qu'on appelle *boîte verbale*. Contrairement au complexe verbal, une boîte verbale comporte uniquement deux champs : un *Oberfeld* et un *champ tête*. En fait, cette distinction entre complexe verbal et boîte verbale est nécessaire afin d'assurer le blocage de la permutation d'auxiliaire à un auxiliaire qui se trouve dans l'*Oberfeld* de la parenthèse droite. C'est la raison pour laquelle les phrases (117) sont agrammaticales. En fait, une fois que le verbe a été placé dans l'*Oberfeld*, il n'est plus possible pour un autre verbe d'effectuer la permutation d'auxiliaire.

- (117) a. \*Ich glaube, dass Peter das Buch haben<sub>2</sub> lesen<sub>4</sub> können<sub>3</sub> wird<sub>1</sub>  
 je pense, que Pierre le livre avoir<sub>2</sub> lire<sub>4</sub> pouvoir<sub>3</sub> va<sub>1</sub>

<sup>66</sup> Phrase originale produite par Leo Wanner, à Paris, Avril 2002.

b. \*Ich glaube, dass Peter das Buch haben<sub>2</sub> wird<sub>1</sub> lesen<sub>4</sub> können<sub>3</sub>

je pense, que Pierre le livre avoir<sub>2</sub> va<sub>1</sub> lire<sub>4</sub> pouvoir<sub>3</sub>  
 'Je pense que Pierre aura à lire livre.'

Dans le cas du *Zwischenstellung*, la modélisation est faite en permettant à un verbe qui veut se placer dans un Oberfeld de traverser les boîtes verbales et les complexes verbaux afin d'atterrir dans le Oberfeld d'un verbe enchâssé qui a été ouvert par un auxiliaire et qui est resté vacant parce que le dépendant de cet auxiliaire a été placé dans l'Unterfeld. Comme nous pouvons le voir, la description que fait notre grammaire de ce phénomène considéré comme complexe est très naturelle et facile à comprendre. La possibilité du *Zwischenstellung* dépend de la possibilité ou non pour un verbe de se placer non seulement au Oberfeld de son gouverneur, mais aux Oberfeld de tout le complexe verbal. Ceci peut également expliquer la variation d'acceptabilité qui existe chez les locuteurs d'allemand.

Nous considérons que le néerlandais ou l'allemand suisse ont la même structure topologique que l'allemand. Dans ces langues, l'ordre standard à l'intérieur de la parenthèse droite est similaire au Oberfeldumstellung allemand en ce qu'un verbe infinitif nu tend à être placé à droite de son gouverneur<sup>67</sup>. L'ordre qui en résulte implique des *dépendances sérielles croisées* (de l'anglais *cross serial dependencies*, voir Evers 1975 et Bresnan *et al.* 1982). En fait, l'ordre du Mittelfeld du néerlandais est syntaxiquement contraint et les constituants doivent apparaître dans leur ordre hiérarchique : ainsi, un constituant C<sub>1</sub> qui dépend d'un verbe V<sub>1</sub> doit apparaître avant le constituant C<sub>2</sub> qui dépend de V<sub>2</sub> et ainsi de suite. A cause de l'ordre interne de la parenthèse droite, nous obtenons des ordres comme C<sub>1</sub>C<sub>2</sub>... C<sub>n</sub>V<sub>1</sub>V<sub>2</sub>...V<sub>n</sub>, comme dans l'exemple ci-dessous que nous empruntons à Bresnan *et al.* 1982 :

(118) ...dat Jan<sub>1</sub> Piet<sub>2</sub> Marie<sub>3</sub> de kinderen<sub>4</sub> zag<sub>1</sub> helpen<sub>2</sub> laten<sub>3</sub> zwemmen<sub>4</sub>  
 ...que Jan<sub>1</sub> Piet<sub>2</sub> Marie<sub>3</sub> les enfants<sub>4</sub> vu<sub>1</sub> aider<sub>2</sub> laisser<sub>3</sub> nager<sub>4</sub>  
 '...que Jan a vu Piet aider Marie à faire nager les enfants.'

Les constructions de *dépendances sérielles croisées* ont beaucoup été étudiées à cause de leur complexité supposée. Notre grammaire de l'allemand peut être facilement adaptée au néerlandais (si on laisse de côté la description de l'ordre dans le Mittelfeld) en gardant exactement les mêmes constituants topologiques. Ainsi un verbe placé dans la parenthèse droite ou dans un Unterfeld ouvre un complexe verbal avec les trois mêmes champs. Par contre, contrairement à l'allemand, la règle qui permet à un verbe de se placer dans un Oberfeld est contraignante (il semblerait que seuls les participes passés peuvent aller dans ce champ) tandis que la règle qui permet à un verbe de se placer dans un Unterfeld est plus permmissible : en effet, le Oberfeldumstellung semble s'appliquer à toute la classe des verbes qui gouvernent un infinitif nu (contrairement à l'allemand où cette règle est restreinte aux auxiliaires). Notons qu'une particule se place également dans l'Oberfeld. Par ailleurs, comme pour l'allemand, certains locuteurs natifs du néerlandais acceptent les cas où l'Oberfeld de tous les constituants qui comportent la tête de la particule est accessible, permettant ainsi les cas de *Zwischenstellung*. Nous illustrons ce cas par les exemples en (119), que nous empruntons à Bouma et van Noord 1998 (exemple 58) :

(119) a. ... dat Jan Marie zou<sub>1</sub> hebben<sub>2</sub> aan<sub>4</sub> gesproken<sub>3</sub>  
 ... que Jan Marie aura<sub>1</sub> avoir<sub>2</sub> particule-sur<sub>4</sub> parlé<sub>3</sub>  
 ... 'que Jan aura parlé à Marie.'

<sup>67</sup> Les infinitifs en *-te*, l'équivalent néerlandais des infinitifs en *zu* allemands, ne peuvent pas se placer directement dans la parenthèse droite et doivent donc obligatoirement ouvrir un domaine enchâssé.

- b. ... dat Jan Marie zou<sub>1</sub> aan<sub>4</sub> hebben<sub>2</sub> gesproken<sub>3</sub>
- c. ... dat Jan Marie aan<sub>4</sub> zou<sub>1</sub> hebben<sub>2</sub> gesproken<sub>3</sub>
- d. \*... dat Jan Marie zou<sub>1</sub> hebben<sub>2</sub> gesproken<sub>3</sub> aan<sub>4</sub>

Nous voyons donc que le modèle topologique permet d'expliquer de manière élégante la structure interne du complexe verbal. Cette extension du modèle topologique est, à notre connaissance, quelque chose de relativement nouveau. Des approches antérieures qui se basent sur le modèle topologique comme Kathol 1995, 2000 utilisent d'autres processus pour régler le problème de l'ordre des mots dans la parenthèse droite.

### 3.3 Complexité du calcul topologique et la « troisième construction »

Avant de clore cette section sur l'analyse topologique du complexe verbal, revenons sur la question du calcul topologique. La complexité de la structure topologique qui est impliquée par la création d'un domaine ne correspond pas toujours à la complexité qu'éprouve une personne (ou une machine) qui a comme tâche de formuler une telle structure. D'une manière intuitive, la coordination de deux phrases peut être ressentie comme plus complexe que la même information contenue dans deux phrases distinctes : autrement dit, l'information contenue dans une seule structure semble être plus complexe que lorsque la même information est distribuée dans deux structures différentes. De la même manière, la création d'un nouveau domaine verbal dans le Nachfeld, autrement dit l'extraposition, créant une structure avec deux domaines, peut paraître d'une plus grande complexité formelle qu'un complexe verbal qui constitue une seule structure, mais les deux types de constructions sont couramment utilisés. En particulier, la réalisation des chaînes subordinatives des verbes non finis qui sont très longues devient beaucoup plus facile (en production et en compréhension) lorsqu'une partie de cette chaîne est « extraposée » dans le Nachfeld.

La construction topologique particulière, qu'on trouve souvent sous le nom de « troisième construction »<sup>68</sup> présente un problème similaire. Le terme de « troisième construction » reste très peu et très mal défini dans la littérature ; la seule chose que l'on peut retenir et que la « troisième construction » renvoie au cas où le dernier verbe de la chaîne subordonative est un infinitif en *zu* qui ouvre un nouveau domaine dans le Nachfeld et dont les arguments nominaux se placent dans le Vorfeld ou le Mittelfeld. Nous donnons à la Figure 36 une représentation topologique d'une « troisième construction » de notre exemple de base.

Grewendorf 1988 considère que la « troisième construction » renvoie à un cas d'*Oberfeldumstellung*. Cependant, il est clair qu'une telle analyse est fautive (au moins si l'on suit notre définition) car, comme le montre aussi Müller 1999:345, dans le cas de la subordination de la phrase complète, l'ordre de l'*Oberfeldumstellung* nous obligerait à avoir un ordre  $V_1V_2V_3$  alors que  $V_2V_1V_3$  est le seul ordre possible (phrases sous (120)).

(112)

Vorfeld	(	Mittelfeld	)	Nachfeld		
				MF	)	NF
Niemand	hat	diesem Mann den Roman	versprochen		zu lesen	
personne (nom.)	a	à cet homme (dat.) le roman (acc.)	promis		de lire	

Figure 36: Structure topologique de la « troisième construction ».

<sup>68</sup> Troisième car il s'agit des constructions après les constructions cohérentes et incohérentes, selon Bech 1955. Le terme de « troisième construction » a été introduit pour la première fois par Besten & Rutten 1989 pour une construction similaire du néerlandais.

- (120) a. Ich glaube, dass niemand diesem Mann den Roman versprochen<sub>2</sub> hat<sub>1</sub> zu lesen<sub>3</sub>.  
 je pense que personne (nom.) à cet homme (dat.) le roman (acc.) promis<sub>2</sub> a<sub>1</sub> de lire<sub>3</sub>  
 'je pense que personne n'a promis à cet homme de lire le livre.'
- b. \* Ich glaube, dass niemand diesem Mann den Roman hat<sub>1</sub> versprochen<sub>2</sub> zu lesen<sub>3</sub>.  
 je pense que personne (nom.) à cet homme (dat.) le roman (acc.) a<sub>1</sub> promis<sub>2</sub> de lire<sub>3</sub>

Par contre, bien que Müller constate qu'il s'agit d'une construction incohérente, il considère que le verbe à la « troisième construction » (le verbe *zu lesen* dans l'exemple (120)) fait partie du complexe verbal. Or cette affirmation est en contradiction avec l'exemple (121) qu'il donne, où on voit clairement la grande liberté que nous avons de placer des arguments directement devant le verbe à la « troisième construction » :

- (121) Jetzt hab' ich dir vergessen, einen Löffel mitzubringen<sup>69</sup>.  
 maintenant ai je te oublié, une cuillère à apporter\_avec  
 'Maintenant, j'ai oublié de t'amener une cuillère.'

Müller classe cette phrase comme faisant partie des « troisièmes constructions » en expliquant que le dépendant au datif de *mitbringen* 'apporter avec' se trouve dans le Mittelfeld de la phrase principale. Or, si on suppose comme Müller que la « troisième construction » a lieu, comme pour l'*Oberfeldumstellung*, dans le complexe verbal, non seulement faut-il ajouter un phénomène d'ordre dans le complexe verbal qui est très différent de l'*Oberfeldumstellung* (c'est ce que fait précisément Müller 1999 à l'aide d'un trait spécial dans sa structure HPSG), mais il faut aussi expliquer pourquoi une série quelconque d'éléments non verbaux peut apparaître dans le complexe verbal dès qu'un argument du verbe à la « troisième construction » se trouve dans le Mittelfeld<sup>70</sup>. Il nous semble donc préférable de traiter les « troisièmes constructions » comme un cas d'émancipation.

Bien que dans notre approche topologique l'ordre des mots de cette construction ne présente rien d'inhabituel, la « troisième construction » pose néanmoins un problème explicatif : la facilité avec laquelle on produit de telles phrases ne semble pas correspondre à la complexité de la structure topologique (qui contient un nouveau domaine enchâssé et une émancipation). Nous n'avons pas de réponse simple à cette question, mais il semble clair que les locuteurs ont plus de mal à gérer la complexité communicative que la complexité de la structure topologique. En effet, l'exemple (121) a une structure communicative très simple, si la phrase est accompagnée de la bonne structure prosodique, où on voit clairement apparaître un premier grouper formant le topique (la première partie de la phrase jusqu'à la virgule) et un autre formant le focus (le domaine enchâssé). Le fait que *dir* 'te' a été émancipé pour aller se placer dans le Mittelfeld simplifie la phrase puisque *dir* en tant que pronom est clairement thématique, permettant ainsi une séparation nette entre le topique (ce dont on parle), et le focus (ce qu'on en dit) de la phrase.

L'ordre des mots de l'allemand semble donc jouer un rôle dans la simplification et la clarification de la structure communicative, si la structure topologique le permet. La complexité de la structure topologique ne se fait entendre que dans des rares cas, comme par exemple dans le cas où on essaie d'éviter un complexe verbal complexe.

<sup>69</sup> D'après Müller, il s'agit d'un énoncé original de Detmar Meurers.

<sup>70</sup> Nous supposons que la raison principale de cette analyse que fait Müller repose sur le fait qu'il ne veut pas permettre le positionnement d'éléments nominaux dans le Mittelfeld au même titre que leur topicalisation dans le Vorfeld, car, de manière générale, c'est beaucoup plus difficile à motiver communicativement.

## 4 Complétives et Relatives

### 4.1 Les complétives

#### 4.1.1 Le traitement des complétives

Dans cette section, nous incluons une analyse des complétives. Nous considérons en suivant Tesnière 1959, Mel'čuk 1988 et les récents travaux de la syntaxe X-barre que la tête d'une complétive est le complémenteur. Ainsi, dans l'analyse topologique, le complémenteur, tête de la phrase complétive ouvre un nouveau domaine qui consiste d'un champ de complémenteur (qui correspond à la parenthèse droite de la proposition principale), d'un Mittelfeld et d'une parenthèse droite.



Figure 37: Les champs du domaine de la complétive ouvert par le complémenteur.

Comme dans le domaine principal, le champ du complémenteur est le champ tête. Le complémenteur se place donc dans le champ du complémenteur tandis que le verbe principe occupe la position finale du domaine, la parenthèse droite (qui peut être suivie de ses constituants extraposés exactement de la même manière que la proposition principale et former ainsi un complexe verbal). Notons que le verbe fini qui dépend d'un complémenteur ne peut pas ouvrir un nouveau domaine. Qu'un mot ne puisse pas ouvrir est un phénomène assez rare. Dans un sens, nous considérons que le complémenteur et le verbe fini sont liés d'une manière inséparable, tout comme le sont le nom et son déterminant qui doivent aussi rester dans le même domaine (voir la notion de « translation » chez Tesnière 1959 et Abney 1987 pour une hypothèse DP dans le cadre de la théorie X-barre).

Nous illustrons le traitement des complétives en prenant comme exemple les linéarisations que l'on peut obtenir à partir de l'arbre syntaxique de surface que nous présentons dans la Figure 38.

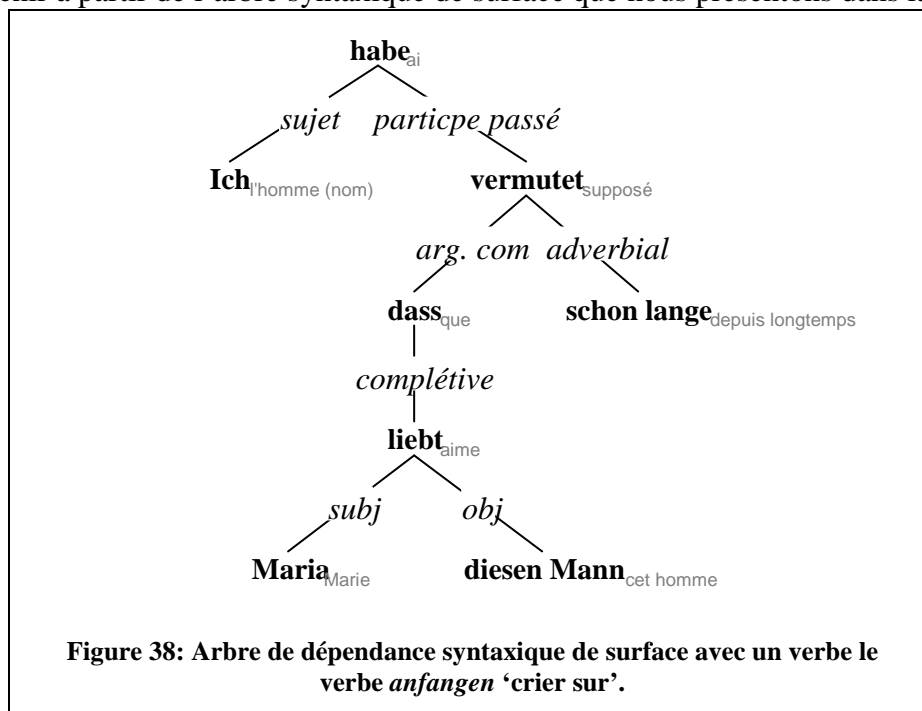


Figure 38: Arbre de dépendance syntaxique de surface avec un verbe le verbe *anfangen* 'crier sur'.

Le complémenteur ouvre de préférence un domaine dans le Nachfeld. Dans la Figure 39 nous donnons la structure topologique avec domaine du complémenteur ouvert dans le Nachfeld.

<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>			
				<i>KF</i>	<i>MF</i>	)	<i>NF</i>
Ich	habe	schon lange	vermutet,	dass	Maria diesen Mann	liebt	
Je (nom.)	Ai	depuis longtemps	supposé	que	Marie (nom.) cet homme (acc.)	aime	

**Figure 39: Domaine du complémenteur ouvert dans le Nachfeld :**  
 ‘Je me suis douté depuis longtemps que Marie aime cet homme.’

Contrairement aux domaines des infinitifs, les arguments nominaux du domaine complétif ne peuvent pas s’émanciper dans un domaine supérieur, et donc la linéarisation dans la Figure 40 est impossible :

<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>			
				<i>KF</i>	<i>MF</i>	)	<i>NF</i>
* Ich	habe	schon lange diesen Mann	vermutet,	dass	Maria	liebt	
Je (nom.)	ai	depuis longtemps cet homme (acc.)	supposé	que	Marie (nom.)	aime	

**Figure 40: Les dépendants nominaux dont le gouverneur se trouve dans un champ complémenteur ne peuvent pas s’émanciper. Cette linéarisation est donc impossible.**

La seule exception se trouve dans les pronoms interrogatifs : ces derniers peuvent en effet aller dans le Vorfeld du domaine principal. Nous illustrons ce cas avec la phrase (122) dont nous donnons l’analyse topologique dans la Figure 41.

- (122) ? Wen hast du schon lange vermutet, dass Maria liebt.  
 qui as tu depuis longtemps supposé, que Marie aime.  
 ‘Qui supposais-tu depuis longtemps que Marie aime ?’

<i>Vorfeld</i>	(	<i>Mittelfeld</i>	)	<i>Nachfeld</i>			
				<i>KF</i>	<i>MF</i>	)	<i>NF</i>
? Wen	hast	du schon lange	vermutet,	dass	Maria	liebt	
Qui	as	tu depuis longtemps	supposé	que	Marie (nom.)	aime	

**Figure 41: Les dépendants nominaux dont le gouverneur se trouve dans un champ complémenteur ne peuvent pas s’émanciper. Cette linéarisation est donc impossible.**

#### 4.1.2 Les complétives sans complémenteurs

En allemand, les complétives sans complémenteurs représentent un lieu de confusion dans la littérature. Dans notre approche topologique, nous considérons que les complétives sans complémenteurs peuvent être analysés de la même manière que les complétives : on considère qu’en l’absence de complémenteur, le même domaine complétif est ouvert par le verbe qui représente la tête de la complétive. On peut illustrer ce cas par la fameuse phrase de Boèce que nous reprenons sous (123).

- (123) Hättest du geschwiegen, wärst du ein Philosoph geblieben  
 as t- être\_silencieux, serais tu un philosophe resté  
 ‘Eusses-tu gardé ton silence, tu serais devenu un grand philosophe.’

La forme au conditionnel du verbe HABEN *hättest* 'aurais' incorpore en quelque sorte le complément *wenn* 'si' ce qui rend possible la paraphrase que nous donnons sous (124) :

- (124) Wenn du geschwiegen hättest, wärest du ein Philosoph geblieben  
 si tu être\_silencieux aurais, serais tu un philosophe resté  
 'Si tu avais gardé ton silence, tu serais devenu un grand philosophe.'

Et comme une complétive peut apparaître soit dans le Nachfeld, soit dans le Vorfeld, l'ordre des mots tel que la phrase (125) est aussi possible :

- (125) Du wärest ein Philosoph geblieben hättest du geschwiegen  
 tu serais un philosophe resté      aurais tu être\_silencieux  
 'Tu serais devenu un grand philosophe eusses-tu gardé ton silence.'

Comme pour les complétives habituelles l'intrapolation est plus difficile car elle requiert une structure communicative particulière. Frey 2000 propose comme test permettant de marquer la frontière entre un topique et un focus, l'insertion d'un adverbial de phrase comme *wahrscheinlich* 'probablement'<sup>71</sup>. Or, ce test fait clairement apparaître que *hättest du geschwiegen* fait clairement partie du topique (phrase

- (126) Du wärest, hättest du geschwiegen, wahrscheinlich ein Philosoph geblieben  
 tu serais, aurais tu être\_silencieux, probablement un philosophe resté  
 'Eusses-tu gardé ton silence, tu serais devenu probablement un grand philosophe.'

Comme le remarque déjà Drach 1937:34, les phrases comme (127) et (128) constituent un cas légèrement problématique pour la composition du domaine complétif.

- (127) Er benimmt sich, als      wäre er der Herr      im      Haus<sup>72</sup>.  
 il comporte se,      comme était il le maître dans\_le maison  
 'Il se comporte comme s'il était le chef de maison.'

- (128) Er benimmt sich, als      ob er der Herr      im      Haus      wäre  
 il comporte se,      comme si il le maître dans\_le maison était  
 'Il se comporte comme s'il était le chef de maison.'

De la même manière que la structure de la phrase (123), la forme conditionnelle fait que le complément *ob* 'si' est optionnel (phrase (128)). Les dépendants verbaux qui se trouvent dans le champ complétif (dans le cas de notre exemple, ouvert par *wäre* 'était') se placent de manière normale dans la parenthèse droite, comme en témoigne l'exemple

- (129) Er benimmt sich, als wäre er aus dem Osten gekommen  
 il comporte se, comme était il dehors de\_le Est venu  
 'Il se comporte comme s'il venait de l'Est.'

<sup>71</sup> Nous reviendrons sur cette question dans la section IV.5, lorsque nous aborderons la question de l'organisation topologique dans le Mittelfeld.

<sup>72</sup> Nous empruntons cet exemple à Drach 1937:33

En fait, le problème de la phrase (127) repose sur la place topologique qu'on attribue à la conjonction *als* 'comme'. Deux choix se présentent : soit nous considérons que *als* 'comme' ouvre un domaine spécifique dont il occupe le champ à gauche X et qui propose une place Y au domaine complétif ouvert par *ob* 'si' ou par le verbe (voir Figure 42), soit on considère que *als* 'comme' occupe le champ du complémenteur ensemble avec *ob* 'si' (ou le verbe qui occupe habituellement seul le champ du complémenteur).

En fait, nous optons pour la deuxième analyse, parallèlement à l'analyse topologique que nous proposons pour le syntagme nominal dans la section IV.6<sup>73</sup>. Nous considérons donc que la conjonction *als* 'comme' ouvre lui-même le domaine complétif dont il occupe le champ de

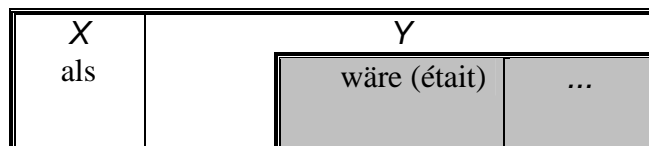


Figure 42: Domaine spécifique ouvert par la conjonction *als* 'comme'.

complémenteur mais ce domaine est limité dans le champ du complémenteur. Ce domaine propose une seule place à la droite de *als* 'comme' pour son dépendant « complémenteur », comme dans le cas du complexe verbal, où le verbe dans la parenthèse droite ouvre une place pour son argument verbal. Nous donnons une représentation de cette structure topologique dans la Figure 43.

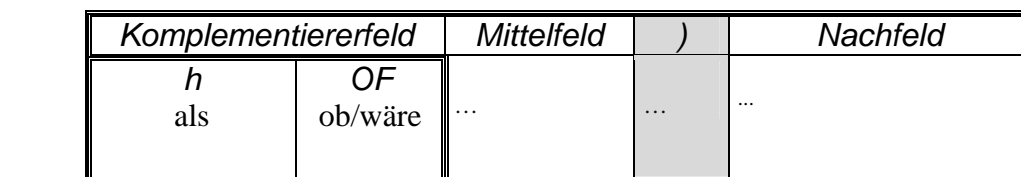


Figure 43: Domaine spécifique ouvert par la conjonction *als* 'comme'.

### 4.1.3 Parenthétique et complémentation

La phrase (130), bien que possible en allemand, semble très peu élégante et on trouve de manière générale très peu d'extraction hormis celle de la complétive dans la langue écrite.

(130) <sup>?</sup>Wen vermutest du seit langem, dass Peter liebt?

qui suppose tu depuis longtemps, que Pierre aime  
 'Qui supposes-tu depuis longtemps que Pierre aime ?'

On pourrait trouver une réponse à ce phénomène dans l'existence d'alternatives d'ordre. D'abord, on peut placer le pronom interrogatif *in situ*, c'est-à-dire à l'intérieur du domaine complétif (131) et dans le cas d'une double interrogation ((132)ab) ou d'une construction dite « W...W » (133) où on place au moins un mot *qu- in situ*. Nous ne discutons pas ces constructions plus en détail car d'un point de vue d'analyse topologique, elles ne posent pas de problème<sup>74</sup>.

(131) Du vermutest, dass Peter wen liebt?

tu supposes, que Pierre qui aime

<sup>73</sup> Comme en français, *als* 'comme' peut aussi introduire des syntagmes nominaux et le même problème se pose pour l'analyse topologique des syntagmes nominaux (voir section IV.6).

<sup>74</sup> Pour une discussion détaillée sur les restrictions sous-jacentes, voir Höhle 1996 et Reis 2000.



- 'Qui supposes-tu que Pierre aime ?
- (132) a. <sup>??</sup>Wer vermutest du, dass wen liebt?  
 qui supposes tu, que Pierre qui aime  
 'Que supposes-tu sur qui aime qui ?'
- b. <sup>?</sup>Du vermutest, dass wer wen liebt?  
 tu supposes, que qui qui aime  
 'Que supposes-tu sur qui aime qui ?'
- (133) Was vermutest du, wen Peter liebt?  
 quoi supposes tu, qui Pierre aime  
 'Que supposes-tu ? Pierre aime qui ?'

Une possibilité de paraphraser (130) consiste d'utiliser une parenthétique comme dans (134).

- (134) Wen, vermutest du, liebt Maria?  
 qui, supposes tu aime Maria  
 'Qui supposes-tu que Marie aime.'
- (135) Maria, vermute ich, liebt Peter.  
 Marie, suppose je aime Pierre  
 'Marie, je suppose, aime Pierre.'

Cette possibilité est aussi accessible sans mot *qu-*, comme on peut le voir dans l'exemple (135). Une condition qui doit être vérifiée est que la parenthétique doit avoir une valeur sémantique affirmative, et fonctionne donc comme une adverbiale de phrase qui renforce ou modifie légèrement la phrase principale ; par contre, elle ne peut pas contenir d'information qui change complètement la sémantique de la phrase, comme c'est le cas dans les phrases (136) et (137) :

- (136) a. \*Wen, vermutest du nicht, liebt Maria?  
 qui, supposes tu non aime Maria  
 '\*Qui, ne supposes-tu pas, aime Marie ?'
- b. \*Maria vermute ich nicht liebt Peter.  
 Marie, suppose je non aime Pierre  
 '\*Marie, je ne suppose pas, aime Pierre.'
- (137) a. <sup>??</sup>Wen, bezweifelst du liebt Maria?  
 qui, doutes tu aime Maria  
 '\*Qui, doutes-tu, aime Marie ?'
- b. <sup>??</sup>Maria bezweifele ich liebt Peter.  
 Marie, doute je aime Pierre  
 '\*Marie, je doute, aime Pierre.'

Dans la plupart des cas, une parenthétique est courte et le contenu est affirmatif (on trouve souvent *glaub'ich* 'je crois') mais on peut également trouver des constructions auxiliaires comme dans (138) et (139).

- (138) Maria, hat er gesagt, liebt den Peter.  
 Maria, a il dit, aime Pierre  
 'Marie, a-t-il dit, aime Pierre.'

(139) Maria, hat er gesagt als er kam, liebt den Peter.

Maria, a il dit comme il venait, aime Pierre  
'Marie, a-t-il dit en arrivant, aime Pierre.'

D'un point de vue de l'analyse topologique, il semble donc claire que la structure topologique est identique à celle du domaine complétif, à la différence que le premier champ n'est plus un champ complétif mais une parenthèse gauche :



Figure 44: Domaine spécifique de la phrase parenthétique.

Cependant, un problème persiste quand on essaie de décrire le placement de ce domaine dans le domaine principale, car contrairement aux adverbiaux, ce nouveau domaine apparaît entre le Vorfeld et la parenthèse gauche comme il apparaît dans les phrases (134) et (135) mais cela impliquerait qu'on ajoute un nouveau champ entre les champs de Vorfeld et de parenthèse gauche ; par ailleurs, il ne peut pas remplir lui-même le Vorfeld (voir phrases sous (140)). D'ailleurs, ce placement autour de la parenthèse gauche est préférable à un placement ailleurs dans le domaine principal (phrase (140)d.).

- (140) a. \* Glaube ich liebt Maria Peter.  
pense je aime Marie Pierre
- b. # Glaube ich Maria liebt Peter.  
pense je Marie aime Pierre
- c. Maria liebt, glaube ich, Peter.  
Marie aime, pense je, Pierre  
'Marie aime, je pense, Pierre.'
- d. ? Maria liebt Peter, glaube ich.  
Marie aime Pierre pense je  
'Marie aime Pierre, je pense.'

Pour résoudre ce problème, nous devrions introduire de nouveaux champs dans le domaine principal juste pour traiter les parenthétiques. Or nous considérons que ni les données, ni les analyses que l'on trouve de ce phénomène sont suffisamment claires pour modifier la structure topologique de base. Nous laissons donc l'analyse topologique des parenthétiques pour une recherche future.

## 4.2 Les relatives et pronoms résomptifs

### 4.2.1 Les phrases relatives dans une grammaire de dépendance

On réserve une place importante au traitement des relatives en raison des difficultés qu'il présente dans la description syntaxique des langues. Les relatives, en l'allemand comme dans beaucoup d'autres langues, reçoivent une analyse de « dépendances de longue distance » et constituent ainsi des cas intéressants pour une grammaire basée sur une structure de constituants comme TAG mais aussi pour la structure topologique que nous proposons. Dans cette section, nous présentons le traitement des relatives dans une grammaire de dépendance. Les idées développées ici sont essentielles pour montrer comment le modèle topologique réussit à gérer d'une manière simple et élégante les constructions relatives enchâssées dites de « pied-

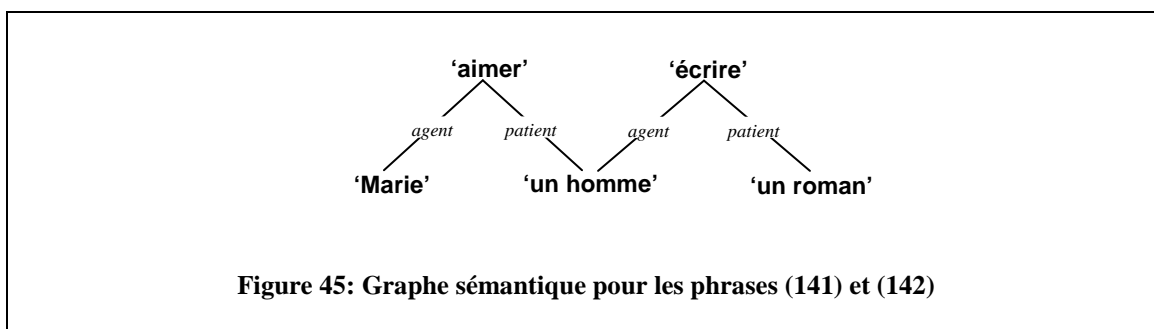
piping<sup>75</sup> », ainsi que pour le développement de l'analyse TAG des relatives dans la section III.2.4. La conception de la phrase relative que nous présentons reprend des conceptions de Tesnière 1959 et des travaux basées sur ses idées de Mel'čuk 1988 et de Kahane 2001. Notre but est d'adapter ses analyses pour nos besoins dans une analyse de l'allemand.

Dans cette section, nous discuterons principalement des structures de dépendance syntaxique pour des phrases comme (141) et (142).

(141) Marie aime un homme qui écrit un roman.

(142) Un homme que Marie aime écrit un roman.

Notons d'abord que selon le point de vue de la structure prédicat-argument, si on laisse de côté les distinctions communicatives ou de portée des éléments, les deux arbres ont une structure de dépendance sémantique qui est identique (Figure 45, cf. Kahane et Mel'čuk 1999).



Les deux phrases se distinguent sémantiquement dans le sémantème qui résume leur contenu sémantique. Dans le sens de Polguère 1990, on peut affirmer que le prédicat *aimer* prédomine dans la phrase (141) tout comme le prédicat *écrire* prédomine dans la phrase (142). Dans la Théorie Sens-Texte, on note des prédicats prédominants dans les représentations de ce type par le soulignement. Le correspond syntactique du terme prédominant sera donc la racine de l'arbre de dépendance syntaxique correspondant. Une manière de représenter mentalement ce qui se passe est de s'imaginer que l'on empoigne le graphe par le terme prédominant et par la suite, la syntaxe fait en sorte que le graphe sémantique devienne un arbre, c'est-à-dire que tout « pende » à partir de ce nœud prédominant (cf. Candito & Kahane 1998). Dans l'arbre de dépendance syntaxique, afin de faire « pendre » *écrire* de *un homme*, nous devons « transférer » le verbe *écrire* en un modifieur syntaxique du nom : on introduit des mots spécialement conçus pour cette tâche de transfert, ou, pour utiliser le terme introduit par Tesnière, de translation.

Au niveau de la syntaxe de surface, entre la sémantique et la sortie finale de la génération du texte, on voit donc apparaître des mots « translatifs », comme par exemple le pronom relatif. Bien sûr, il nous faut nous poser la question du statut des éléments « translatifs » qui sont introduits au niveau syntaxique. Selon Tesnière, de la même manière qu'un complémenteur transfère une phrase en nom (ou au moins en un élément avec une distribution nominale), le pronom relatif fait qu'une phrase devient un modifieur d'un nom : c'est ce qui amène Tesnière à donner aux phrases relatives le nom de *proposition adjective*. Cependant, contrairement aux complémenteurs, les pronoms relatifs varient en fonction d'un deuxième rôle syntaxique qui est interne à la phrase relative même: ces derniers portent le cas que leur assigne le verbe (ou un autre gouverneur) dans la relative. « C'est donc que le pronom relatif est un mot de **nature double**, composé de deux éléments syntaxiques fondus ensemble, que l'analyse syntaxique révèle comme syntaxiquement distincts. » (Tesnière 1959:560, mise en valeur reprise de l'original). Ce double rôle du pronom relatif amène Tesnière soit à le redoubler dans la structure syntaxique, soit à le couper en deux : « qu- » représente la partie « complémentante », qu'il

<sup>75</sup> On reviendra sur ce terme à la fin de cette section.

appelle *transférème* et « -e » la partie nominale sous-catégorisée dans la phrase relative, appelée *anaphorème*.

Pour les deux phrases (141) et (142), nous obtenons donc les structures de la Figure 46<sup>76</sup>.

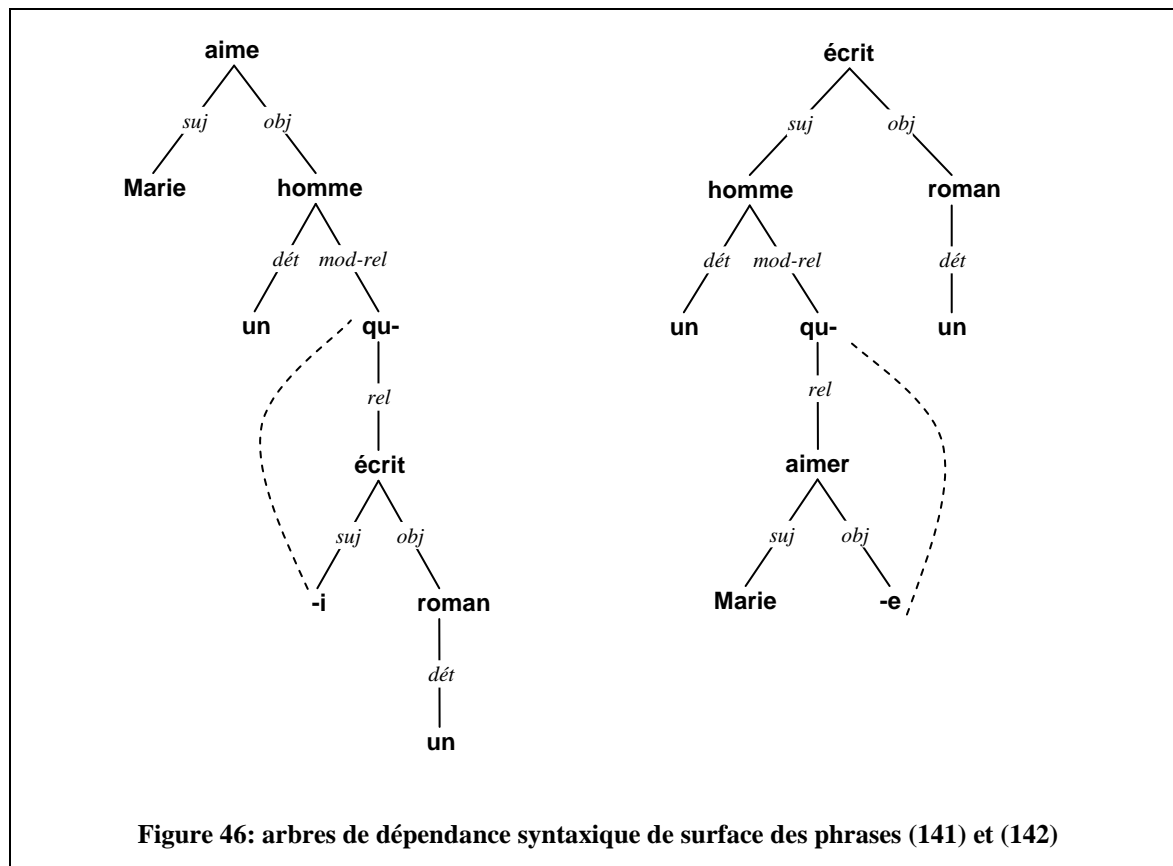


Figure 46: arbres de dépendance syntaxique de surface des phrases (141) et (142)

Cette analyse qui propose une séparation entre transférème et anaphorème est facilement traduisible en allemand : le pronom relatif de l'allemand fait l'accord en nombre et en genre avec l'antécédent et le cas est assigné par la place syntaxique de l'anaphorème. Prenons comme exemple la phrase (143), qui représente la traduction de la phrase phrases (141). Le pronom relatif est à la 3<sup>e</sup> personne masculin du singulier, dont la forme coïncide avec le déterminant et le pronom démonstratif et partage les même traits morphologiques, comme l'illustrent les phrases (144) et (145).

Dans la représentation syntaxique de la phrase (143) (premier arbre de la Figure 47), nous utilisons « d- » comme transférème (car le « d » apparaît aussi dans le complémenteur de l'allemand *dass* 'que') et « -er » comme anaphorème, car la terminaison indique le cas du pronom relatif.

(143) Maria liebt einen Mann, der einen Roman schreibt.

Marie aime un homme, qui un roman écrit.  
'Marie aime un homme qui écrit un roman.'

(144) Der Mann schreibt einen Roman.

'L'homme écrit un roman.'

(145) Der schreibt einen Roman.

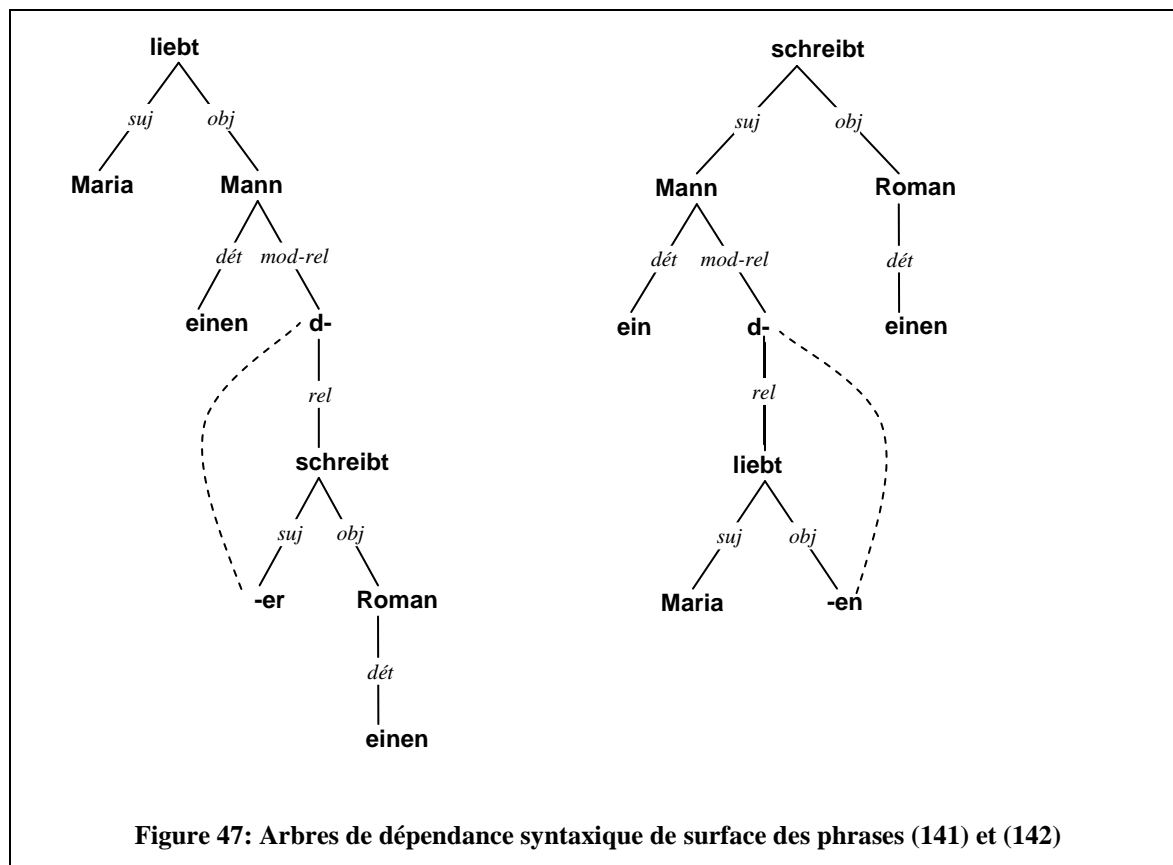
'Celui-ci écrit un roman.'

<sup>76</sup> Il s'agit de représentations légèrement modifiées par rapport aux idées de Tesnière, car nous n'utilisons pas le signe de « translatif ».

La structure de la phrase (146), qui représente une traduction de la phrase (142), donne aussi une structure identique à celle du français (deuxième arbre de la Figure 47).

(146) Ein Mann, den Maria liebt, schreibt einen Roman.

Un homme, que Marie aime, écrit un roman.  
 'Un homme, qu' aime Marie, écrit un roman.'



Un fait qui corrobore la double nature des pronoms relatifs est le fait qu'ils peuvent être dupliqués dans les constructions à la première ou la deuxième personne. Alors que, le pronom relatif qui, en allemand, est identique à des formes de déterminants et de pronoms démonstratifs, a un caractère intrinsèque de troisième personne. C'est la raison pour laquelle la modification des pronoms personnels à d'autres personnes que la troisième est difficile. Néanmoins, il existe deux possibilités pour réaliser une construction relative à la première ou la deuxième personne. Soit le pronom relatif garde sa double nature de *transférème* et d'*anaphorème* faisant en sorte que le verbe s'accorde à la troisième personne qui est implicite dans le pronom relatif, soit le pronom relatif est dupliqué : le pronom relatif ne garde que la fonction de *transférème* faisant en sorte que l'*anaphorème*, donc la reprise du nom modifié, se réalise par un pronom résomptif. Le premier cas est illustré par les exemples (147), (148) et (149) ; (150) et (151) sont des exemples de « redoublement » du pronom, le pronom résomptif est mis en gras.

(147) du, die wandelt wie Frühwind<sup>77</sup>

tu, qui se balade [3 pers. sing] comme tôt-vent  
 'Toi, qui te promène comme le vent du matin'

(148) Du, der so etwas erlebt hat<sup>78</sup>

tu, qui ainsi quelque chose vécu a

<sup>77</sup> R.M Rilke, Troisième élégie des *Duineser Elegien*

<sup>78</sup> Duden Grammatik §1246

‘Toi, qui as vécu de telles choses’

(149) Und ich, die gehofft hatte, Sie heute ausnahmsweise zu einem vernünftigen Gespräch zu bringen!<sup>79</sup>

Et je, qui [fém] espéré avait, vous aujourd’hui exceptionnellement à une raisonnable conversation de inciter

‘Et moi qui avais espéré pouvoir vous inciter aujourd’hui exceptionnellement, à une conversation raisonnable’

(150) Und du, der **du** so schön still sitzen kannst, hast du alles verstanden, was jetzt gesprochen wurde?<sup>80</sup>

et tu, qui tu ainsi beau immobile être-assis peux, as-tu tout compris, quoi maintenant parlé était ?

‘Et toi, qui peux rester si bien immobile, as-tu compris tout ce qui a été dit maintenant ?’

(151) Ihr, die **ihr** selten seid, was ihr sein wollt, niemals, was ihr sein solltet.<sup>81</sup>

vous, qui vous rarement êtes, que vous être voulez, jamais, que vous être devrez.

‘Vous qui êtes rarement ce que vous voulez être, et jamais ce que vous devrez être.’

Les deux choix présentés ci-dessus sont presque identiques (avec, selon la Duden Grammatik §1246, une légère préférence pour le « redoublement ») et ils constituent des paraphrases presque parfaites, bien que d’un point de vue sémantique, la troisième personne entraîne naturellement une plus grande réification de la personne (ou les personnes) à laquelle le pronom relatif fait référence. Notons par ailleurs que les exemples de relatives à la première ou la deuxième personne que nous avons trouvés, réalisent presque toujours la reprise du pronom au nominatif. Toutefois, il est possible de construire d’autres exemples, pas toujours très naturels, qui illustrent des reprises avec d’autres cas grammaticaux, comme en (152), (153) et (154)

(152) a. ? Mich, der mir vor Klößen graut, liebt keiner.

Me, que à\_moi de boulettes fait\_la\_hantise, aime personne.

b. ? Mich, dem vor Klößen graut, liebt keiner.

Me, à\_qui de boulettes fait\_la\_hantise, aime personne.

‘Personne ne m’aime, moi qui a horreur des boulettes’

(153) a. ?? Ihr, die euch mal wieder keiner hilft, seid arm dran.

Vous, que à\_vous encore à\_nouveau personne aide, êtes pitoyable

b. ? Ihr, denen mal wieder keiner hilft, seid arm dran.

Vous, à\_qui encore à\_nouveau personne aide, êtes pitoyable

‘Vous, qui n’êtes aimé par personne, êtes pitoyable.’

(154) a. ? Dir, der dich keiner liebt, graut vor Klößen.

Me, que à\_moi de boulettes fait\_la\_hantise, aime personne.

---

<sup>79</sup> M. von Ebner-Eschenbach, *Er lässt die Hand küssen*, propos tenu par la duchesse

<sup>80</sup> F. Kafka, *Der Gruftwächter*

<sup>81</sup> J. W. Goethe, Acte 2 scène 2, *Götz von Berlichingen*

b. ? Dir, den keiner liebt, graut vor Klößen.

Me, que à\_moi de boulettes fait\_la\_hantise, aime personne.  
 'Personne ne m'aime, moi qui a horreur des boulettes.'

Notons que le pronom relatif à la troisième personne doit indiquer le genre bien que les pronoms à la première ou la deuxième personne ne porte pas le genre. On voit donc clairement son caractère anaphorique. En effet, dans l'exemple cité ci-dessus, seul le pronom relatif de la phrase (149) est marqué par le genre féminin, indiquant ainsi qu'il s'agit d'un propos tenu par une femme. Cette remarque est importante : elle suggère que le caractère grammatical de l'accord entre le pronom relatif et son antécédent n'est peut-être pas de type syntaxique mais plutôt de type anaphorique.

Cependant, il est impossible de faire l'accord avec l'antécédent (à une personne autre que la troisième) sans introduire un pronom, comme on peut le voir dans la phrase (155) :

(155) ?? Du, die wandelst wir Frühwind

tu, qui te\_balades [2 pers. sing] comme tôt-vent  
 'Toi, qui te promènes comme le vent du matin'

Par ailleurs, la reprise par un pronom résomptif n'est pas possible à la troisième personne (exemples (156) et (157)), car le pronom relatif porte déjà les traits de la troisième personne :

(156) \* Peter, der **er** wandelt wir Frühwind

Peter, qui il [pronom personnel] se\_balade [3 pers. sing] comme tôt-vent  
 'Pierre, qui se promène comme le vent du matin'

(157) ?? Er, der **er** wandelt wir Frühwind

Lui, qui il [pronom personnel] se\_balade [+ 3 pers. sing] comme tôt-vent  
 'Lui, qui se promène comme le vent du matin'

L'examen des exemples ci-dessus illustre bien l'utilité de la distinction que fait Tesnière de la partie « complémentante » et la partie argumentale du pronom relatif. Dans le cas de la reprise pronominale, les deux nœuds d'anaphorème et de transfère correspondent tout simplement à deux mots différents dans la phrase. Dans le cas d'un pronom relatif habituel, nous avons un mot qui correspond à ces deux nœuds.

Malgré différents indices pour un traitement du pronom relatif en tant qu'anaphore, nous considérons que l'accord du pronom relatif est un phénomène syntaxique.

Nous avançons trois raisons : premièrement, la relative peut apparaître à deux positions différentes de la phrase différentes : adjacente au nom modifié ou extraposée dans le Nachfeld. Nous discuterons le problème de l'extraposition dans la section III.2.4 L'extraposition, donc le placement non adjacent de la relative peut nous amener à une analyse anaphorique du pronom relatif dans ce cas. Pourtant, si l'on veut proposer un traitement uniforme pour la relative il sera non souhaitable de traiter un cas comme anaphorique et le deuxième cas comme syntaxique.

Deuxièmement, contrairement aux anaphores réalisées par les pronoms personnels, l'antécédent du pronom relatif se trouve toujours à l'intérieur de la même phrase, même dans le cas des relatives extraposées. Nous illustrons ces cas dans les exemples suivants.

(158) Ein Mann schloss die Tür ab, den ich noch nie gesehen habe.

Un homme fermait\_à\_clé la porte PARTICULE, que je encore jamais vu ai.  
 Un homme que je n'avais jamais vu fermait la porte à clé.

(159) Ein Mann schloss die Tür ab, die frisch gestrichen war.

Un homme fermait\_à\_clé la porte PARTICULE, qui fraîchement peint était.

Un homme fermait la porte à clé qui était fraîchement peinte.

(160) Ich habe sein Taschentuch gesehen. Er schloss die Tür ab. Es fiel auf den Boden.

Je ai son mouchoir vu. Il fermait\_à\_clé la porte PARTICULE. Il (neutre et donc coréférent avec le mouchoir) tombait sur la terre.

J'ai vu son mouchoir. Il fermait la porte à clé. Il (le mouchoir) tombait par terre.

(161) \*Ich habe sein Taschentuch gesehen. Er schloss die Tür ab, das auf den Boden fiel.

Je ai son mouchoir vu. Il fermait\_à\_clé la porte PARTICULE, qui sur la terre tombait.

J'ai vu son mouchoir. Il fermait la porte à clé. Il tombait par terre.

Bien que la relative extraposée puisse être coréférente avec n'importe quel élément nominal de la phrase (pourvu qu'il fasse l'accord en genre et en nombre, cf. (158) et (159)), les pronoms relatifs, contrairement aux pronoms personnels, ne peuvent pas être coréférents avec un éléments nominal externe à la phrase (cf. (160) et (161))

Enfin, la troisième raison repose sur la différence topologique entre le pronom relatif et le pronom démonstratif. En effet, bien qu'identique en forme, le pronom relatif déclenche un domaine enchâssé avec le verbe dans la parenthèse droite tandis que le pronom démonstratif apparaît dans un champ principal et le verbe peut occuper la parenthèse gauche. Cette distinction est illustrée par les phrases (159) et (162).

(162) Ein Mann schloss die Tür ab, und die war frisch gestrichen.

Un homme fermait\_à\_clé la porte PARTICULE, qui fraîchement peint était.

Un homme fermait la porte à clé qui était fraîchement peinte.

Nous considérons qu'il est souhaitable de traiter la question de la relative comme phénomène syntaxique en prenant en compte sa double nature telle qu'elle a été proposée par Tesnière. Par la suite, nous discutons du cas des modifieurs relativisés. Puis, dans la section suivante, nous proposerons une analyse topologique des relatives.

#### 4.2.2 Relativisation des modifieurs

Il est facile d'élargir cette analyse aux cas de relativisations des modifieurs illustrées par la phrase suivante :

(163) Maria klettert auf den Balkon, auf dem Peter singt.

Marie grimpe sur le balcon sur lequel Pierre chante.

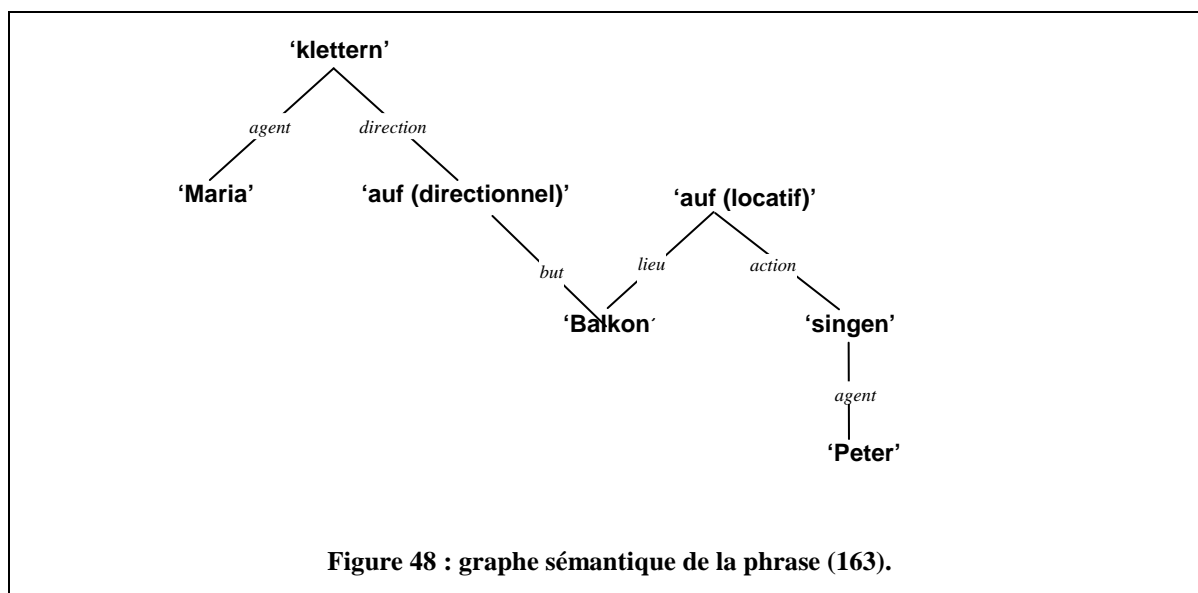
'Marie grimpe au balcon où chante Pierre.'

Au niveau sémantique, il est essentiel de faire la distinction entre, d'une part la direction qui est un argument du verbe *klettern* 'grimper' et, d'autre part le locatif qui prend lui-même deux arguments, l'endroit et l'action qui y a lieu. La Figure 48 montre un graphe sémantique possible de la phrase (163)<sup>82</sup>.

---

<sup>82</sup> Il est possible, sans que cela soit essentiel pour la présente analyse, de représenter *Balkon* directement comme argument de *klettern* 'grimper', mais une telle analyse a l'inconvénient que le verbe sous-catégorise plutôt une direction, représenté par le terme 'auf (directionnel)' que l'objet 'Balkon'.





Par contre, au niveau sémantique, le fait qu'un syntagme prépositionnel soit argument ou modifieur ne joue aucun rôle ; les deux syntagmes prépositionnels gouvernés par *auf* se comportent de manière identique aussi bien par rapport à leurs arguments nominaux que par rapport aux verbes qu'ils modifient. Nous obtenons donc la structure syntaxique de surface de la Figure 49.

Le fait que le *auf* locatif soit placé au début de la phrase relative est appelé « pied-piping », un terme introduit par Ross 1967. L'idée transformationnelle derrière ce terme est l'image que le pronom relatif est extrait et placé au début de la phrase relative, et qu'il tire derrière lui son gouverneur syntaxique, la préposition. Une telle analyse est compréhensible sous l'impression de liberté qu'offre l'anglais, illustrée par les phrases (164) et (165) : Soit on peut faire comme en français ou l'allemand et placer la préposition devant le pronom relatif, c'est-à-dire le cas (8), soit on peut le placer à la fin de la phrase relative, comme dans (165). Ce dernier cas est nommé *preposition stranding* ('échouage de préposition').

(164) Mary climbs on the balcony that Peter sings on.

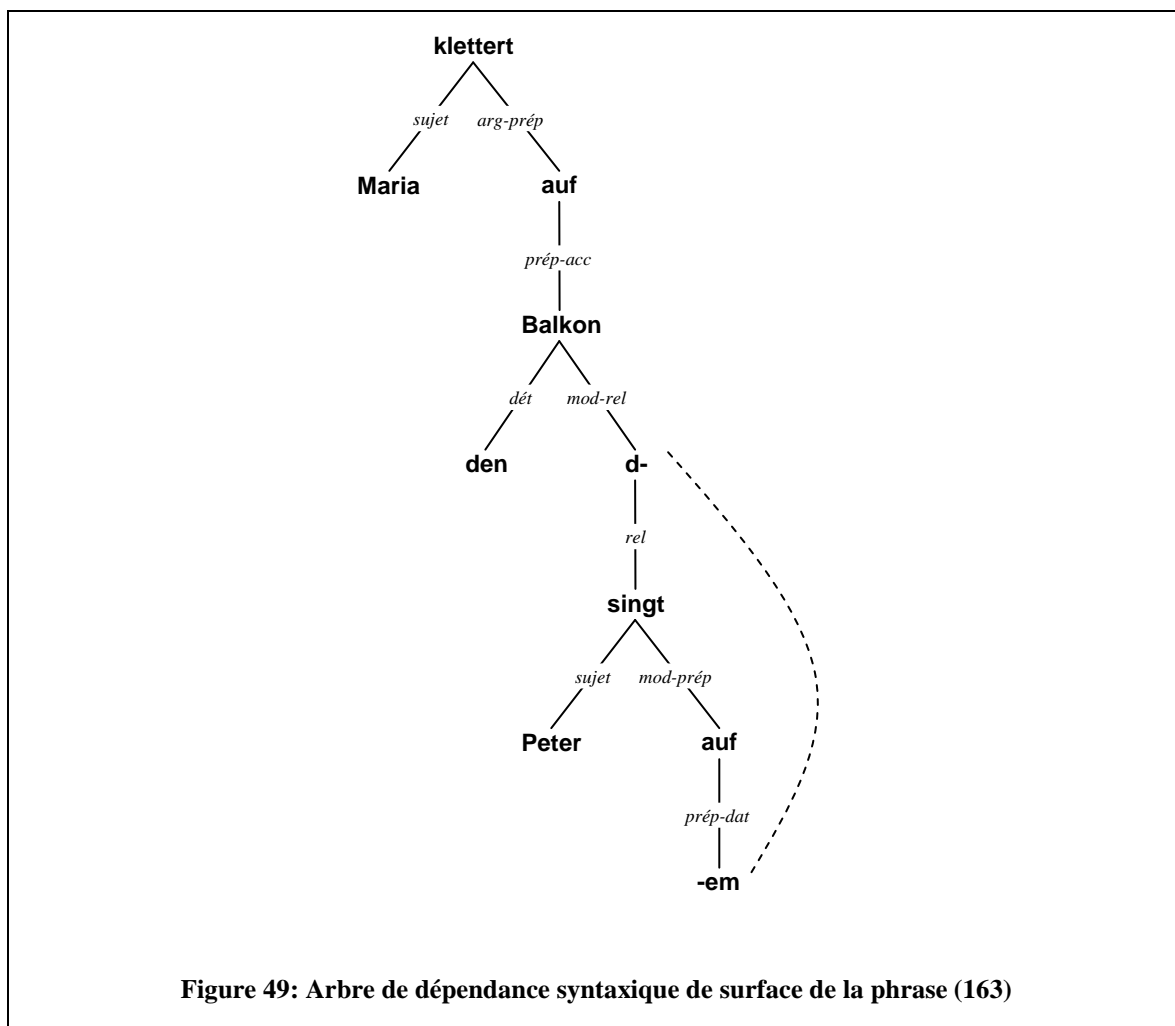
Marie grimpe sur le balcon que Pierre chante dessus.  
'Marie grimpe le balcon où chante Pierre.'

(165) Mary climbs on the balcony on which Peter sings.

Marie grimpe sur le balcon sur lequel Pierre chante.  
'Marie grimpe le balcon où chante Pierre.'

En utilisant ces termes, on peut donc affirmer que le français et l'allemand sont des langues à pied-piping prépositionnel obligatoire.

Nous verrons dans la section 4.2.3 que l'allemand présente en plus, la possibilité du pied-piping verbal, c'est-à-dire qu'autour du pronom relatif, un syntagme est créé, qui peut même inclure un verbe dominant le pronom relatif.

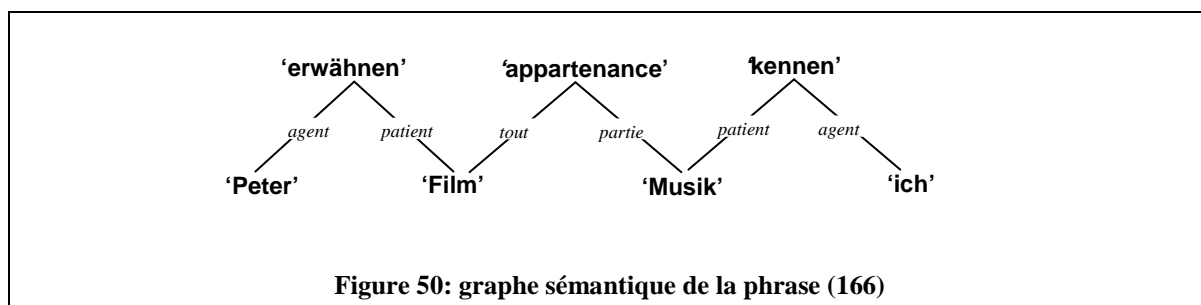


Il reste à invoquer les cas où l'élément relativisé ne dépend pas du verbe mais d'un élément enchâssé, que ce soit un argument ou un modifieur du verbe. De tels cas ne sont pas d'un intérêt particulier d'un point de vue de la dépendance syntaxique parce que la structure de dépendance syntaxique ne rencontre pas de problème à incorporer un niveau supplémentaire ; par contre, ces constructions posent de grands problèmes pour les approches basées sur des structures syntagmatiques, comme nous le verrons dans la section III.2.4 où nous présentons l'analyse des relatives dans le formalisme des grammaires d'arbres adjoints. La phrase (166) illustre un tel cas.

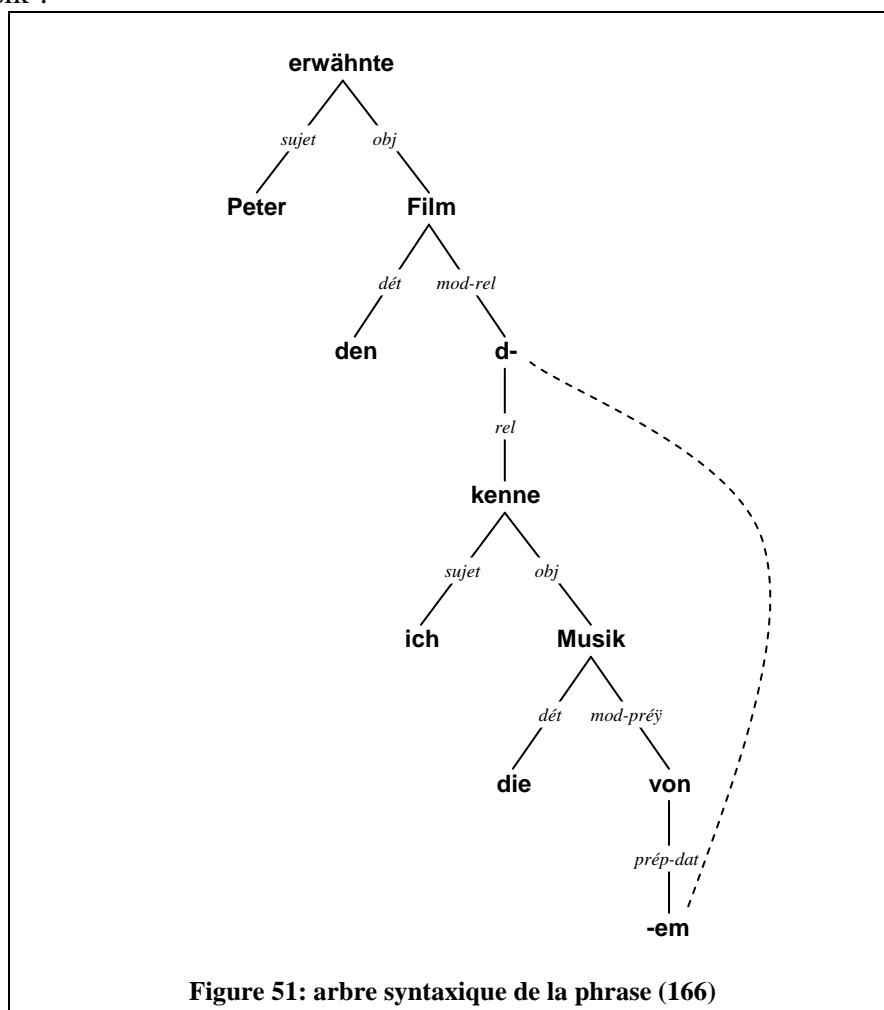
(166) Peter erwähnte den Film, von dem ich nur die Musik kenne.

Pierre mentionnait le film, de qui je seulement la musique connais  
 'Pierre a mentionné le film dont je ne connais que la musique.'

Le graphe de la Figure 50 montre une possibilité de représenter la structure prédicat-argument sous-jacente : 'le film' et 'la musique' sont liés par un lien d'appartenance, la musique faisant partie du film.



L'arbre de dépendance syntaxique de surface qui correspond à la phrase (166) est montré dans la Figure 51. Comme pour les autres cas de relatives, le transféréme « d- » du pronom relatif gouverne toute la relative, et la partie anaphorème est gouvernée par la préposition *von*, qui modifie 'Musik'.



### 4.2.3 Topologie et relatives

A la base de l'analyse des propositions relatives dans les grammaires de dépendance, nous allons maintenant introduire les règles (très simples) qui permettant de placer les relatives suivant le bon ordre des mots.

En fait, la seule condition supplémentaire que nous devons ajouter à notre grammaire topologique est le fait que la partie transféréme et la partie anaphorème du pronom relatif doivent se retrouver dans un seul mot à la fin de la dérivation. La transféréme (la partie « complémentante » du pronom relatif) ouvre un domaine complétif et se place dans le champ du complémenteur. Cependant, comme cette fois-ci il ne s'agit pas d'un mot qui ouvre un domaine dans le champ du complémenteur, ce dernier reste accessible pour un (et seul) autre élément. on permet l'entrée aux parties anaphorèmes du pronom relatif, aux prépositions et aux verbes. Exactement de la même manière que pour les complétives, le dépendant verbal du transféréme droit se placer dans la parenthèse droite. Les dépendants du verbe se placent de la même manière qu'avant : ils peuvent occuper les champs majeurs ou les parenthèses (champ du complémenteur et complexe verbal), et ce en fonction des restrictions qui contraignent leur positionnement.

Dans l'exemple (167), la racine de la relative, le transféréme *d-* ouvre un domaine complétif et se place dans le champ du complétif, sans y créer un domaine et son dépendant verbal *schreibt* se place dans la parenthèse droite. Nous présentons dans la Figure 231'arbre de dépendance

syntactique de surface et dans la Figure 53, la structure topologique correspondante à cette relative.

(167) ... der einen Roman schreibt.

... qui un roman écrit.  
 '... qui écrit un roman.'

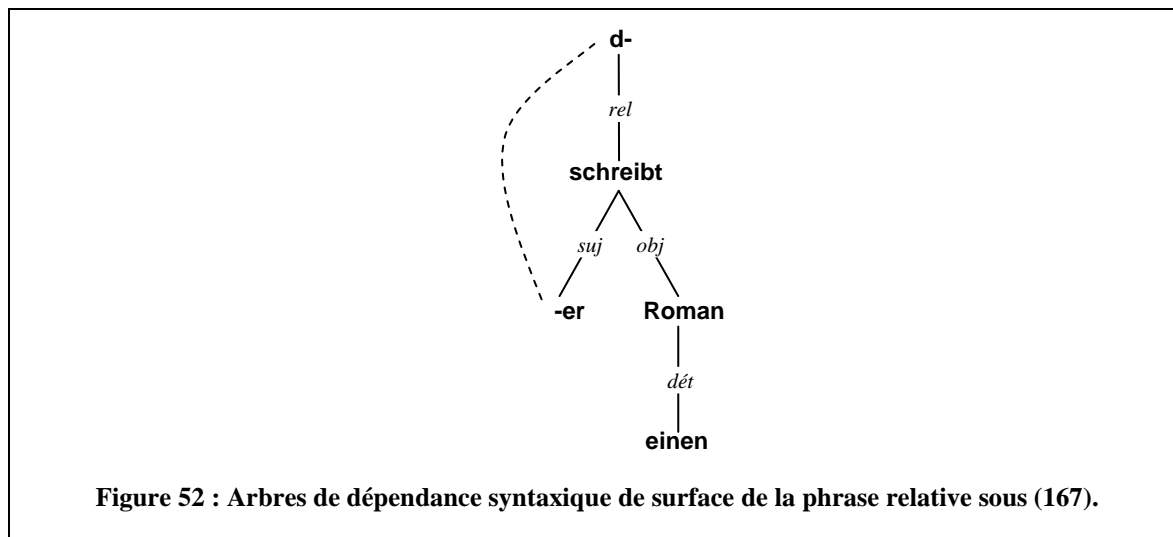


Figure 52 : Arbres de dépendance syntaxique de surface de la phrase relative sous (167).

Komplementierfeld	Mittelfeld	)	Nf
d-		schreibt	...

Figure 53 : Représentation de la structure topologique d'une relative : placement du transféréme et de son dépendant verbal.

Le dépendant nominal *den Roman* 'le roman' se place dans un champ majeur (et ouvre un domaine nominal ; la discussion sur la structure du domaine nominal sera le sujet de la section IV.6). L'autre dépendant nominal, l'anaphorème *-er* pourrait aussi se placer dans un des champ majeur mais, afin de rejoindre sa partie transféréme, il doit se placer dans le champ du complément. La structure topologique résultante, qui ressemble beaucoup à celle d'une complétive, est présentée dans la Figure 54 :

Komplementierfeld	Mittelfeld	)	Nf
d- -er	den roman	schreibt	...

Figure 54 : Représentation de la structure topologique d'une relative : placement des arguments nominaux<sup>83</sup>.

De la même manière, la structure topologique permet de régler facilement le cas du pied piping prépositionnel. Reprenons l'exemple (163) que nous avons explicité dans les sous-sections précédentes. Nous reprenons ci-dessous l'exemple ainsi que l'arbre de dépendance syntaxique de surface (uniquement la partie de la relative) :

<sup>83</sup> Nous ne présentons pas ici la structure interne des syntagmes nominaux.

- (168) ... auf dem Peter singt.  
 ... sur lequel Pierre chante.  
 '... où chante Pierre.'

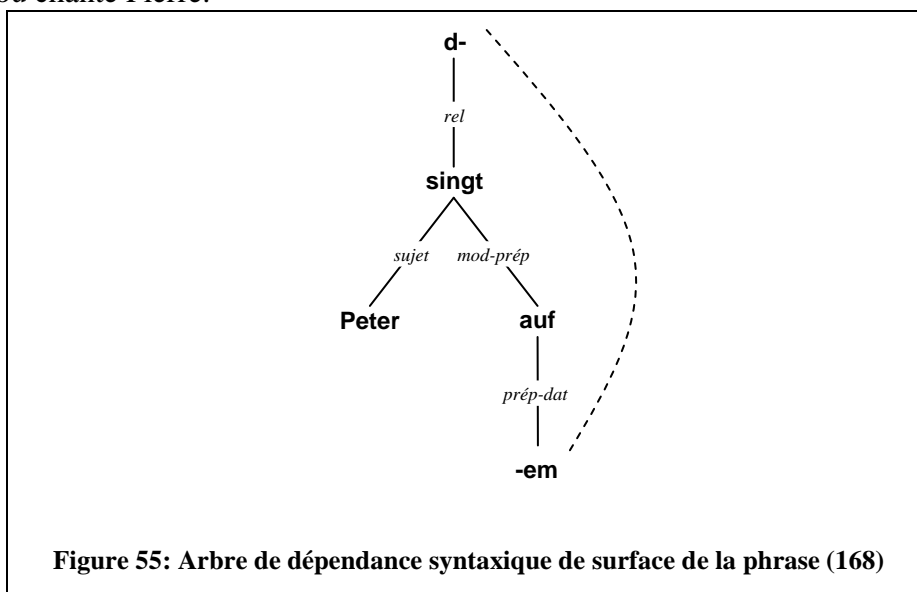


Figure 55: Arbre de dépendance syntaxique de surface de la phrase (168)

Une fois que le transféréme et son dépendant verbal ont été placés, le dépendant prépositionnel peut se placer soit dans le champ du complémenteur, soit dans le Mittelfeld, soit dans le Nachfeld pour y ouvrir un domaine nominal. Mais ce n'est que dans le champ du complémenteur que son dépendant, l'anaphorème *-em* pourra retrouver sa partie transféréme car l'anaphorème ne peut pas s'émanciper du domaine nominal.

<i>Komplementiererfeld</i>	<i>Mittelfeld</i>	)	<i>Nf</i>
auf d- -em	Peter	singt	...

Figure 56: Représentation de la structure topologique de la relative dans le cas du pied piping.

Pour expliquer topologiquement le cas de la phrase (169) (la même que (166)), il faut savoir que quelques syntagmes prépositionnels peuvent sortir du domaine nominal de leur gouverneur, comme d'ailleurs c'est le cas en français. Les mêmes conditions permettant de construire une relative avec pied piping prépositionnel (171) doivent être remplies pour qu'on puisse placer un constituant prépositionnel qui dépend d'un nom dans le Vorfeld comme nous le montrons dans la phrase (170).

- (169) Peter erwähnte den Film, von dem ich nur die Musik kenne.  
 Pierre mentionnait le film, de qui je seulement la musique connais  
 'Pierre a mentionné le film dont je ne connais que la musique.'
- (170) Über Syntax habe ich noch kein Buch gelesen.  
 sur syntaxe ai je encore aucun livre lu  
 'Sur la syntaxe, je n'ai lu encore aucun livre.'
- (171) ein Thema, über das ich noch kein Buch gelesen habe.  
 un thème, sur lequel je encore aucun livre lu ai  
 'Un thème sur lequel je n'ai encore lu aucun livre.'

Kuthy 1998 discute quelles sont les conditions (plutôt de nature sémantique) sous lesquelles un syntagme prépositionnel peut sortir d'un domaine nominal. Il discute également pourquoi la topicalisation en (172) par exemple (et de la même manière le pied piping en (173)) est exclue :

(172) ??Über Syntax habe ich noch kein Buch geklaut.

sur syntaxe ai je encore aucun livre volé  
 '??Sur la syntaxe, je n'ai encore jamais volé un livre.'

(173) ??ein Thema, über das ich noch kein Buch geklaut habe.

un sujet, sur lequel je encore aucun livre volé ai  
 '??Un sujet sur lequel je n'ai encore jamais volé un livre.'

(174) Ich habe noch kein Buch über Syntax geklaut.

je ai encore aucun livre, sur syntaxe volé  
 'Je n'ai jamais volé encore de livres sur la syntaxe.'

Sans rentrer dans les détails des raisons que Kuthy 1998 donnent pour l'agrammaticalité de ces phrases, nous pensons que cette différenciation nécessaire entre des prépositions qui sont « libres de sortir » et celles qui sont « assignées » au domaine de leur gouverneur doit être calculé aux niveaux adéquats (syntaxiques et sémantiques) et donner comme résultat un marquage nécessaire sur la relation syntaxique de surface qu'entretiennent le nom et la préposition. Si cette relation permet l'extraction, alors nous pouvons construire une relative avec pied piping. Dans le cas de la phrase (169), qui a parallèle à l'exemple donné sous (171), la linéarisation topologique procède de la même manière, et donc avant le placement de la préposition, nous avons la situation suivante (Figure 57) :

<i>Komplementiererfeld</i>	<i>Mittelfeld</i>	)	<i>Nf</i>
d-	Ich die Musik	kenne	...

Figure 57: Première étape dans la linéarisation topologique de la partie relative de la phrase (169).

Puis la préposition peut se placer dans le champ du complémenteur ou dans un champ majeur, mais seul son placement dans le champ du complémenteur permet à son dépendant l'anaphorème de rejoindre la partie transférée. La préposition n'a donc en quelque sorte pas de choix que de se placer dans ce champ (Figure 58).

<i>Komplementiererfeld</i>	<i>Mittelfeld</i>	)	<i>Nf</i>
von d- -em	Ich die Musik	kenne	...

Figure 58: Placement de la préposition dans le champ complétif pour la linéarisation topologique de la partie relative de la phrase (169).

Sans être amené à faire un quelconque ajustement, nous pouvons analyser le cas de « pied piping verbal ». La restriction d'accès pour le champ complétif que nous avons posée exclue les noms

de s'y placer, mais les verbes peuvent y accéder et y ouvrir un domaine. Si un infinitif avec *zu* ne prend pas la place de la parenthèse droite qui est proposée par son gouverneur et cherche à créer un nouveau domaine ailleurs, il peut aussi bien se placer dans le champ complétif sous condition qu'à la fin, le transfère qui y est déjà présent puisse garder sa place et que la partie complémentante l'anaphorème puisse rejoindre son transfère.

C'est ainsi que nous arrivons d'une manière naturelle à la linéarisation topologique sans pied piping (Figure 60) et avec pied piping (Figure 62) correspondant aux phrases (175) et (176) dont nous donnons l'arbre syntaxique de surface dans la Figure 59 :

(175) Der Roman, den zu lesen niemand dem Mann versprochen hat.

le roman, que de lire personne à\_le homme promis a  
 'le roman que personne n'a promis de lire à cet homme'

(176) Der Roman, den niemand dem Mann zu lesen versprochen hat.

le roman, que personne à\_le homme de lire promis a  
 'le roman que personne n'a promis de lire à cet homme'

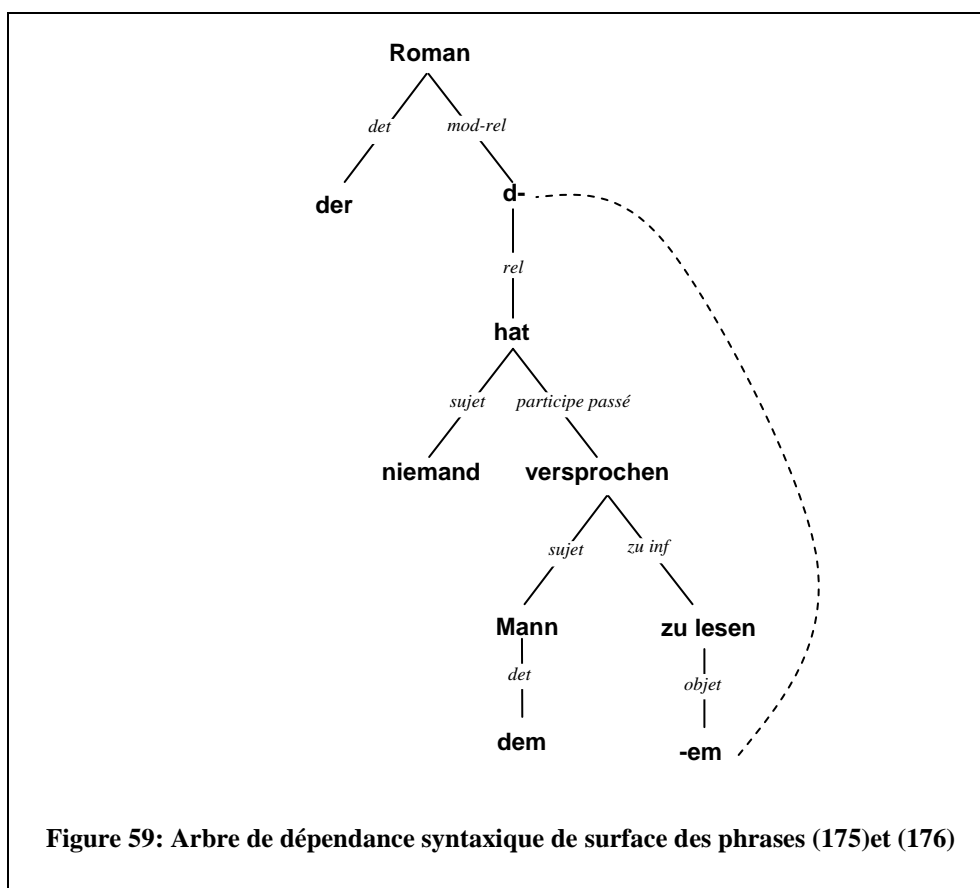


Figure 59: Arbres de dépendance syntaxique de surface des phrases (175) et (176)

L'arbre syntaxique de surface que nous avons présenté dans la Figure 59 aussi une autre linéarisation topologique pour le même ordre des mots que les phrases (175) et (176). En effet, l'infinitif en *zu lesen* 'de lire' peut aussi s'émanciper dans le Mittelfeld du domaine de son gouverneur : ce phénomène s'appelle aussi intraposition. On peut voir qu'il s'agit en fait d'une

Komplementiererfeld		Mittelfeld	)	Nf
d-en	zu lesen	niemand dem Mann	versprochen hat	...

Figure 60: Linéarisation topologique sans pied piping (phrase (175)).

construction différente que le contraste des phrases (175) et (176), les cas avec intraposition étant communicativement plus complexes. Nous présentons les structures topologiques dans les figures suivantes.

<i>Kf</i>	<i>Mittelfeld</i>	)			<i>Nf</i>
d-en	niemand dem Mann	Zu lesen	versprochen	hat	...

Figure 62: Linéarisation topologique avec pied piping (phrase (176)).

<i>Kf</i>	<i>Mittelfeld</i>		)		<i>Nf</i>
d-en	zu lesen	niemand dem Mann	versprochen	hat	...

Figure 61: Linéarisation topologique avec intraposition phrase (175)).

<i>Kf</i>	<i>Mittelfeld</i>			)		<i>Nf</i>
d-en	niemand	Zu lesen	dem Mann	versprochen	hat	...

Figure 63: Linéarisation topologique avec pied piping (phrase (176)).

Il est possible que les champs majeurs soient occupés par d'autres dépendants verbaux et non verbaux du verbe qui a ouvert le domaine dans le champ du complémenteur. Nous donnons des exemples sous (177).

(177) a. Alles dies sind Fragmente des gigantischen postsowjetischen Syndroms, [[von dem sich zu kurieren] [bislang] [nur sehr wenigen] [gelungen ist]]<sup>84</sup>.

tout ceci sont fragment de\_le gigantesque postsoviétique syndrome, [[de quoi se de guérir] [encore] {seulement très peu} [réussi est]  
 ‘ Tout cela sont des fragments du syndrome post-soviétique gigantesque, duquel seulement peu s’ont arrivés à se guérir.’

<sup>84</sup> journal Taz Nr. 6636, 28.12.2001, page 4, article *Die Beerdigung ist nicht zu Ende* par Michail Ryklin, Berlin.



- b. Aber auch im Allgemeinen würden sich an unsere Darlegung noch weitere Fragen anschließen, [[auf die weiter einzugehen] [der mir gewährte Raum nicht gestattet]]<sup>85</sup>.  
 mais aussi en\_général voudrait se à notre description encore d'autres questions  
 connecté, [[sur qui plus\_loin entre] [le à\_moi accordé place pas permis]]  
 'Mais aussi en général, d'autres questions s'ajoutent à notre description dont l'espace  
 qui m'est accordé ne me permet pas d'entrer dans les détails.'

#### 4.2.4 Une place devant le Vorfeld ?

Certaines analyses topologiques de l'allemand (entre autres Höhle 1986) posent l'existence d'un champ entre la place des connecteurs de phrase qui sont très restreints (seuls les connecteurs comme *und* 'et', *denn* 'car', *aber* 'mais' etc.) et la place du Vorfeld. Cette place est appelée le *Vorvorfeld* 'devant le Vorfeld' ou encore le *Außenfeld* 'champ externe' et est occupée par des éléments qui ressemblent à des éléments disloqués à gauche du français)<sup>86</sup>. Pourtant, cette position est beaucoup plus restreinte que ce qui se passe en français : en allemand, l'élément « disloqué » doit toujours être repris par un pronom résomptif dans le champ Vorfeld. Ce pronom peut toujours apparaître sans la spécification de l'élément disloqué. L'accord de ce pronom avec son antécédent est de nature anaphorique et beaucoup moins strict que pour les pronoms relatifs (179). Dans l'exemple donné par Altmann 1981 (que nous présentons ci-dessous sous (178)), le pronom ne fait pas l'accord en genre.

(178) Freundschaft, das ist wie Heimat.

amitié (fém), cela (neutre) est comme patrie  
 'L'amitié, c'est comme la patrie.'

(179) Eine Freundschaft, die (\*das) das wie Heimat ist.

amitié (fém), que (fém) comme patrie est  
 'L'amitié qui est comme la patrie.'

Si le pronom n'est pas repris, alors nous avons affaire à deux phrases, la première étant clairement une interrogation. et clairement distincte d'une point de vue prosodique de la deuxième.

(180) a. Meinem Vater, dem schenkt niemand Blumen.

à\_mon père, lui offre personne fleurs  
 'A mon père, personne ne lui offre de fleurs.'

b. \*Meinem Vater, niemand schenkt (dem) Blumen.

à\_mon père, personne offre (lui) fleurs

c. Meinem Vater ? niemand schenkt (dem) Blumen.

à\_mon père, personne offre (lui) fleurs  
 'A mon père ? personne ne lui offre de fleurs.'

d. dem schenkt niemand Blumen.

lui offre personne fleurs  
 'Personne ne lui offre de fleurs.'

<sup>85</sup> Ce sont les remarques concluantes du fameux article de Behaghel 1909:142 sur la « loi des segments augmentants (*Gesetz von den Gliedern mit steigender Schwere*).

<sup>86</sup> la notion de *dislocation gauche* est un autre terme qui provient de Ross 1967.

De plus, comme d'ailleurs en anglais, la dislocation ne peut avoir lieu que dans une phrase principale.

(181) a. \*Ich bin sicher, dass meinem Vater dem niemand Blumen schenkt.

je suis sûr, que mon père lui personne fleurs offre

b. Ich bin sicher, dass meinem Vater niemand Blumen schenkt.

je suis sûr, que mon père personne fleurs offre

'Je suis sûr que personne n'offre de fleurs à mon père.'

Hirschbühler 1975 donne un exemple pour le français.

(182) a. Je peux prouver que la chasse à l'étudiant, la police a toujours considérée cette activité comme un sport agréable.

b. \*Ich kann beweisen, dass die Studentenjagd (,die) Polizei immer diese Aktivität als angenehmen Sport angesehen hat.

je peux prouver, que la chasse\_à\_l'étudiant (la) police toujours cette activité comme agréable sport considérer a

c. Ich kann beweisen, dass die Polizei die Studentenjagd immer diese Aktivität als angenehmen Sport angesehen hat.

je peux prouver, que (la) police la chasse\_à\_l'étudiant toujours cette activité comme agréable sport considérer

Si l'élément disloqué est une complétive, il est repris par le pronom qui peut la remplacer.

(183) Dass er regnet, damit rechnet keine<sup>87</sup>.

que il pleut, avec\_cela calculer personne

'Qu'il pleuve, personne ne peut le prévoir.'

Pourtant, l'ordre inverse entre complétive et pronom résomptif est également possible :

(184) Damit, dass er regnet, rechnet keine.

avec\_cela que il pleut calculer personne

'Qu'il pleuve, personne ne peut le prévoir.'

Dans le cas de la phrase (184), il semble clair qu'on a affaire à un seul constituant dans le Vorfeld. Nous ne voyons pas d'inconvénient de considérer que le Vorfeld est dans les deux cas occupé par un seul constituant. Autrement dit, un tel pronom résomptif qui se trouve dans le Vorfeld ouvre une place pour son dépendant devant et derrière lui. Si ce dernier ne se trouve pas dans le Vorfeld mais ailleurs comme par exemple le Mittelfeld, il n'ouvre une place que derrière lui pour ses dépendants<sup>88</sup>. Cette analyse est complètement compatible avec notre système de

---

<sup>87</sup> Exemple de Hohler 1986.

<sup>88</sup> Une légère complication qui semble être en faveur de champ de Vorvorfeld à part repose sur la position des infinitifs phrastiques repris par un pronom (exemple (i)). En effet contrairement aux complétives, de telles phrases doivent toujours se placer devant le pronom résomptif.

(l'exemple est tiré du journal taz Nr.6594 du 7.11.2001, page 16, 186 Interview de Farin Urlaub

(i) Zu sagen, dass man keine simple Liebeslyrik schreiben darf, wenn man erwachsen sein will, das finde ich unfair.

de dire, que on aucun simple lyric\_d'\_amour écrire a\_droit quand on adulte être veut, cela trouve je injuste

'De dire qu'on n'a pas le droit des poèmes d'amour simples quand on veut être adulte, je trouve cela injuste.'

\* Das, zu sagen, dass man keine simple Liebeslyrik schreiben darf, wenn man erwachsen sein will, finde ich unfair.

cela, de dire, que on aucun simple lyric\_d'\_amour écrire a\_droit quand on adulte être veut, trouve je injuste

règles. Cependant, nous laissons ouvert la possibilité qu'un champ de Vorvorfeld (très restreint d'accès) puisse exister dans le domaine principal, et ce suivant des analyses plus fines, d'ordre syntaxiques et prosodiques.

## 5 Sur les traces du modèle topologique

Comme nous l'avons indiqué dans la partie introductive de ce chapitre, notre approche se base sur le modèle classique introduit par Drach 1937 et Bech 1955 pour une description de l'allemand. Une telle approche a également été proposée pour le traitement de l'ordre des mots de langues non germaniques, notamment l'ancien français (Skårup 1975) et le Warlpiri (Donohue & Sag 1999). Le modèle topologique a également été implémenté dans la grammaire HPSG (Kathol 1995, 2000) ainsi que dans les grammaires de dépendance (Debusmann and Duchier 2001, Gerdes and Kahane 2001).

En fait, l'idée de place fixe est apparue très tôt dans la description de l'allemand : on peut la retracer au moins jusqu'aux travaux de Herling 1821 qui construisait la première théorie globale sur la structure hiérarchique des phrases complexes. En 1886, Erdmann proposait des listes de places de la phrase en énumérant le type d'éléments que ces places pouvaient contenir, sans pour autant donner des noms spécifiques à ces places. A cause de la complémentarité de ces deux travaux pionniers de ce qu'on appelle aujourd'hui *la théorie des champs topologiques*, Höhle 1986 va même à proposer d'appeler cette théorie le système de Herling/Erdmann. Le terme de « champ » pour désigner ces places dans la phrase apparaît pour la première fois chez Drach 1937 dans un livre intitulé *idées fondamentales de la science de la phrase allemande*<sup>89</sup> destiné aux enseignants de l'allemand comme langue maternelle et langue étrangère.

Le livre de Drach prône, avec une teinte légèrement nationalisante, une émancipation de la grammaire, qui était jusqu'à là sous une influence latine forte, en faveur d'une « construction d'une présentation et d'un système de règles basé sur la nature de la langue allemande »<sup>90</sup> (Drach 1937:6). Ce qui est particulièrement intéressant dans cet ouvrage, c'est la liaison que l'auteur fait entre la structure de la phrase séparée en champs et la phonétique<sup>91</sup>. Dans une tradition explicite de Bally et de Taine, Drach énonce : « nous avons besoin d'une science du son de la phrase. La constatation doit s'imposer que les moyens biologiques d'expressions de la langue vivante parlée – l'accent, l'intonation, les pauses – sont partie essentielle de l'intégralité de la langue »<sup>92</sup>. (Drach 1937:8).

Bech 1955 a adapté par la suite la terminologie de Drach pour la construction de la structure interne du complexe verbal. Griesbach 1960 fait la connexion explicite entre la terminologie de Drach et le système de Herling/Erdmann, encore avec des noms légèrement différents. Enfin, ce n'est qu'avec Engel que la terminologie telle qu'elle est utilisée aujourd'hui sera (presque) définitivement établie. Höhler 1986 reviendra sur cette terminologie en remplaçant le noms des champs par des variables abstraites comme X et Y et en introduisant l'idée de vorvorfeld (KL). Nous avons essayé de faire un récapitulatif de la terminologie employée par les différents auteurs, afin de montrer les similarités et les différences. Nous donnons dans le tableau la

<sup>89</sup> *Satzlehre* 'science de la phrase' est le terme allemand, bien que peu usité, de *syntax* 'syntaxe'.

<sup>90</sup> texte original : « Aufbau einer im Wesen der deutschen Sprache begründeten Darstellung und Regelfassung. » (ma propre traduction)

<sup>91</sup> ce que Drach n'appelle bien entendu pas phonétique (*Phonetik*) mais *Schall-lehre* (science du son)

<sup>92</sup> Texte original : « Notwendig ist eine Schall-Lehre des Satzes. Die Erkenntnis muss sich durchsetzen, dass die biologischen Ausdrucksmittel lebendigen Sprechens – die Betonung, die Intonation, die Einschnitte – als notwendiger Bestand mit zum Sprachganzen gehören. »

terminologie que nous adoptons sans rentrer dans les détails. Cette terminologie sera reprise et expliquée dans la section suivante.

notre terminologie	Champ de coordination	Vorfeld	[	Mittelfeld	]	Nachfeld	
Avant Drach	?	Position devant la copule, position du sujet ; chez Erdmann : dit explicitement que ce n'est pas la position du sujet		Position de copule		prédicat	
Drach (seulement pour proposition finie)	?	Vorfeld		Milieu/ parenthèse		Nachfeld parenthèse	
Erben 1954		Vorfeld		Verbe fini	Mittelfeld	complexe verbal obligatoirement rempli Nachfeld	
Griesbach 1960		Vorfeld		?	Satzfeld	?	
Engel 1970		Vorfeld		Parenthèse gauche FINIT	Mittelfeld	Parenthèse droite Complexe verbal	Mittelfeld
Höhle 1986	KOORD	K L		X		Y	
Kathol 1995	?	Vorfeld (vf)		champ du complémenteur (cf)	Mittelfeld (mf)	Complexe verbal (vc) Nachfeld (nf)	

Par ailleurs, l'idée de place fixe dans la phrase est aussi apparue dans des travaux sur des langues non germaniques ; par exemples Skårup 1975 utilise un tel modèle pour la description de l'ancien français<sup>93</sup>, Donohue & Sag 1999 ont proposé une analyse basée sur l'idée de place pour le Warlpiri, et Cori et Marandin 1993 et Marandin 2001 introduisent l'idée de « site » dans l'analyse du français.

D'autres descriptions où le terme de « place » ou « champ » inclut bien sûr les travaux sur des langues qui ont des phénomènes d'ordres des mots voisins à celui du verbe second de l'allemand : par exemple, on peut citer le verbe second en Kashmiri (Hook 1985) ou le clitique second du serbo-croate (Milićević 2001). A nos connaissances, le travail de Milićević est

<sup>93</sup> D'après les informations qui l'ont été communiquées personnellement par Sophie Prévost, l'ancien français est une langue avec un verbe second qui est dans un sens beaucoup moins stricte que l'allemand. En effet, contrairement à l'allemand, la dislocation gauche en ancien français n'est pas restreint (comme d'ailleurs en français moderne) et seule la place devant le verbe, comme le Vorfeld de l'allemand ; par contre, la place comparable à celle du Vorfeld est restreinte car elle doit être remplie par un seul de ses constituants. En français moderne, la place devant le verbe est généralement occupé par le sujet sauf dans les cas d'inversion, ce qui est lexicalement et stylistiquement très marqué. Dans une analyse topologique de l'ancien français, on pourrait proposer un Vorvorfeld qui n'est pas rempli d'une manière obligatoire (pour la dislocation) et un Vorfeld, nécessairement rempli par un constituant de fonction syntaxique quelconque.

d'ailleurs le seul travail qui essaie de modéliser des « places » dans la Théorie Sens-Texte. Elle définit, de la même manière que nous le faisons, une structure de constituants au niveau de la représentation morphologique profonde. Cependant, elle considère que les constituants qu'elle propose servent à « représenter la façon par laquelle une langue exprime linéairement les relations syntaxiques »<sup>94</sup> (Milićević 2001:11). Ces constituants doivent correspondre à un sous-arbre de l'arbre de dépendance (la définition est donnée dans Milićević 2001:11), ce qui correspond à des critères de projectivité légèrement plus stricts que ceux que nous proposons dans notre approche. Il serait intéressant de vérifier en détail la compatibilité de ses règles avec un modèle topologique.

Enfin, le travail sur la topologie le plus proche et le plus influent sur notre travail est sans doute la description de l'allemand en HPSG (Pollard & Sag 1994) proposée par Kathol 1995. Kathol prône une syntaxe orientée vers la linéarisation et introduit un nouveau trait dans la théorie HPSG, le trait TOPO, dans lequel les éléments sont placés à l'aide de règles de précédence l'un par rapport à l'autre plus ou moins indépendamment de la structure de constituant à laquelle est attaché ce trait. Il ajoute donc un outil de linéarisation en sus de la structure de constituants habituelle de HPSG. Ainsi, une des prémisses de son travail est que « le composante linéaire d'une langue comme l'allemand a d'une certaine manière une existence qui dépend de distinctions catégoriques des notions de constituance... Dans aucun cas, cela implique qu'on peut se débarrasser de la notion de constituants en soi »<sup>95</sup> (Kathol 1995 :25).

La structure de constituants n'a donc plus la tâche de représenter l'ordre des mots, elle reste toutefois la colonne vertébrale de la construction, car la combinaison habituelle des signes en HPSG se fait justement au niveau de la structure des constituants. Comme l'ordre n'est plus directement exprimé par la structure de constituants, Kathol combine les schémas habituels de dominance de HPSG (Pollard & Sag 1994), basés sur la distinction entre l'ordre respectif des arguments, le *schéma tête-sujet* et le *schéma tête-complément*, en un seul *schéma tête-argument*. La structure de constituants qui en résulte reste toujours binaire, et n'est donc pas équivalente à une structure de dépendance (plate) où tout argument dépend directement d'une tête.

Kathol 1995 ne construit jamais une véritable structure topologique, et le formalisme HPSG lui permet de garder l'ordre topologique des éléments comme des ajouts à la structure des constituants. Ce que Kathol appelle « domaine » (d'ordre) ne renvoie pas à l'idée d'un constituant proposant des places fixes. Il s'agit d'un trait du nom « *DOM* » sous lequel tout participant à la phrase envoie le contenu de leur trait *DOM*, pour y être ordonné. Kathol doit proposer une règle générale de précédence linéaire, valable pour toute combinaison d'éléments sous le trait *DOM*. Il donne ainsi à la place du V2 le nom de « champ de complémenteur ». Sous *DOM*, on trouve donc l'information *TOPO*, des fois non résolu entre deux champs, par exemple *verbal*, qui entraîne *cf* (champ de complémenteur) ou *vc* (complexe verbal). Par ailleurs, les champs sont des classes d'équivalence du trait *TOPO*, qui ne peuvent pas porter eux-mêmes des restrictions de placements. Ainsi, si l'on veut exprimer qu'un champ ne peut prendre un seul élément, on indique par une règle de précédence linéaire que tout élément sous un trait *DOM*, portant le trait *TOPO vf*, doit précéder tout autre élément ayant la même valeur pour ce trait (*TOPO vf* < *TOPO vf*), ce qui ne peut être satisfait que par un seul élément. De même, le formalisme ne permet pas d'exprimer directement si un champ est obligatoirement rempli. C'est ainsi que Kathol est amené à ajouter la « contrainte de phrase » (« clause constraint », Kathol

---

<sup>94</sup> texte original : « to represent the way a language linearly expresses syntactic relations », ma propre traduction.

<sup>95</sup> Texte original : « One of the premises of the current work ... is that the linear component of a language like German to some extent has an existence that is independent of categorical distinctions or notions of constituency. ... By no means does this entail that the notion of constituency itself will be dispensed with ». Ma traduction

1995:167), qui dit que dans une phrase complète de l'allemand, le champ topologique du complémenteur (*cf*) doit être instantié.

Le lien entre syntaxe de surface et structure topologique pourrait donc être extrait de la structure HPSG proposée par Kathol : la structure X-barre ; démunie de l'information sur l'ordre des mots, représente essentiellement la dépendance syntaxique (la tête gouverne les autres éléments d'un constituant) et les différentes structures topologiques plates que propose Kathol pourraient être analysées dans une structure hiérarchique similaire à la notre, en se basant sur les relations de dépendances syntaxiques sous-jacentes. Dans ce sens, on pourrait dire Kathol construit déjà une structure topologique sans jamais l'explicitier.

On peut noter que malgré les différences importantes de formalisation, la plupart des analyses proposées par Kathol ressemblent aux nôtres. C'est principalement au sujet de la parenthèse droite qu'apparaissent des différences, justement parce que ce dernier ne construit pas de structure topologique. En effet, dans notre propre analyse nous construisons une structure topologique interne au complexe verbal tandis que Kathol règle la question du complexe verbal à l'aide de règles de précedence « plates », c'est-à-dire non structurées d'une manière interne. Nous ne rentrerons pas dans le détail de la différence, ceci nous obligeant à discuter en détail le formalisme HPSG.

## 6 La notion de syntagme topologique

Nous avons vu jusqu'à présent que une analyse topologique permettait de traiter d'une manière simple et cohérente les phénomènes d'ordre de mots de l'allemand, souvent considérés comme très complexes. Dans cette section, nous essayerons de caractériser et justifier la notion de syntagme topologique. Le but de montrer que ce constituant ne représente pas juste un outil qui s'adapte bien dans le mécanisme topologique mais qui représente une entité réelle, présente dans la conscience du locuteur, correspondant à un groupement communicatif et prosodique. Nous présentons donc une série de critères (syntaxiques et prosodiques) permettant de justifier cette notion puis nous discutons la place du syntagme topologique dans un modèle linguistique.

### 6.1 Critères syntaxiques

Dans le cadre des grammaires syntagmatiques, beaucoup de tests ont été proposé pour valider la notion de syntagme (voir les précurseurs Bloomfield 1933, Wells 1947 et Harris 1951 ; pour un survol et un critique dans les travaux récents, voir Abeillé 1991, 2002 et Bonami 1999). Ces tests permettent de déterminer les constituants syntaxiques, une notion qui diffère de celle de syntagme topologique que nous défendons dans ce présent travail. Néanmoins, il est intéressant et instructif de revoir les différents tests proposés. Les trois tests principaux sont les tests d'énoncé, de commutation et de position :

- *test d'énoncé* : un groupe de mots est considéré comme un constituant syntaxique potentiel s'il peut constituer un énoncé, par exemple lors d'une réponse à une question (185):

(185) Qu'est-ce que Pierre essaie de faire ? –**de lire le livre**

- *test de commutation* : un groupe de mot est un constituant syntaxique potentiel si il peut communiquer avec un mot unique, par exemple un pronom (186) :

(186) Pierre essaie **de lire le livre** → Pierre **l'**essaie.

- *test de position* : un groupe de mot est un constituant syntaxique potentiel s'il peut apparaître dans une position syntaxique qui ne peut être occupée que par un seul constituant. La topicalisation (187) et le clivage (188) constituent de bons tests de position.

(187) **Lire le livre**, Pierre essaie de le faire.

(188) C'est **lire le livre** que Pierre essaie de faire.

Pour l'allemand, l'occupation de la place du Vorfeld peut donner lieu à un test similaire, comme on peut le voir dans l'exemple (189) mais ce test est affaibli à cause de la possibilité d'émancipation comme il apparaît dans l'exemple (190) : dans cet exemple, le groupe *zu lesen* 'de lire' occupe le Vorfeld mais n'est pas considéré comme un constituant syntaxique car son dépendant syntaxique *das Buch* 'le livre' se trouve ailleurs que le Vorfeld.

(189) Das Buch zu lesen versucht Peter.

Le livre de lire essaie Pierre.  
'Lire le livre, Pierre essaie de le faire.'

(190) **Zu lesen** versucht Peter Das Buch.

de lire essaie Pierre le livre  
'Pierre essaie de lire le livre.'

En fait, il s'agit là de tests qui peuvent être utiles dans le choix de la représentation syntaxique. La représentation syntaxique que nous adoptons repose sur l'arbre de dépendance syntaxique de surface. Or cet arbre est plus ou moins équivalent à une structure syntagmatique X-barre. Ainsi, les tests syntaxiques proposés pour la détermination de constituants syntaxiques peuvent être directement adaptés afin de déterminer les dépendances syntaxiques. Nous renvoyant à Mel'čuk 1988 pour des tests équivalents permettant de déterminer la dépendance.

Les syntagmes topologiques ne remplissent pas les critères présentés ci-dessus pour la détermination des constituants syntaxiques. Si l'on prend comme exemple un complexe verbal, il ne peut généralement pas être un énoncé complet, il ne peut pas être commuté avec un seul et il ne peut pas occuper une autre place que celle de la parenthèse droite d'un domaine (et d'une manière inverse, seul un complexe verbal peut occuper cette position). Et pourtant, la position d'un complexe verbal peut être occupé par un verbe unique et donc, d'une certaine manière, un complexe verbal se comporte comme un verbe unique : ainsi, les dépendants des différents verbes qui forment le complexe verbal se comportent comme s'ils étaient les dépendants d'un seul verbe, leur ordre relatif étant libre. Ainsi, dans l'exemple il est très difficile de remplacer un complexe verbal par un seul verbe car aucun verbe allemand ne sous-catégorise deux arguments au datif :

(191) ... weil dem Mann seinem Sohn niemand **zu helfen anbietet**.

...parce\_que à\_le homme à\_son fils personne de aider offre  
'parce que personne n'offre à l'homme d'aider son fils.'

Cependant, bien qu'un complexe verbal ne constitue pas un constituant syntaxique, la plupart des syntagmes topologiques coïncident avec des constituants syntaxiques, et ceci est vrai pour tous les domaines. Un domaine représente le syntagme topologique maximal qu'un mot peut ouvrir et correspond à sa projection maximale (la projection maximale d'un mot est la projection d'un sous arbre complet dont il est la tête dans l'arbre de dépendance syntaxique de surface). Cette dernière propriété est très importante car elle caractérise les contraintes mutuelles qui existent entre la représentation syntaxique et la structure topologique. Nous pensons que cette relation

étroite qui existe entre constituants syntaxiques et syntagmes topologiques peut expliquer pourquoi beaucoup de théories ne font pas la distinction entre ces deux entités. En effet, ces théories commencent par poser que chaque élément réalise une projection maximale et que par conséquent, on doit utiliser des processus complexes (comme le mouvement) lorsqu'on est confronté à une non congruence entre constituants syntaxiques et syntagmes topologiques. Dans notre approche, le domaine représente le plus grand constituant topologique qu'un mot peut ouvrir, ce qui ramène à dire que ce constituant peut accueillir tous les dépendants de ce mot, mais cela veut pas dire que l'accueil de ces éléments est obligatoire (en cas d'émancipation par exemple).

En fait, nous pouvons avoir recours à d'autres critères pour caractériser les syntagmes topologiques : il s'agit de critères prosodiques, souvent mis en avant pour déterminer les constituants syntaxiques (Chomsky & Halle 1968, Di Cristo 1981, Selkirk 1986 et Truckenbrodt 1995) bien que la non congruence entre prosodie et syntaxe a été et continue à être démontrée (voir une présentation générale dans Hirst & Di Cristo 1998).

## 6.2 Critères communicatifs et prosodiques

Une particularité de notre approche est qu'une phrase syntaxiquement non ambiguë peut avoir plusieurs structures topologiques. D'un point de vue des grammaires syntagmatiques, ces dernières sont considérées comme des ambiguïtés et par conséquent, ne sont pas acceptées (voir par exemple Müller 1999, section 17.5). Nous allons montrer que ces ambiguïtés structurelles sont justifiées d'un point de vue communicatif et prosodique.

Reprenons notre exemple de base sous (192) :

(192) Die Männer haben dieser Frau den Roman zu lesen versprochen.

les hommes ont à\_cette femme le roman de lire promis  
'Les hommes ont promis à cette femme de lire ce livre.'

La phrase (192) peut avoir trois structures topologiques différentes (elles ont bien sûr le même arbre de dépendance syntaxique de surface). Nous indiquons les groupements possibles sous (193) ainsi que la représentation des trois structures topologiques dans la Figure 64 :

- (193) a. [Die Männer] [haben] [dieser Frau] [den Roman] [zu lesen versprochen]  
b. [Die Männer] [haben] [dieser Frau] [den Roman zu lesen] [versprochen]  
c. [Die Männer] [haben] [dieser Frau] [den Roman] [zu lesen] [versprochen]

La première structure est une structure plate, où les verbes *versprochen* 'promis' et *zu lesen* 'de lire' se placent dans la parenthèse droite. La seconde structure illustre un cas d'intraposiion : le verbe *zu lesen* 'de lire' ouvre un domaine enchâssé dans le Mittelfeld. Enfin, la troisième structure illustre un cas plus complexe, avec l'intraposiion de *zu lesen* 'de lire' et l'émancipation de *den Roman* 'le roman', qui continue à être placé à côté de son gouverneur, mais dans un autre domaine, notamment dans le Mittelfeld du domaine principal.

Nous considérons que l'ambiguïté structurelle correspond à une ambiguïté sémantique, d'ordre communicatif. En effet, dans les trois structures, *die Männer* 'les hommes' se trouve dans le Vorfeld. Les phrases qui ont *die Männer* 'les hommes' dans la place du Vorfeld peuvent être des réponses naturelles à une question du type *Was ist mit den Männern* 'Qu'est ce qui se passe avec les hommes?'. Dans ce cas, on considère *die Männer* 'les hommes' comme le topique de la phrase. En fait, le constituant qui se trouve dans le Vorfeld peut soit être marqué comme un topique, soit comme un focus (voir la discussion sur la structure communicative dans le chapitre I.6.3 ; voir également Choi 1999, Gibbon 1998 et Büring 1997 pour une discussion plus détaillée). Dans les phrases sous (193), si on considère que *die Männer* 'les hommes'



représentent le topique de la phrase, dans la phrase a. le reste de la phrase est rhématique tandis que dans la phrase b. le fait de 'lire un livre' est marqué. Pour expliciter cet exemple, comparons la phrase (193)b à la phrase 0.

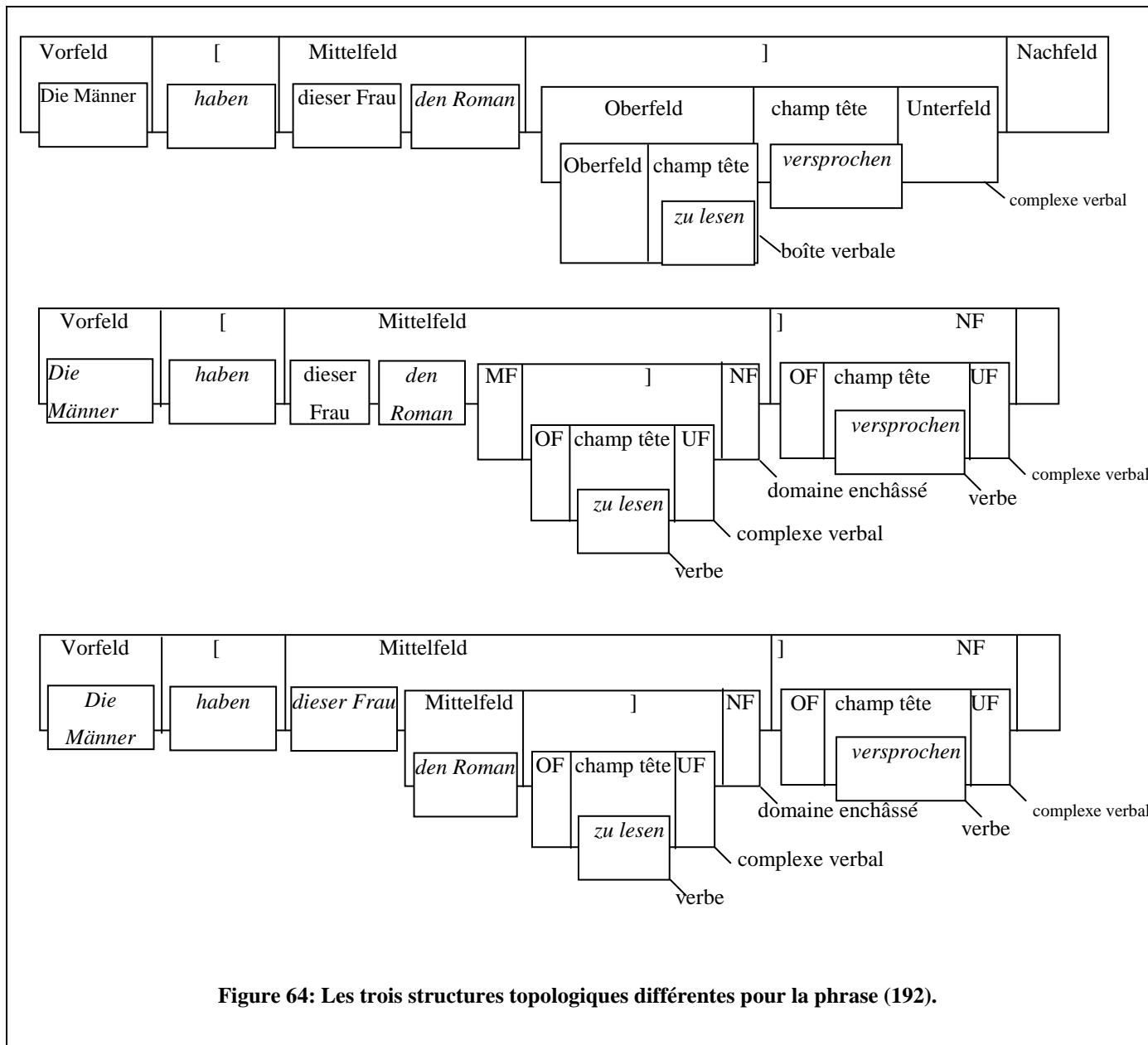


Figure 64: Les trois structures topologiques différentes pour la phrase (192).

(194) [Den Roman zu lesen] [haben] [dieser Frau] [die Männer] [versprochen].

Dans 0, le constituant qui se trouve dans le Vorfeld peut, comme pour les phrases sous (193), peut soit être thématique (correspondant grosso modo à 'De lire le roman, les hommes l'ont promis à cette femme') soit rhématique (correspondant à 'C'est lire le roman que les hommes ont promis à cette femme'). La phrase (193)b., est plus complexe : en effet, une telle phrase requiert un contexte où *den Roman zu lesen* 'le roman de lire' apparaît comme une unité communicative distincte, soit parce qu'elle contraste avec une autre action (par exemple *ihre Zimmer aufzuräumen* 'de nettoyer leur chambre') comme dans **Erreur ! Source du renvoi introuvable.**, soit parce que leur gouverneur *versprochen* 'promis' contraste avec un autre verbe de contrôle comme dans (196), défendant la formation d'un complexe verbal.

(195) [Die Männer] [haben] [dieser Frau] [den Roman zu lesen] [versprochen] und [die Kinder] [ihre Zimmer aufzuräumen]

les hommes ont à\_cette femme le roman de lire promis et les enfants leur chambre nettoyer

'Les hommes ont promis à cette femme de lire ce livre, et les enfants de nettoyer leur chambre.'

(196) [Die Männer] [haben] [dieser Frau] [den Roman zu lesen] [versprochen] und [die Kinder] [gedroht]

les hommes ont à\_cette femme le roman de lire promis et les enfants menacé

'Les hommes ont promis à cette femme de lire ce livre, et les enfants ont menacé de le faire.'

Enfin, la structure (193)c. est difficile à motiver communicativement. Le domaine enchâssé qui est ouvert par *zu lesen* 'de lire' ainsi que son complément émancipé *den Roman* 'le roman' ont été créés, mais la raison de cette position reste peu claire. En effet, si pour une raison communicative on avait cherché à indiquer qu'ils forment une entité, on aurait opté plus pour un placement dans le Vorfeld ou le Nachfeld, ou encore dans le Mittelfeld<sup>96</sup>. Par conséquent, nous pouvons prédire qu'une phrase avec une structure telle que (193)c. n'est que rarement prononcée, ce qui semble être le cas.

D'un point de vue prosodique, les syntagmes prosodiques semblent être étroitement liés à la structure prosodique de la phrase. Considérons les exemples suivants :

(197) a. [Die Männer] [haben] [dieser Frau] [den Roman] [zu lesen versprochen]

b. [Die Männer] [haben] [den Roman zu lesen] [dieser Frau] [versprochen]

c. [Dieser Frau] [haben] [die Männer] [den Roman zu lesen] [versprochen]

Dans le cas de nos exemples, le complexe verbal *zu lesen versprochen* 'de lire promis' occupe la parenthèse droite du domaine principal, comme on peut le voir dans la première configuration prosodique de la Figure 65<sup>97</sup>.

Ce complexe verbal se caractérise par un seul accent (sur la première syllabe du radical du premier verbe hors préfixe non accentué) et une chute régulière de la courbe mélodique. De la même façon, dans la phrase 100b. (la deuxième configuration prosodique de la Figure 65), la parenthèse droite du domaine principal est réduite à *versprochen* 'promis', qui porte maintenant un accent d'intensité sur la première syllabe du radical verbal (*spro*) et une chute régulière de la courbe mélodique. On observe également un domaine enchâssé *den Roman zu lesen* 'le roman de lire', initié par une descente de la fréquence fondamentale, et une accentuation sur la parenthèse droite, occupé par *zu lesen* 'de lire'.

<sup>96</sup> Les éléments isolés [den Roman] et [zu lesen] auraient pu rester côte à côte dans le Mittelfeld selon une structure communicative extrêmement complexe, où la position de Vorfeld n'est pas accessible ou occupée par des éléments qui sont communicativement plus proéminents. La phrase (i) a comme seule structure topologique une intraposition couplée d'une émancipation dans le Mittelfeld pendant que le Vorfeld et le Nachfeld semblent être accessibles, comme dans le cas de la structure topologique de la phrase (193)c. Le manque de contextes communicatifs pour la phrase (i) explique peut-être le fait qu'elle est souvent considérée comme agrammaticale. Nous préférons juger cette phrase tout comme (193)c. comme des phrases correctes, mais communicativement non motivées.

(i) <sup>97</sup>[Die Männer] [haben] [zu lesen] [dieser Frau] [den Roman] [versprochen]

<sup>97</sup> Dans les figures, sont présentées le contour de la fréquence fondamentale, le texte aligné, le contour d'intensité et le signal acoustique. Les phrases analysées ont été obtenues lors d'un enregistrement d'un corpus composé de différentes linéarisations que l'on voulait tester. Nous avons enregistré une locutrice native de l'allemand dit standard, qui habite à Cologne et qui n'a pas d'accent particulier.

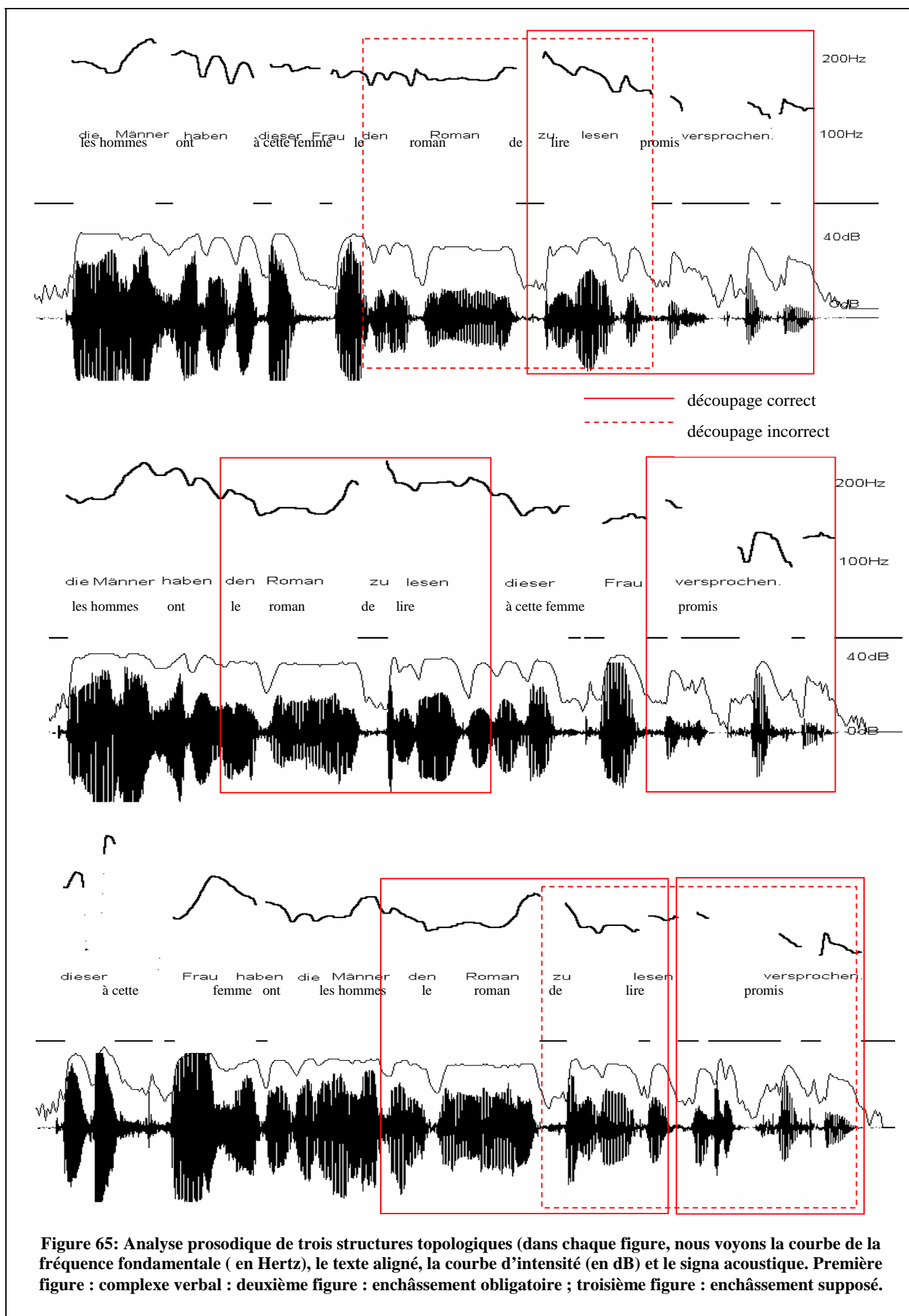


Figure 65: Analyse prosodique de trois structures topologiques (dans chaque figure, nous voyons la courbe de la fréquence fondamentale ( en Hertz), le texte aligné, la courbe d'intensité (en dB) et le signal acoustique. Première figure : complexe verbal ; deuxième figure : enchaînement obligatoire ; troisième figure : enchaînement supposé.

De plus, la courbe mélodique de *zu lesen* reste plate après l'ictus initial, ce qui contraste avec la configuration prosodique du complexe verbal (première configuration) dans lequel il forme un complexe verbal avec le verbe qui suit (*versprochen*).

La phrase 100c. (troisième configuration prosodique de la Figure 65) se termine comme la première sur *den Roman zu lesen versprochen* 'le roman de lire promis'. Pourtant, cette dernière s'apparente plutôt à la phrase 100b.: *den Roman zu lesen* possède exactement la même courbe mélodique que la phrase 100b., tandis que *versprochen* présente un accent d'intensité et une courbe mélodique descendante, ce qui nous fait dire qu'il occupe à lui seul la parenthèse droite du domaine principal.

Alors que les phrases 100a. et b., été obtenues par lecture sans consigne spéciale, la phrase 100c., est le résultat d'une consigne où *den Roman zu lesen* a été précédemment activé comme entité. Des résultats similaires ont pu être obtenus en oral spontané en réponse à des questions favorisant ce regroupement, telles que *Was haben die Männer dieser Frau versprochen? 'Qu'ont promis les hommes à cette femme?'*

Un autre cas d'ambiguïté structurelle peut être illustré par le contraste qui existe entre les infinitifs en *zu* et les infinitifs nus (sans *zu*) : en effet, les infinitifs nus ne peuvent pas créer de domaine enchâssé en dehors du Vorfeld. Par conséquent, il y a deux configurations prosodiques pour la phrase (selon qu'il y a ou pas détachement de *den Roman* 'le roman' de *zu lesen* 'de lire') tandis qu'un seul contour prosodique est permis pour la phrase, bien que ces deux phrases ont le même arbre de dépendance syntaxique de surface. Cette différence est aussi marquée à l'écrit, où une virgule est recommandée pour la phrase (où la structure avec enchâssement est préférée) tandis que la virgule est interdite dans le cas de la phrase :

(198) a. Niemand versucht(,) den Roman zu lesen.

Personne essaie le roman de lire  
'Personne n'essaie de lire le roman.'

b. Niemand will den Roman zu lesen.

Personne veut le roman de lire  
'Personne ne veut de lire le roman.'

Nous espérons avoir donné suffisamment d'arguments pour montrer que le syntagme topologique n'est pas une unité abstraite juste « utile » pour notre formalisation mais bel et bien une entité linguistique, motivée par des raisons syntaxiques, communicatives et linguistiques. Dans la section suivante, nous discutons plus généralement de la question de structure syntagmatique à l'intérieur d'un modèle linguistique.

## 7 La place de la structure syntagmatique dans un modèle linguistique

Une des choses que notre modèle met en valeur, est la nécessité de bien séparer deux types d'unités linguistiques, les constituants syntaxiques d'une part et les constituants prosodiques d'autre part. Les premiers sont des projections (syntaxiques) maximales de têtes lexicales, et sont directement liés à la notion de dépendance syntaxique. Un constituant syntaxique correspond donc à une chaîne ordonnée de nœuds dominés par un nœud « tête » dans l'arbre de dépendance syntaxique. Autrement dit, une dépendance syntaxique exprime la relation qui existe entre la tête lexicale d'un constituant et la tête lexicale d'un sous-constituant, laissant de côté l'ordre linéaire de ces deux éléments. Par contre, les syntagmes topologiques sont des groupes de mots linéairement ordonnés qui apparaissent naturellement dans la linéarisation de la phrase. Ils sont

motivés par des critères syntaxiques et communicatifs. Ils représentent des segments continus de la phrase et sont marqués d'un contour prosodique spécifique.

En fait, nous pensons que le deuxième type d'unité, c'est-à-dire les syntagmes topologiques, constituent réellement et légitimement des unités « syntagmatiques » selon la terminologie saussurienne.

En quelque sorte, la structure syntagmatique traditionnelle encode dans une même structure, deux types d'information, notamment une dépendance syntaxique et une structure de constituants topologiques, bien que ces derniers, soient deux notions bien distinctes.

Comme le dit bien Kathol 1995, « dans notre conception des relations syntagmatiques, il n'existe pas de connexion inhérente entre ce qui peut occuper un champ topologique et leur composition catégorielle. Dans d'autres termes, il n'est pas nécessaire que les éléments qui sont éligibles pour un champ topologique forment également une classe naturelle pour d'autres aspects. [...] De même, nous rejetons l'idée communément acceptée en GB que les positions sont intrinsèquement limitées à un seul niveau de projection »<sup>98</sup>. Jacobs poursuit dans la même direction : « Même si on accepte la supposition – en premier lieu contre-intuitive – que le verbe fini en première ou seconde position se place dans la position du complémenteur, dans le but d'exprimer de manière élégante les relations grammaticales multiples entre les verbes finis dans ces positions et la conjonction de subordination, [...] il n'est pas obligatoire de présumer que le verbe fini a la *fonction grammaticale* d'un complémenteur dans les phrases de ce type »<sup>99</sup>.

Ce n'est que dans les formes récentes du cadre théorique chomskyen qu'on considère que la description linguistique consiste à mettre en pair, des sens avec des textes (un niveau « conceptuel-intentionnel » et un niveau « articulatoire-perceptuel » dans les termes de Chomsky 1995). Suivant les suppositions des héritiers de la Théorie Standard Étendue (angl. *Extended Standard Theory*), on considère que « chaque langage détermine une série de paires  $(\pi, \lambda)$  [ $\pi$  pour la Forme Phonétique (*sic* !) et  $\lambda$  pour le Forme Logique] comme ses représentations formelles de sons et de sens, tant que ces derniers sont déterminés par le langage même » (Chomsky 1995:169)<sup>100</sup>.

Cette idée représente un point de convergence entre l'approche chomskyenne et l'approche de Mel'čuk. Cette mise en paire entre textes et sens représente en effet les bases de la Théorie Sens-Texte. Le premier postulat de cette théorie « signifie qu'un langage naturel est vu comme un processus logique qui établit la correspondance entre une infinité de séries de tous les sens possibles et une infinité de séries de tous les textes possibles, et vice-versa. Ainsi, pour sens donné, ce processus doit produire d'une manière idéale tous les textes qui, selon le jugement des locuteurs natifs, sont jugé, expriment correctement ce sens, simulant ainsi LA PRODUCTION DE LA PAROLE ; en partant d'un sens donné, ce processus dit extraire tous les sens qui, selon

---

<sup>98</sup> texte original : « In our conception of syntagmatic relations, there is no *inherent* connection between what can be assigned to certain topological fields and their categorial make-up. In other words, it is not necessary that the elements eligible for topological fields also form a natural class in other respects [...] Along the same lines, we reject the idea commonly accepted within GB that positions are intrinsically limited to constituents of only one projection level » Kathol 1995:139.

<sup>99</sup> texte original : « Auch wenn man die – zunächst ziemlich kontraintuitive – Annahme teilt, dass bei Verberst- und Zweitstellung das Finitum in der Komplementiererposition sitzt, um damit auf elegante Weise die zahlreichen grammatischen Bezüge zwischen erst- oder zweitgestellten Finita und subordinierenden Konjunktionen zu erklären [...] muß man durchaus nicht der Meinung sein, dass damit das Finitum in Sätzen des fraglichen Typs die *grammatische Funktion* eines Komplementierers hat. » Jacobs 1988:12

<sup>100</sup> texte original : « each language will determine a set of pairs  $(\pi, \lambda)$  [ $\pi$  drawn from Phonetic Form and  $\lambda$  from Logical Form] as its formal representations of sound and meaning, insofar as these are determined by language itself » (Ma propre traduction)

les jugement des locuteurs natifs, peuvent être correctement exprimés par le texte, simulant ainsi LA PERCEPTION DE LA PAROLE » (Mel'čuk 1988:44)<sup>101</sup>.

Néanmoins, les deux approches diffèrent dans l'architecture construite entre les deux niveaux d'interface des modèles linguistiques : tout particulièrement, leurs idées divergent sur le statut de la structure syntagmatique.

En effet, dans les modèles chomskyens, la structure syntagmatique ainsi que sa fondation théorique, la théorie X-barre, représente le centre de leur description structurelle. « Le système computationnel prend les représentations d'une forme donnée et les modifie. De la même manière, la Grammaire Universelle doit fournir un moyen de présenter une série d'items du lexique sur une forme accessible au système computationnel. Nous pouvons considérer que cette forme correspond à une version de la théorie X-barre. Les concepts de la théorie X-barre sont donc fondamentaux. Dans une théorie minimaliste, les propriétés cruciales ainsi que les relations seront exprimés dans les termes simples et éléments de la théorie X-barre<sup>102</sup> » (Chomsky 1995:172, le soulignement vient de notre part). On s'aperçoit que l'idée de X-barre est prise comme acquise et non justifiée si ce n'est que son utilisation large à l'intérieur de la théorie. Ni le concept en soit, ni sa place dans le modèle linguistique n'est jamais mise en cause.

La structure syntagmatique de la théorie X-barre porte le fardeau d'avoir à justifier la correspondance complète entre sens et textes. Examinons le diagramme classique de l'architecture des modèles chomskyens (des travaux allant de 1965 à 1995) que nous représentons dans la Figure 66. Il apparaît que La D-Structure, une structure X-barre, permet de dériver la S-Structure, une autre structure X-barre qui contient des catégories vides. La position des mots dans la S-structure est proche, mais pas nécessairement identique, à la position finale des mots qui apparaissent dans la Forme Phonétique (en effet, en dehors des règles

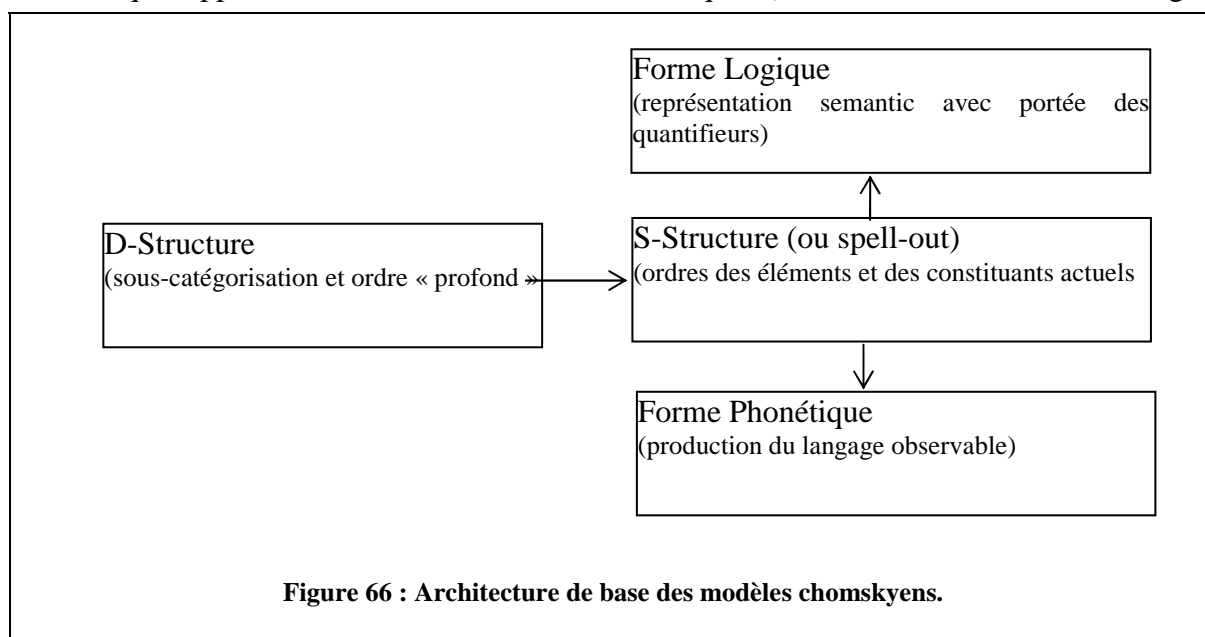


Figure 66 : Architecture de base des modèles chomskyens.

<sup>101</sup> texte original : « means that a natural language is viewed as a logical device that establishes the correspondence between the infinite set of all possible meanings and the infinite set of all possible texts and vice versa. For a giving meaning, this device must ideally produce all the texts that, in the judgment of native speakers, correctly express this meaning, thus simulating SPEAKING; from a given text, the device must extract all the meanings that, according to native speakers, can be correctly expressed by the text, thus simulating SPEECH UNDERSTANDING » (ma propre traduction).

<sup>102</sup> texte original : « The computational system takes representations of a given form and modifies them. Accordingly, Universal Grammar must provide means to present an array of items from the lexicon on a form accessible to the computational system. We may take this form to be some version of X-bar theory.

phonologiques qui s'appliquent, quelques approches considèrent qu'un mouvement de clitiques peut avoir lieu entre la structure de surface et la structure phonologique). De la même manière, La S-structure permet de dériver directement la Forme Logique.

La S-Structure encode donc l'ordre des éléments, leur place antérieure dans la S-Structure, (les traces), la structure argumentale entre les différents nœuds, et les relations de portée. Les choix impliqués dans la formation de la S-Structure sont guidés par l'idée que les dérivations impliquées devraient être aussi simples que possibles. Ainsi, la S-Structure doit être aussi proche que possible que les structures qui encodent des informations qui sont aussi différentes que la relation de portée, la linéarisation des éléments et la sous-catégorisation. Il semble évident que cela ne peut être réalisé que par une S-Structure extrêmement complexe, où on trouve une abondance de nœuds vides et de co-indexations.

Notons que dans les évolutions les plus récentes du courant chomskyen, le minimalisme, on retient la perspective traditionnelle que les langues humaines lient le son et le sens, et on considère donc que les « niveaux d'interface », la forme phonétique et la forme logique, ne peuvent pas être éliminés. « Le minimalisme cherche à établir ces niveaux nécessaires de représentation en tant que seuls niveaux [considérés ...] Il n'y a pas une représentation à la suite de toutes les insertions lexicales et avant toutes les transformations. C'est-à-dire qu'il n'y a pas de D-structure. De l'autre côté, la S-structure persiste dans un sens trivial : c'est le point où la dérivation se divise, en branchant vers la forme logique dans une direction et vers la forme phonétique dans l'autre direction »<sup>103</sup> (Lasnik 1999). L'idée est donc que l'insertion lexicale a toujours lieu à un niveau profond qu'on appelle maintenant « structure de base », mais on ne considère jamais cette structure dans sa totalité, car les mouvements à partir de la structure de base se produisent avant une autre insertion lexicale se combinant ainsi avec la structure existante (où les mouvements peuvent avoir eu lieu). Pourtant dans la S-structure du minimalisme, les traces persistent de la même manière que dans les idées pré-minimalistes, indiquant les positions de base (et intermédiaires) des éléments pendant la dérivation. La D-structure, bien qu'elle ne soit pas considérée en tant que telle reste présente, au moins implicitement.

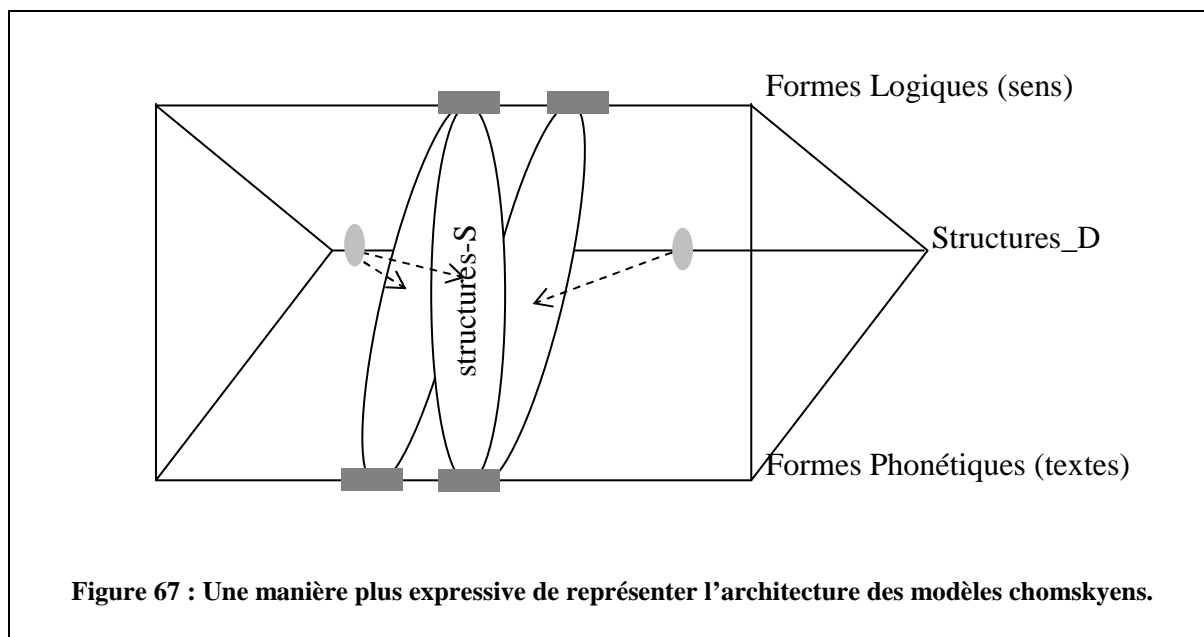
D'un point de vue de la Théorie Sens-Texte, une structure syntagmatique telle que la S-Structure comporte non seulement des représentations du sens (la Forme Logique) mais aussi des représentations du texte (la Forme Phonétique), ainsi que tous les autres représentations intermédiaires et les correspondances entre ces différents niveaux. Autrement dit, une telle structure ne contient pas uniquement les représentations, mais leurs interrelations.

C'est ce que nous avons tenté de représenter dans la Figure 67 : chaque paire sens-texte  $(\lambda, \pi)$  est réalisée par l'intermédiaire d'une S-Structure, qui est elle-même une projection de la S-Structure. Nous représentons les S-Structure par des ovales entre les représentations du sens (Forme Logique) et du texte (Forme Phonétique) qui sont représentés par des rectangles gris. Chaque S-Structure est liée à une D-Structure dans une troisième dimension.

---

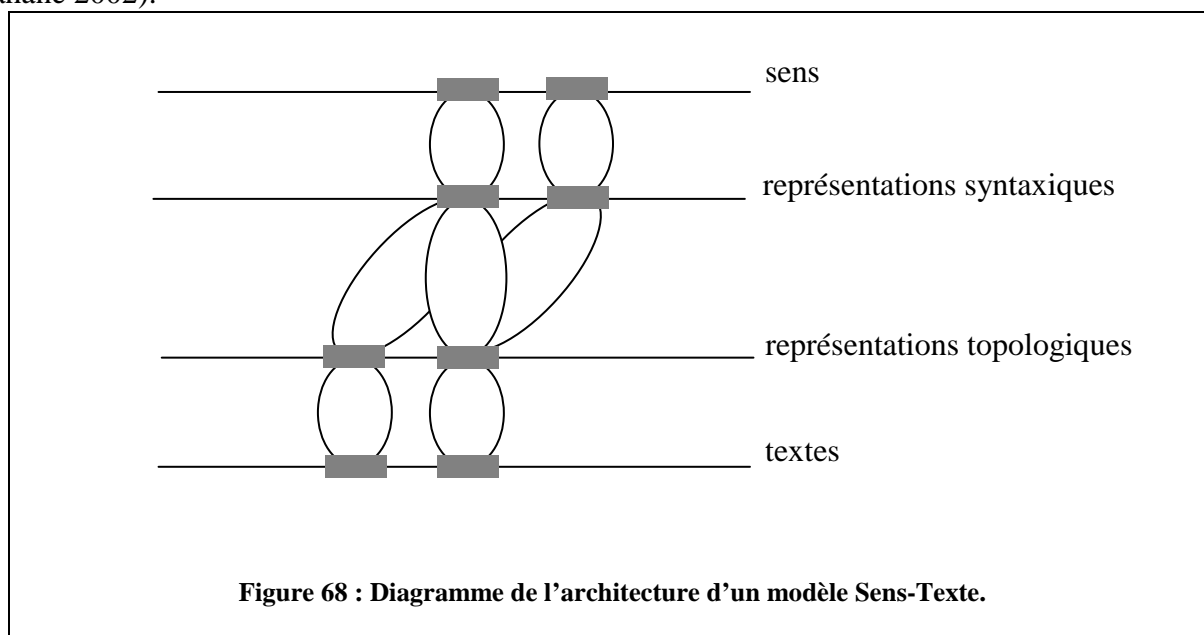
The concepts of X-bar theory are therefore fundamental. In a minimalist theory, the crucial properties and relations will be stated in the simple and elementary terms of X-bar theory ». (Ma propre traduction).

<sup>103</sup> texte originale : « Minimalism seeks to establish that these necessary levels of representation are the only levels. [...] There is no one representation following all lexical insertion and preceding all transformations. That is, there is no D-Structure. On the other hand, S-Structure, persists in one trivial sense: it is the point where the derivation divides, branching toward LF on one path and toward PF on the other. » (Lasnik 1999:549)



Dans l'approche de la Théorie Sens-Texte que nous adoptons, la correspondance entre sens et textes se passe via des représentations intermédiaires, dont la représentation syntaxique de surface et la structure topologique. Contrairement à une approche chomskyenne, nous pouvons clairement voir comment un même sens peut amener à des choix lexicaux différents (cela se passe dans un choix qui apparaît tôt dans le processus de la génération de la langue), des constructions syntaxiques différentes, des groupements de mots différents et des patrons intonatifs différents (un choix tardif dans le processus de génération de la langue). De plus, la structure communicative d'un message est présente et influence les différents choix qui sont faits, allant de la sémantique à la sortie phonétique.

Nous donnons une illustration de l'architecture de la Théorie Sens-Texte dans la Figure 68. De la même manière que dans la figure précédente, nous dessinons les représentations par des rectangles grisés et leurs correspondances par des ovales. Notons que ces correspondances ne reçoivent pas de représentation propre (pour une présentation alternative à l'intérieur de la Théorie Sens-Texte qui assignent des structures à des correspondances, nous renvoyons à Kahane 2002).





Il apparaît clairement que l'information qu'une tradition chomskyenne veut coder dans la structure syntagmatique est très différente que la tâche que nous donnons à notre structure de constituants (topologiques). La représentation topologique représente juste un niveau intermédiaire dans le chemin du sens vers le texte, représentant la manière les mots sont groupés ensemble à la surface. Inversement, un nœud d'un arbre d'une structure phrastique chomskyenne est conditionné par différents buts, dont le groupement de mots, mais aussi des conditions syntaxiques et sémantiques comme la portée des quantifieurs<sup>104</sup>.

Nous adhérons à l'idée qu'une science, doit être capable de représenter séparément des informations qui sont séparables. La Théorie Sens-Texte est à notre connaissance, l'une des seules, sinon l'unique théorie qui permet une telle séparation des représentations linguistiques. Les syntagmes (topologiques), c'est-à-dire des groupements ordonnés de mots, doivent être séparés de la dépendance syntaxique et les deux notions doivent être pris en charge par deux structures différentes, la tâche de la syntaxe étant de décrire et formaliser l'interface entre ces deux niveaux de représentations. La structure syntagmatique est donc le résultat de la combinaison de la structure communicative et la dépendance syntaxique (qui est elle-même liée à la sous catégorisation) selon les règles internes à la langues ; par contre cette structure syntagmatique n'inclut pas la dépendance syntaxique, comme c'est le cas dans les grammaires syntagmatiques basées sur la Théorie X-barre. Les syntagmes n'interviennent que lorsque l'ordre des mots est en jeu, à un niveau intermédiaire de la représentation de l'énoncé entre la représentation syntaxique et la représentation phonologique.

Dans le chapitre suivant nous tenterons de décrire les phénomènes de base de l'allemand à l'aide d'un formalisme basé sur les structures X-barre, et il sera à nouveau question de l'interprétation qu'on donne à ces structures.

---

<sup>104</sup> Suivant Mel'čuk, nous considérons que la portée des quantifieurs est intimement liée à la structure communicative, qui est importante pour le choix des différentes structures topologiques. La portée des quantifieurs peut donc influencer l'ordre des mots d'une phrase.



## Chapitre III

# **TAG**

## **Vers une grammaire d'arbres adjoints de l'allemand**

# TAG

## Vers une grammaire d'arbres adjoints de l'allemand

*Jeder Tag hat seine Plage<sup>105</sup>.  
(Matthieu 6,34)*

Le but de ce chapitre est de présenter une grammaire d'arbres adjoints lexicalisée (LTAG) de l'allemand.

Dans le domaine universitaire, peu de grammaires allemandes de large couverture ont été développées. Les deux grammaires les plus importantes sont probablement la grammaire LFG (Grammaire lexicale fonctionnelle) de l'Université de Stuttgart et la grammaire HPSG (Grammaire syntagmatique guidée par les têtes) du DFKI à Sarrebruck. Pourtant, toutes les deux sont nées dans des projets de coopération industrielles et elle ne sont donc pas facilement accessibles. De ce fait, à nos connaissances, aucun travail de comparaison a été réalisé.

La grammaire allemande LFG, provenant de l'Institut de traitement du langage naturel (IMS) de l'Université Stuttgart s'inscrit dans un projet de développement parallèle de grammaires LFG dans plusieurs langues (projet ParGram), l'anglais, l'allemand, le japonais, le norvégien et l'ourdou. La recherche actuelle autour de cette grammaire LFG porte essentiellement sur un développement le plus rapproché possible des grammaires des différentes langues du projet (Butt et al. 1999 et 2002) ainsi que sur des questions d'intégration de la Théorie de l'optimalité dans une grammaire de large couverture (Kuhn 2001)

En Allemagne, la recherche en HPSG est particulièrement active, les deux centres étant l'Université de Tübingen et le Centre allemand de recherche en intelligence artificielle (DFKI) à l'Université de Sarrebruck. La grammaire de large couverture que Stefan Müller a implémenté dans son propre système d'analyse HPSG, la grammaire du système Babel (Müller 1996, 1999) a servi de base pour le développement de la grammaire d'analyse du système de traduction de la parole Verbmobil (Müller & Kasper 2000). Cette dernière grammaire, contrairement à la grammaire Babel, est adaptée aux types de phrases apparaissant dans Verbmobil, c'est-à-dire aux prises de rendez-vous d'hommes d'affaires et ne se veut pas une grammaire générale. Pourtant, dans le but d'un traitement plus efficace dans ce projet de traduction « en temps réel », c'est cette grammaire qui a été convertie automatiquement dans le formalisme TAG par Tilman Becker (Becker & Lopez 2000). Malgré l'inadaptabilité linguistique de cette grammaire sur laquelle les auteurs-mêmes insistent, elle constitue la grammaire allemande la plus importante dans le formalisme TAG. Nous reviendrons dans la section 2.4 sur d'autres grammaires expérimentales de l'allemand qui ont été réalisées dans le formalisme TAG et ses extensions.

Avant de définir ce qu'est une LTAG, nous discuterons les notions qui sont fondamentales pour la compréhension d'une grammaire lexicalisée : nous aborderons d'abord la questions de la lexicalisation, puis de la distribution de l'information linguistique dans une métagrammaire, un outil pour la conception d'une grammaire lexicalisée. Ceci nous permettra de mettre en évidence les difficultés (parfois insurmontables) que pose la création d'une telle grammaire.

Afin de garder une validité pour la définition ultérieure d'une alternative aux LTAG, nous abordons les notions de lexicalisation et de métagrammaire sous un angle général et détaché du formalisme syntaxique. Nous introduirons également les critères qui ont été retenus pour l'utilisation du formalisme en linguistique.

---

<sup>105</sup> A chaque jour suffit sa peine.

La seconde section de ce chapitre discute les problématiques générales du formalisme TAG et certains problèmes qui sont spécifique à l'allemand. Nous verrons l'importance de l'idée de la capacité générative forte dans l'évaluation d'un formalisme donné en se reportant sur un exemple tiré des la littérature sur les grammaires lexicales fonctionnelles (LFG).

La notion de l'arbre de dérivation nous permettra de donner une interprétation sémantique aux résultats d'une dérivation TAG, mais nous verrons ensuite deux cas où cette interprétabilité sémantique est limitée : les verbes à contrôle et la modification multiple.

Les problèmes de l'interprétation de la structure des constituants que TAG produit, nous amène ensuite à discuter la nécessité des critères linguistiques pour l'élaboration d'une grammaire dans le formalisme TAG.

Dans la troisième section, nous présentons et discutons l'outil utilisé pour la création concrète d'un grand nombre d'arbres de la TAG de l'allemand, le compilateur de méta-grammaire TAG. Nous montrons pourquoi il a été nécessaire de revoir et d'améliorer la définition du référent minimal, notion essentielle pour la description d'arbres dans une méta-grammaire.

Enfin, nous arrivons à la section consacrée à la DTAG proprement dite. Nous commencerons par présenter la raison d'être d'une TAG de l'allemand malgré des défaillances importantes du formalisme pour cette langue : nous aborderons les phénomènes de brouillage des arguments et des modifieurs ainsi que différents problèmes liés au propositions relatives. Sur ces bases, il nous est possible de définir la tâche très limitée d'une TAG de l'allemand : la génération des textes.

Dans un but concret d'application informatique de la grammaire et à l'aide d'une hiérarchie de critères pour sa création, nous pouvons ensuite présenter les premiers arbres simples d'une TAG de l'allemand. Nous présentons donc les différents choix de forme des arbres élémentaires que nous avons faits dans la construction de notre grammaire. Enfin la fin de cette section présente l'organisation proprement dite de la DTAG, avec la description en classes de l'information linguistique contenue dans les arbres de la TAG de l'allemand, que l'outil de méta-grammaire nous permet de concevoir.

## 1 Les grammaires d'arbres adjoints lexicalisées

Le but de cette section est de donner une présentation générale des grammaires d'arbres adjoints lexicalisés. Nous présentons les notions de lexicalisation et de méta-grammaire et aboutiront à la définition de TAG.

### 1.1 La lexicalisation

Une grammaire générative classique a comme but principal de générer toutes les phrases acceptables, et seulement les phrases<sup>106</sup> acceptables, d'une langue donnée. Pour ce faire, on dispose d'un ensemble de règles qui s'appliquent (à partir d'un symbole initial) suivant un ordre donné jusqu'à la génération complète d'une phrase. Il est possible de garder une trace de cette dérivation et la désigner comme structure syntaxique de la phrase générée.

Un grand nombre de linguistes cherchent à écrire des grammaires de réécriture dont les structures de dérivation coïncident avec les structures syntagmatiques, et c'est en partie possible. De manière générale, une analyse linguistique consiste en une ou plusieurs structures plus complexes qui donnent plus d'information que juste l'ordre dans lequel les règles ont été appliquées. Les structures syntaxiques ou sémantiques ne sont pas justifiées uniquement par le fait qu'elles permettent d'obtenir une chaîne de mots, mais par d'autres considérations linguistiques. Par exemple, on peut vouloir exprimer la portée des éléments, leur relation

---

<sup>106</sup> Une phrase peut être définie soit comme une suite de caractères, soit comme un signe linguistique.

syntactique ou leur groupement en segments. Ce qui importe finalement dans le formalisme linguistique n'est pas quelles règles ont été appliquées pour obtenir une phrase, mais quelle structure (ou quel ensemble de structures) on attribue à cette phrase. Une phrase n'est grammaticale que si elle possède une structure ; les règles ne produisent pas de structure pour des phrases agrammaticales.

Afin de générer l'ensemble (infini) des phrases d'une langue, il est nécessaire de construire des règles qui soient générales, c'est-à-dire qui servent à la construction d'un nombre infini de phrases. Par exemple, la règle suivante est une règle générale : si A et B sont des phrases du français, alors l'expression « A et B » constitue aussi une phrase de cette langue.

L'application de telles règles doit être limitée à certains éléments linguistiques. En effet, dans la construction de la structure syntaxique d'une phrase, on est très souvent amené à considérer des catégories de mots qui partagent certaines propriétés, sans pour autant partager toutes leurs propriétés. Par exemple, considérons le verbe français *coûter*. Il s'agit d'un verbe transitif qui porte les propriétés de la catégorie *verbe transitif* (il a deux arguments, il s'accorde avec le premier, il se combine avec tous les temps verbaux composés ou non) à une exception près : les règles du passif ne doivent pas lui être appliquées, puisque ce verbe ne peut pas apparaître avec des formes passivées. Cette information sur la *passivation* doit donc être transportée à travers toutes les règles générales (de réécriture par exemple) qu'on exprime sur des verbes transitifs jusqu'à l'application de la règle du passif. Dans des analyses syntaxiques faites à l'aide de grammaires hors contexte, cette information est habituellement mise dans des traits, qui sont passés et unifiés de règle en règle. Un tel système de traits permet d'exprimer des sous-classes de l'ensemble des catégories aussi spécifiques que nécessaires.

Cependant, un tel système de traits peut mener à une certaine inefficacité dans le processus de construction de la structure linguistique pour deux raisons principales. D'une part, l'incompatibilité d'une règle avec une autre apparaît souvent tardivement, à la suite de l'application d'un grand nombre de règles, et un algorithme d'analyse doit donc tenter un tout autre chemin. D'autre part, les mêmes règles s'appliquent fréquemment dans le même ordre à des objets de la même catégorie ou sous-classe. D'un point de vue computationnel, il s'avère donc utile de limiter l'espace de recherche de l'algorithme par mot apparaissant dans la phrase à analyser (ou à générer). On fournit ainsi à l'algorithme d'analyse (ou de génération) le mot muni d'un ensemble de règles prenant en compte le contexte dans lequel le mot apparaît. Puisqu'on ne s'intéresse pas aux règles mêmes mais aux structures résultantes, on peut encore améliorer l'algorithme en remplaçant l'ensemble de règles par la structure que cet ensemble peut générer. Cette structure doit garder une *place libre* pour des éléments obligatoires ou optionnels qui peuvent intervenir dans les règles en question<sup>107</sup>. Le fait de savoir comment ces *places libres* peuvent être occupées par des mots (ou par des structures de mots) dépendra de la formalisation retenue.

Suite à ces remarques, nous pouvons maintenant définir une grammaire lexicalisée : une *grammaire lexicalisée* est une grammaire organisée par mots, c'est-à-dire une grammaire où chaque mot porte un ensemble (fini) de structures (ou des ensembles de règles qui permettent d'obtenir ces structures) correspondant aux différents contextes dans lequel le mot apparaît.

Nous illustrons l'idée de métagrammaire dans la Figure 69 : chaque mot possède une liste de structures partielles qui peuvent être combinées en une structure finale à l'aide de simples règles de connexion. Un résultat possible d'une telle structure finale est donné dans la Figure 70.

---

<sup>107</sup> La structure générée par une règle de combinaison de plusieurs sous-structures est attachée à la tête du syntagme introduit par la règle. La notion de gouvernement est donc aussi indispensable pour une grammaire lexicalisée.

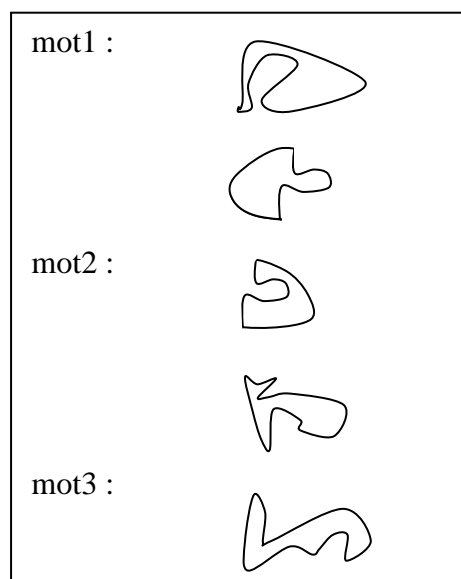


Figure 69 : principe général d'une grammaire lexicalisée

La définition d'une grammaire lexicalisée reste vague non seulement parce que le terme *contexte* reste non défini mais aussi parce qu'il est nécessaire de spécifier encore comment les fragments de structures précalculées sont finalement combinés. On retrouve différents degrés de lexicalisation en fonction de la définition concrète que l'on donne à cette combinaison de fragments de structures.

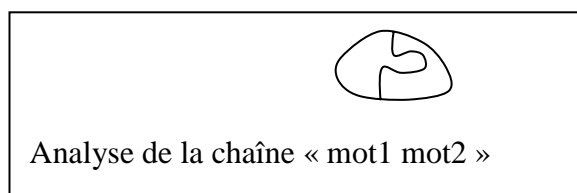


Figure 70 : exemple d'une analyse avec une grammaire lexicalisée

La seule façon d'obtenir une grammaire complètement lexicalisée est d'attacher au mot toutes les constructions dans lequel le mot peut apparaître ; or, comme une langue comporte un ensemble infini de phrases et un ensemble fini de mots (si on ne tient pas en compte la dérivation morphologique, qui pose un grand problème pour les grammaires lexicalisées), la plupart des lexèmes recevront une liste infinie de structures. Une grammaire doit alors comporter un certain nombre de règles de combinaison des structures attachées aux mots. On dira que plus les règles sont simples, moins nombreux et moins paramétrables, plus la grammaire peut être considérée comme lexicalisée. D'un point de vue computationnel, aussi bien pour l'analyse que pour la génération, la lexicalisation n'est avantageuse que si les structures contextuelles attachées au mot sont limitées en nombre et en taille, et que la combinaison de deux structures est une procédure algorithmiquement simple. L'avantage algorithmique découle donc de la possibilité de calculer « off-line » les contextes envisageables, c'est-à-dire avant la véritable analyse (ou génération) d'une phrase, et d'accéder directement aux structures contextuelles précalculées pendant l'analyse (ou la génération) d'une phrase. Ceci nous évite de considérer toutes les règles qui n'apparaîtront jamais avec le mot en question.

D'un point de vue linguistique, l'avantage de la lexicalisation est qu'elle permet d'exprimer facilement les spécificités de chaque mot. Ainsi, si l'on revient à notre exemple du verbe "coûter", on ne trouvera tout simplement pas de structure passivée parmi les structures contextuelles de ce verbe. Le passif ne sera donc jamais considéré pour ce verbe. De la même manière, chaque mot peut porter un ensemble « personnalisé » de structures contextuelles. La

grammaire lexicalisée consiste donc en la liste des mots d'une langue munie de la liste des structures contextuelles pour chaque mot.

Néanmoins, la lexicalisation comporte des problèmes qui sont à la fois pratiques et théoriques :

- Comment créer et maintenir une grammaire lexicalisée ?
- Comment exprimer des généralisations sur le comportement des mots dans une grammaire lexicalisée ?

Ces questions vont trouver une réponse lorsqu'on abordera la question de regroupement des structures par lexème et par *famille* (présenté dans la sous-section 1.1.2) et celle de métagrammaire (présenté dans la section suivante 1.2).

### 1.1.1 La délexicalisation partielle

L'impossibilité d'une grammaire complètement lexicalisée nous amène à la question de savoir où les généralisations nécessaires doivent intervenir afin de générer l'ensemble (infini) des phrases. Une *délexicalisation* est un regroupement des structures contextuelles effectué non par la forme du mot mais par une généralisation qui porte sur les formes des mots. De cette manière, une délexicalisation nous permet d'exprimer des généralités sur l'ensemble des contextes représentés par les structures regroupées.

L'exemple le plus simple de délexicalisation consiste en des généralités qu'on veut exprimer sur toutes les formes d'un lexème<sup>108</sup>, sans être obligé de les répéter pour chaque contexte dans lequel une forme du lexème peut apparaître. A ma connaissance, cette première étape de délexicalisation est réalisée dans toutes les grammaires lexicalisées : les entrées sont organisées par lexème (ou par lemme) et non par formes pleines, d'où le nom de « grammaire lexicalisée »<sup>109</sup>. Il n'y a donc qu'une seule liste de contextes par lexème.

Prenons un exemple. Le lexème [*coûter*] comporte les structures contextuelles de toutes ses formes : *coûter*, *coûtons*, *coûtée*<sup>110</sup>, ... Ce n'est que pendant l'analyse (ou pendant la génération) que chaque forme trouvera parmi ces différentes structures celles qui lui conviennent. Le calcul en temps réel devient alors tout de suite beaucoup plus compliqué car il faut, à l'aide de traits, sélectionner le bon arbre. Par ailleurs, dans certains cas, il faut faire passer dans la structure lexicalisée, des informations qui ne sont pas communes à toutes les formes, comme l'accord d'un verbe fini avec son sujet.

Le regroupement des mots par lexèmes est un choix commun, linguistiquement motivé mais formellement arbitraire. On pourrait aussi bien imaginer de regrouper les mots par forme, c'est-à-dire de mettre tous les participes passés féminins singuliers sous une entrée commune, sous une autre tous les infinitifs, ou toutes les formes finies du présent de la 1<sup>ère</sup> personne du pluriel, etc. Les ressemblances de comportement, et donc les ressemblances dans les contextes dans lesquels apparaissent les mots d'un tel ensemble sont plus grandes qu'à l'intérieur d'un lexème. (Il n'y a par exemple aucun contexte dans lequel peuvent apparaître en même temps aussi bien *coûtons* que *coûté*, et il faut soit adapter soit bloquer mutuellement les structures de chacun des deux mots à l'aide de traits.)

En fait, le regroupement en lexèmes est plutôt guidé par la sémantique ; *coûtons* et *coûté* porte le même sens, et il est difficile d'imaginer un système qui « comprenne » *coûtons* mais pas *coûté*. On veut savoir dans quels contextes peut apparaître le sens de 'coûter' ; il est moins intéressant d'établir une classification d'après les contextes dans lesquels peuvent apparaître, par exemple, les participes passés au féminin singulier, car la définition de cette catégorie est justement distributionnelle, c'est-à-dire qu'elle dépend du contexte dans lequel la forme apparaît.

<sup>108</sup> Cf. Mel'čuk 1988.

<sup>109</sup> Ce choix d'ordre par lexème ne représente pas la seule possibilité ; on peut aussi bien imaginer d'ordonner la grammaire par d'autres objets apparaissant dans la description linguistique. Cf. le cas des locutions, des auxiliaires, des verbes support, etc.

<sup>110</sup> Par exemple, la forme *coûtée* n'apparaît que dans les relatives comme *Tu ne peux pas imaginer la somme que m'a coûtée cette montre*.



### 1.1.2 La notion de famille

Parallèlement au regroupement par lexème, il existe une autre classification des structures lexicalisées, qui est aussi dépendante de l'analyse linguistique et formellement circulaire : c'est le cas de la classification en *familles*. Une *famille* est un ensemble de structures contextuelles lexicalisées qui apparaissent habituellement avec une certaine catégorie de mots. Par exemple, la famille des verbes transitifs comporte toutes les structures qui peuvent apparaître avec les verbes transitifs (les structures des formes finies, de l'infinitif, du passif, etc.). Pourtant, la définition des verbes transitifs, qui est distributionnelle et non sémantique, renvoie justement à l'ensemble des contextes dans lesquels les verbes transitifs apparaissent habituellement, sauf exception que l'on rencontre par exemple dans le cas du verbe *coûter*.

D'autres critères sont donc nécessaires pour décider de l'utilité de l'existence d'une famille. Une famille peut être définie comme un ensemble de structures contextuelles qui apparaît comme un ensemble de structures contextuelles d'un grand nombre de lexèmes. Elle est donc la traduction de la notion de *catégorie syntaxique* dans les grammaires lexicalisées, et chaque distinction entre deux familles est aussi discutable et dépendante de l'analyse linguistique et de l'utilité pour la description souhaitée que la distinction entre deux catégories.

Par exemple, grâce aux familles, l'ajout d'un nouveau lexème dans une grammaire lexicalisée devient plus facile : il ne faut plus vérifier l'applicabilité de chaque structure contextuelle à ce nouveau lexème, mais il suffit de trouver la famille du lexème (sa catégorie) et vérifier pour chaque structure contextuelle de la famille si elle peut constituer un contexte du lexème. Dans ce cas, on dit que le lexème *ancrer* la structure.

Une lexicalisation n'est donc jamais totale, et les choix nécessaires de délexicalisation, comme le regroupement des structures contextuelles par lexème et leur regroupement en famille, dépendent de l'analyse linguistique souhaitée.

## 1.2 L'idée de métagrammaire

Cependant, arranger des structures contextuelles par lexème et par famille ne résout pas la difficulté qu'on a à exprimer des généralisations linguistiques dans une grammaire lexicalisée. Considérons les familles des verbes transitifs et intransitifs afin d'illustrer ce propos. Ces deux familles partagent beaucoup de propriétés ; autrement dit, les contextes dans lesquels peuvent apparaître les verbes transitifs et les verbes intransitifs, décrits par des structures contextuelles dans leur famille respective, se ressemblent : par exemple, les deux familles sont ancrées par les verbes, ce qui implique de prendre en considération certaines propriétés d'ordre linéaire, et les différents contextes doivent prévoir une place pour le sujet du verbe. Un sujet d'un verbe intransitif se comporte et se place de la même manière qu'un sujet d'un verbe transitif<sup>111</sup>. Ceci représente un indice parmi d'autres pour la notion générique de *sujet*.

Deux choix se présentent pour exprimer cette généralité sur le sujet dans une grammaire lexicalisée : une première possibilité est la délexicalisation par rapport au sujet, c'est-à-dire, que l'information sur les verbes transitifs et intransitifs indiquent que les classes doivent encore ajouter l'information sur le sujet. L'information est stockée dans la classe spécifique de *sujet*. L'avantage que représente la délexicalisation est que l'information sur le sujet est stockée une seule fois ; par contre l'inconvénient est qu'il y a une perte d'efficacité du calcul et de finesse de l'expression qu'apporte justement la lexicalisation.

Le deuxième choix que nous avons pour exprimer la généralité sur le sujet est d'introduire un méta-niveau d'expression linguistique, placé en dehors de la véritable analyse (ou génération), un niveau qui n'intervient qu'au moment de la création des structures lexicalisées.

A ce niveau dit de *métagrammaire*, les propriétés générales qui sont communes, par exemple, à tous les sujets, ne sont ainsi exprimées qu'une seule fois, et pendant le processus de *compilation*

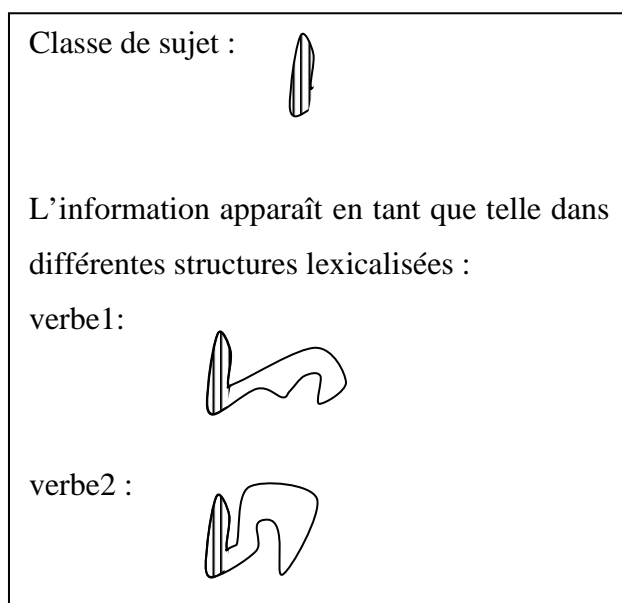
<sup>111</sup> En français, on a certaines restrictions sur l'inversion (Cf. Bonami 2001).

de la grammaire lexicalisée, chaque famille qui a besoin de cette propriété l'ajoute à des structures lexicalisées correspondantes. L'introduction d'un méta-niveau permet donc d'exprimer des généralités linguistiques tout en gardant lexicalisée, la « véritable » grammaire.

En même temps, l'introduction de la métagrammaire pose de nouvelles questions :

La métagrammaire est introduite afin de mieux exprimer des généralités qu'on observe dans une langue. En effet, ses classes correspondent mieux à notre modèle de la langue que les structures lexicalisées. Pourtant, pendant le processus d'analyse ou de génération, on n'a accès qu'aux structures lexicalisées, et les généralisations (comme la notion de *sujet*) n'apparaissent pas en tant que telles.

La Figure 71 illustre de manière très abstraite l'idée de la métagrammaire : tous les verbes qui



**Figure 71 : Illustration du lien entre une classe d'une métagrammaire et des entrées lexicales de la grammaire lexicalisée portant l'information de la classe.**

portent un sujet ont en commun un morceau de leur structure contextuelle ; il s'agit justement de la partie permettant à un sujet de s'introduire dans la structure du verbe. Il est donc suffisant d'exprimer une seule fois l'information contenue dans cette partie, identique à tous les verbes (qui prennent un sujet). Cette information est stockée dans une *classe* à part de la métagrammaire. Ainsi, pour la construction de la structure contextuelle d'un verbe qui prend un sujet, il n'est plus nécessaire de spécifier les détails de cette partie de la structure ; il suffit d'indiquer que la structure du verbe doit inclure l'information de la classe *sujet*. Un *compilateur* de la métagrammaire peut ensuite créer des structures contextuelles finales des verbes en question, toutes contenant une information identique sur le comportement contextuel du sujet. La description devient plus facile et les structures contextuelles résultantes sont plus homogènes, justement parce qu'on a pu exprimer des généralités sur le sujet.

Il apparaît donc que le problème d'un tel modèle est que cette dichotomie entre connaissances profondes d'un côté et la combinaison simple des structures lexicalisées de l'autre, semble être arbitraire (et dépendante de la puissance des règles non-lexicalisées restantes) et manque d'indices empiriques.

Une autre façon possible d'interpréter la métagrammaire est de considérer les structures lexicalisées comme pur produit computationnel sans autre valeur linguistique que d'être la combinaison de différents bouts d'information intervenant dans le contexte d'un mot. Pour évaluer la difficulté d'une telle approche, il faut connaître la complexité du calcul des structures lexicalisées fait à partir des classes de la métagrammaire. S'il s'agit juste de la composition des informations contenues dans différentes classes, alors la structure lexicalisée résultante porte directement et visiblement ces informations (comme il est graphiquement illustré dans la Figure

71). Plus le calcul est complexe, plus la question se pose selon le statut des structures lexicalisées, car après un calcul complexe, les structures lexicalisées comporteront des éléments nouveaux, qui n'apparaissent pas en tant que tels dans les classes de la métagrammaire et le statut linguistique de ces éléments nouveaux reste à définir.

Nous verrons dans la section 3 que les métagrammaires proposées pour le formalisme de la grammaire d'arbres adjoints lexicalisée sont faites de manière à ce que les règles de combinaison de l'information des différentes classes soient extrêmement simples. On trouve donc la même information dans les structures lexicalisées que dans les classes qui sont combinées pour créer ces dernières. La simplicité de ces règles présente l'inconvénient de rendre les classes très dépendantes des structures lexicalisées – on ne peut exprimer les généralités observées sans prêter attention à la nature du formalisme des structures lexicalisées qui doivent finalement en être déduites.

### 1.3 Définition de TAG

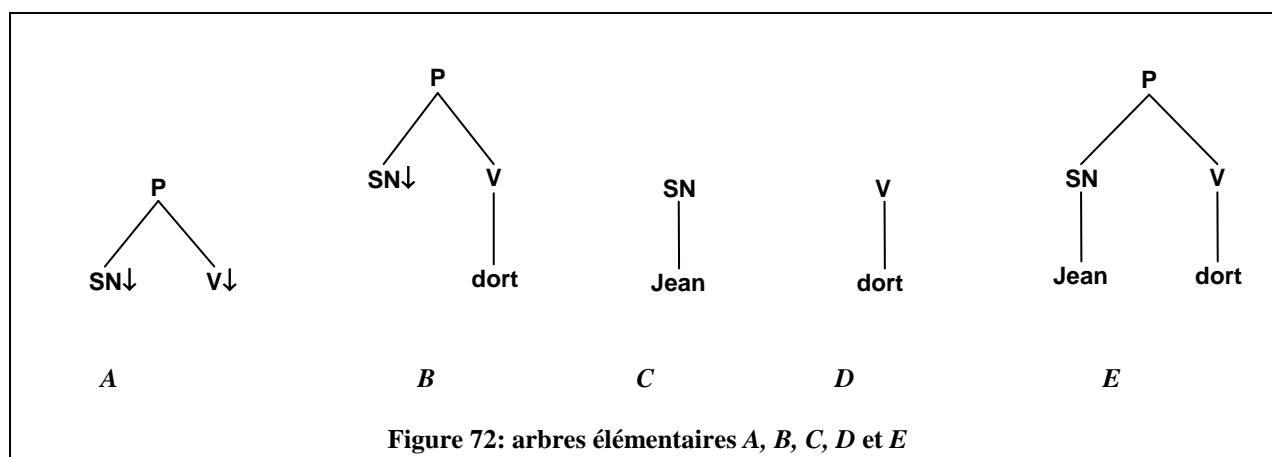
Les grammaires hors-contexte (ou algébriques) sont un cas particulier des *grammaires de réécriture*. Dans une grammaire de réécriture, une règle consiste toujours en un élément (non-terminal) qui est réécrit comme une chaîne d'autres éléments (terminaux ou non-terminaux). De plus, dans une grammaire hors contexte, on ne peut pas spécifier de conditions contextuelles à l'application d'une règle. Par exemple, on ne peut pas dire que telle règle ne s'applique que quand tel symbole est précédé de tel autre symbole.

Une grammaire contextuelle (ou sensible au contexte, angl. : '*context-sensitive*'), où de telles restrictions sont permises, a des propriétés computationnelles notablement moins bonnes que celles d'une grammaire hors-contexte. Il est donc d'un intérêt d'abord purement formel de rechercher des formalismes à mi-chemin entre des grammaires hors-contexte et des grammaires sensibles au contexte qui possèdent une plus grande capacité générative que les grammaires hors contexte tout en restant calculables en temps polynomial.

De manière indépendante, il est intéressant de considérer de tels formalismes pour une modélisation linguistique du fonctionnement syntaxique car d'une part, les grammaires hors-contexte ne sont clairement pas suffisamment puissantes pour les langues naturelles : les dépendances sérielles croisées (angl. *cross serial dependencies*, voir section 2.1 suivante) montrent l'insuffisance de la description faite avec un tel formalisme. Chomsky 1957 a introduit les transformations pour remédier aux problèmes d'une description basée sur des grammaires hors-contexte pures. Par ailleurs, les grammaires contextuelles semblent trop puissantes : en effet, les descriptions syntaxiques de ces grammaires n'utilisent que quelques conditions contextuelles spécifiques. De plus, une modélisation de la langue doit prendre en considération que pour un humain, le temps de l'analyse et de la génération ne semble pas augmenter exponentiellement avec le nombre de mots à traiter. Dans une telle perspective, la recherche d'un formalisme munie de bonnes propriétés computationnelles dans lequel on peut exprimer un modèle syntaxique s'impose.

Les Grammaires d'Arbres Adjoints (*Tree Adjoining Grammars*, TAG) étaient présentées en 1975 par Aravind Joshi comme une classe de grammaires dont la capacité générative se situe entre celle des grammaires hors-contexte et celle des grammaires contextuelles. Leur capacité générative est appelée « légèrement sensible au contexte ». Cette capacité générative permet à TAG de dériver l'ensemble  $\{a^n b^n c^n \mid n \in \mathbb{N}\}$ , ce qui est impossible avec une grammaire hors-contexte.

Une règle hors-contexte peut alors être vue comme une partie prédéfinie d'un arbre. Pour reprendre l'exemple de Abeillé A. 1993:203-204, la règle  $P \rightarrow SNV$  apparaît comme sous-structure dans l'arbre de constituants qui est construit à l'aide de cette dernière (voir la structure A de la Figure 72). Ces parties de l'arbre syntagmatique, décrites par des règles hors contexte, ont toujours la profondeur 1, car une règle réécrit un symbole par la liste de ses fils.



L'idée initiale de TAG est de permettre à une règle de décrire plus d'un seul niveau dans l'arbre de constituants construit. Une règle TAG peut donc avoir une profondeur plus grande. Par exemple, dans l'arbre *B* de la Figure 72, on voit une structure de profondeur 2, qui combine la règle  $P \rightarrow SN V$  (= arbre élémentaire *A*) et la règle  $V \rightarrow dort$  (= arbre élémentaire *D*). Une telle partie de l'arbre syntagmatique final est appelée *arbre élémentaire*. L'arbre élémentaire *C*, correspondant à la règle  $SN \rightarrow Jean$ , peut être combinée avec *B* pour donner la structure syntagmatique souhaitée (structure *E* de la Figure 72).

### 1.3.1 Substitution et adjonction

La combinaison de deux arbres élémentaires se fait au moyen des deux opérations suivantes : la *substitution* et l'*adjonction*. L'opération qui consiste à substituer la racine  $X$  dans un nœud  $X\downarrow$  s'appelle *substitution*. Dans le cas de notre exemple, le placement de l'arbre *C* dans l'arbre *B* pour obtenir l'arbre *E* représente une substitution. Un nœud feuille qui permet cette opération (normalement tous les nœuds feuilles non terminaux) s'appelle *nœud de substitution* et il est marqué par une flèche vers le bas (voir les arbres *A* et *B* de la Figure 72). Pendant cette opération, le nœud de substitution est remplacé par le nœud racine de l'arbre de l'argument du gouverneur où est ancré le nœud de substitution.

La deuxième opération de combinaison de deux arbres élémentaires dans une TAG est appelée *adjonction*. Si un arbre élémentaire porte le même symbole sur sa racine que sur un nœud feuille, le nœud feuille, nommé *nœud pied*, peut être marqué par un astérisque, créant ainsi une structure appelée *arbre auxiliaire*. Un tel arbre, par exemple l'arbre *A* dans la Figure 72), peut s'adjoindre à un nœud d'un autre arbre, par exemple l'arbre élémentaire *B*, portant le même symbole que le nœud pied de *A*. L'adjonction habituelle (voir ci-dessous la problématique d'une définition alternative) consiste en un remplacement du nœud de l'arbre *B* par toute la série de nœuds de l'arbre *B* entre sa racine et le nœud pied. Nous appelons cette série de nœuds qui composent le chemin entre la racine et le nœud pied la *colonne* de l'arbre (d'après le terme anglais *spine*, pour *colonne vertébrale*).

Contrairement au processus de substitution, l'adjonction est itérative ; un arbre auxiliaire peut s'adjoindre à un autre arbre auxiliaire. Un arbre élémentaire peut contenir plusieurs nœuds de substitution et un seul nœud pied<sup>112</sup>.

Une *grammaire d'arbres adjoints* est donc un ensemble d'arbres élémentaires muni d'un symbole initial. Les arbres peuvent être combinés par adjonction ou par substitution. Les structures complètes (appelées *arbres dérivés*) sont le produit de la combinaison d'arbres élémentaires ayant le symbole initial comme racine et dont tous les nœuds pied et de substitution

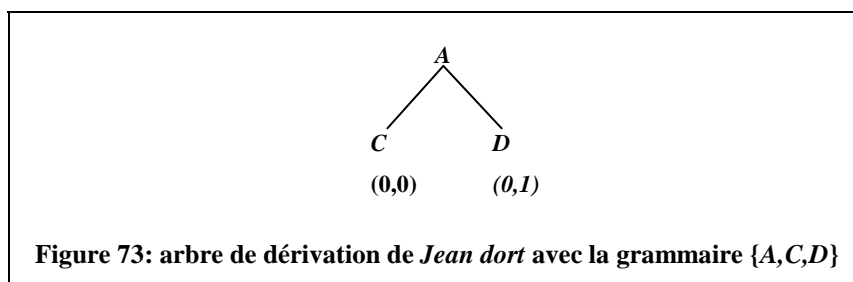
<sup>112</sup> La raison de cette asymétrie est la difficulté de définir la structure résultant d'une double adjonction. Voir la discussion dans la section 2.2.

sont consommés. Le langage généré par une TAG est l'ensemble des chaînes de nœuds feuilles des arbres dérivés.

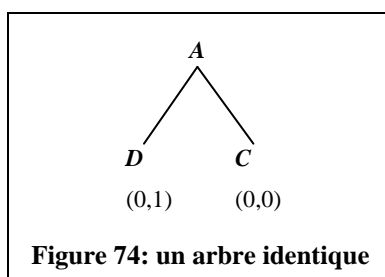
Les ensembles  $\{A,C,D\}$ ,  $\{B,C\}$  et  $\{E\}$  de la Figure 72 sont chacun une TAG avec le symbole initial  $P$  et ils génèrent tous les trois le même langage  $\{(Jean\ dort)\}$  consistant en une seule phrase. Pour distinguer les différents processus à l'œuvre dans les trois manières de dériver la structure finale  $E$ , on introduit un *arbre de dérivation* qui n'est rien d'autre qu'une façon de noter quel arbre a été adjoint ou substitué dans quel autre arbre.

Considérons ces ensembles un par un.

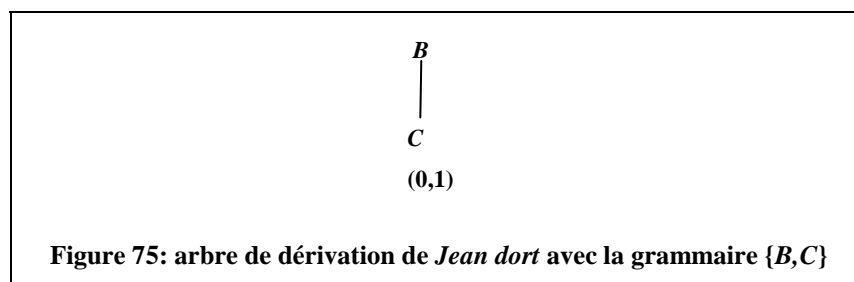
La grammaire  $\{A,C,D\}$  est équivalente à la grammaire hors contexte et l'arbre de dérivation ressemble à l'arbre dérivé :



L'arbre  $C$  est substitué dans l'arbre  $A$  à l'adresse de Gorn  $(0,0)$ , et l'arbre  $D$  à l'adresse  $(0,1)$ . Grâce à cette information sur le lieu d'opération, l'information concernant l'ordre entre les feuilles d'un arbre de dérivation n'est plus pertinente (comme pour un arbre de dépendance) : l'arbre dans la Figure 74 est une autre notation pour la même structure.



Avec la grammaire  $\{B,C\}$  on n'a besoin que d'une seule opération pour obtenir un arbre dérivé complet : l'arbre  $C$  doit être substitué dans le nœud de substitution de  $B$ . On obtient alors l'arbre de dérivation suivant :



La grammaire  $\{E\}$  possède déjà un arbre dérivé complet, l'arbre  $E$ . On n'a donc pas besoin de faire une quelconque opération et l'arbre de dérivation n'est que le nœud  $E$ .

Seule la grammaire  $\{B,C\}$  est une grammaire lexicalisée suivant la définition que nous avons donné au début de la section.

Une *grammaire d'arbres adjoints lexicalisée (LTAG)* est une TAG où chaque arbre élémentaire porte comme nœud feuille au moins un élément lexical. Par ailleurs, deux autres critères doivent être vérifiés :

- *le Principe de sous-catégorisation* : « tout prédicat contient dans sa structure élémentaire au moins un nœud pour chacun des arguments<sup>113</sup> qu'il sous-catégorise » (Abeillé 1993:212). Ce nœud doit être un nœud de substitution ou d'adjonction.
- *le Principe sémantique* : tout arbre élémentaire correspond à une et une seule entité sémantique<sup>114</sup>.

La TAG  $\{B, C\}$  vérifie les critères de validité d'une LTAG : chaque arbre est lexicalisé et correspond à une entité sémantique (*Jean* et *dormir*) et l'arbre du prédicat *dormir* contient un nœud de substitution pour l'agent du verbe.

On nomme les arbres élémentaires en utilisant une représentation des entités sémantiques correspondant à ces arbres. L'arbre de dérivation est donc constitué de nœuds portant des valeurs sémantiques.

Toutefois, ces critères sont très stricts et nous verrons dans la section 2.4.7 qu'aucune TAG de large couverture ne les remplit complètement. De plus, il est clair que les critères de validité, et donc la conformité avec les critères, dépendent de la représentation sémantique choisie.

Chaque nœud d'un arbre élémentaire peut être augmenté d'une structure de traits doublée en *traits amont* et *traits aval*. En cas de substitution, les traits amont et aval du nœud de substitution sont unifiés avec les traits de la racine de l'arbre substitué. Le cas de l'adjonction est un peu plus complexe : les traits amont et aval du nœud sur lequel l'adjonction a lieu sont séparés ; les traits amonts sont unifiés avec les traits amonts de la racine de l'arbre auxiliaire, les traits avals sont unifiés avec les traits du nœud pied (voir discussion dans la section 02.2.2).

Nous voyons donc qu'une LTAG est une grammaire lexicalisée suivant la définition donnée au début de la section on assemble une structure linguistique (ici syntagmatique) à partir de structures lexicalisées. L'avantage de ces opérations de combinaison (adjonction et substitution) est qu'elles sont computationnellement peu chères et permettent donc une analyse efficace<sup>115</sup>.

## 2 Critiques et extensions

Après avoir posé les principes de base des grammaires TAG, nous discutons dans ce chapitre quelques points faibles de ce formalisme. Plus précisément, nous abordons les problèmes de la capacité générative, de l'arbre de dérivation et de l'interprétation de la structure dérivée. En dernier lieu, nous discutons pourquoi, dans un premier temps, l'allemand paraît incompatible avec le formalisme TAG. Pour ce faire, nous reprenons quelques phénomènes problématiques de l'allemand tel que le brouillage (d'arguments et de modificateurs), les relatives. Nous comparons également les problèmes d'une DTAG par rapport à d'autres grammaires existantes comme la XTAG et la FTAG.

---

<sup>113</sup> Cette définition ne spécifie pas s'il s'agit d'un argument sémantique ou syntaxique. Cela ne cause pas de problème tant que les deux coïncident.

<sup>114</sup> D'un point de vue de l'analyse sémantique dans la Théorie Sens-texte (Mel'čuk 2001), on considère que chaque entité sémantique (*sémantème*) est décomposable, à l'exception d'un très petit nombre de sémantèmes de base, qui restent à définir. De manière générale, il est très difficile de donner des critères clairs indiquant quand une partie de sens est suffisamment amalgamée afin de pouvoir justifier l'introduction d'un sémantème le représentant. Cette problématique sémantique ne nous empêche pas d'utiliser le principe sémantique dans les cas non disputés. On peut donc par exemple affirmer que *dormir* possède un seul sens, et que *Jean dort* est compositionnel.

<sup>115</sup> L'algorithme d'analyse pour les TAG est d'ordre  $n^6$ ,  $n$  étant le nombre de mots d'une phrase, cf. Schabes & Joshi 1988.

## 2.1 Capacité générative

Il est important de comprendre que le choix d'un formalisme ne peut être fait sur les bases d'une capacité générative faible. Depuis Chomsky 1957, qui justifiait la nécessité des transformations par le fait que les grammaires hors contexte ne peuvent assigner les descriptions structurelles correctes à une phrase, nous savons qu'il faut d'abord choisir la description qu'on veut donner à une phrase, et que ce n'est qu'après qu'on peut trouver le formalisme qui est capable de la générer.

Les TAG ont été conçues comme un « système de génération d'arbres » (Joshi 1975). Ce système repose à la fois sur des objectifs formels où on cherche la création d'un système ayant une capacité générative qui se situe entre contextuel et hors contexte, mais aussi des objectifs linguistiques. Les objectifs linguistiques que présente Joshi 1975 peuvent être résumés dans les trois points suivants :

1. distinction entre phrase élémentaire et modification dans la dérivation et distinction entre les différentes dérivations de la même structure
2. Idée de compositionnalité de la structure linguistique à partir d'unités phrastiques
3. Un outil permettant de décider sur la nécessité d'enchâssement.

En fait, ces idées ont été formulées dans un contexte de recherche linguistique centré autour de l'arbre syntagmatique. La question essentielle reste encore celle de la capacité générative faible du formalisme proposé et l'arbre syntagmatique est vu comme objet mathématique. On ne trouve pas de discussion sur la relevance hors le formalisme présenté, et des nœuds apparaissant dans la structure. Le troisième point ci-dessus implique même que le formalisme nous aidera à décider sur quelle structure linguistique devra être retenue. Ainsi, le formalisme reste étroitement lié à un point de vue sur la langue, et qui ne donne par conséquent qu'un choix très limité de structures syntagmatiques possibles.

. Les structures qu'on peut générer avec une TAG sont proches des structures souhaitées, mais non tout à fait identiques. Par la suite, nous verrons des exemples de structures dont la forme est principalement influencée par des considérations provenant du formalisme TAG<sup>116</sup>.

Prenons comme première illustration des difficultés de TAG le cas des dépendances sérielles croisées (angl. *cross serial dependencies*) du néerlandais. En néerlandais, le complexe verbal sans *te* (l'équivalent du *zu* allemand) exige que les dépendants des verbes soient placés suivant le même ordre que les verbes gouvernants. L'exemple (143) comporte trois verbes enchâssés, chacun portant un dépendant nominal. La chaîne subordonnante des verbes (et de leurs dépendants) est indiquée par des indices.

- (199) ... dat Jan<sub>N1</sub> Piet<sub>N2</sub> de kinderen<sub>N3</sub> zag<sub>V1</sub> helpen<sub>V2</sub> zwemmen<sub>V3</sub>.  
 ... que Jean Pierre les enfants voyait aider nager.  
 ... que Jean a vu Pierre aider les enfants nager.

Selon Huybregts 1976, un tel cas montrerait que le néerlandais ne peut être un langage hors contexte, car l'ensemble  $\{\omega\omega \mid \omega \in V^*\}$  n'est pas un langage hors contexte (si  $V$  contient au moins deux éléments). Or Bresnan *et alii* 1982 donne une grammaire hors contexte de cette

---

<sup>116</sup> Il est possible de nier complètement l'existence de structures dont la justification ne provient pas directement du formalisme. Autrement dit, on peut défendre qu'un formalisme, qui arrive à générer (ou analyser) seulement les phrases correctes d'une langue, assigne aussi automatiquement de bonnes structures. Un premier problème d'une telle approche, par exemple une analyse statistique, est qu'elle peut donner de bons résultats (souvent même meilleurs que les résultats de systèmes basés sur des règles) sans pour autant donner à beaucoup de linguistes l'envie de considérer les analyses d'un parseur statistique comme structures linguistiques de la phrase (pour un avis contraire voir par exemple Rens Bod 1998). Un deuxième problème se trouve dans le fait que la plupart des formalismes considèrent une analyse linguistique comme la correspondance entre une structure profonde et une structure de surface. Cette structure profonde doit être au moins définie de manière indépendante du formalisme (ce qui permettrait toujours de se débarrasser de toutes les structures intermédiaires).

langue qui peut générer l'ensemble des phrases, incluant les dépendances sérielles croisées. L'affirmation de Huybregts est donc fautive. Néanmoins, Bresnan *et alii* 1982 soulignent que la structure que donne cette grammaire hors contexte est incorrecte, car elle ne reflète pas l'analyse syntaxique souhaitée. La problématique n'est donc pas la génération de l'ensemble, mais la génération de la bonne description structurelle.

Bresnan *et alii* 1982 ont proposé par la suite une meilleure structure tout en montrant qu'une telle structure ne peut pas être générée par une grammaire hors contexte. Ils ont finalement abouti à une grammaire lexicale fonctionnelle (LFG), capable de générer la structure voulue<sup>117</sup>.

Seul un tel procédé peut justifier un formalisme. Il ne suffit pas de proposer un formalisme qui peut analyser la phrase en question sans donner une justification de l'analyse proposée.

La classification des formalismes par rapport à leur capacité générative forte est beaucoup plus complexe qu'une classification par rapport à la capacité générative faible, car il faut savoir comparer deux analyses complètes qui peuvent même consister en plusieurs structures, comme c'est le cas des formalismes comme LTAG ou de LFG (notions d'arbre dérivé et de dérivation pour LTAG, voir section suivante, c-structure et f-structure pour LFG, voir section 6 du quatrième chapitre). Pour une discussion de cette problématique, voir notamment Miller 1999 et Chiang 2001.

Une fois qu'on a établi l'importance de la capacité générative forte d'un formalisme, on peut se poser la question suivante : une LTAG est-elle adaptée aux dépendances sérielles croisées ?

Il est impossible de donner une réponse complète à cette question sans définir les structures reflétant notre analyse linguistique du néerlandais, ce qui dépasse le cadre de cette thèse.

Néanmoins, indépendamment de ce choix linguistique, nous pouvons faire les remarques suivantes.

Tout d'abord, il est possible de traduire la grammaire hors contexte proposée par Bresnan *et alii* 1982 en une TAG qui générerait (faiblement) l'ensemble de phrases du néerlandais contenant des dépendances sérielles croisées. Une telle traduction est toujours possible et l'on peut par ailleurs lexicaliser cette TAG en précalculant l'application d'arbres élémentaires non lexicalisés à un élément lexicalisé au choix. Cependant, une telle analyse n'obéirait pas au principe de sous-catégorisation, comme le remarque Joshi 1985.

Joshi 1985 propose comme alternative une TAG qui est capable d'analyser les dépendances sérielles croisées. La structure dérivée qu'on obtient avec cette analyse ne correspond ni à la structure obtenue par la traduction de la grammaire hors contexte, ni à celle qui a été proposée par Bresnan *et alii* 1982. Par exemple, la structure dérivée de TAG diffère de celle proposée par Bresnan *et alii* 1982 car une TAG ne peut pas créer une structure à « double colonne », ce qui est le cas de la structure générée par la grammaire LFG. Néanmoins, Joshi 1985 admet explicitement que la nouvelle structure ne repose pas sur des justifications linguistiques particulières ; comme c'est souvent le cas en TAG comme dans beaucoup d'autres formalismes (Cf. note de bas de page 117), la fin justifie les moyens, c'est-à-dire que le bon arbre de dérivation permet d'expliquer tout arbre dérivé et la discussion sur l'arbre dérivé est repoussée à une analyse ultérieure. On voit donc que l'analyse linguistique proposée par Joshi 1985 (mais aussi en TAG d'une manière générale à l'exception notable de Abeillé 1991) est centrée sur l'arbre de dérivation, et finalement, seule l'arbre de dérivation est « raisonnable », la justification

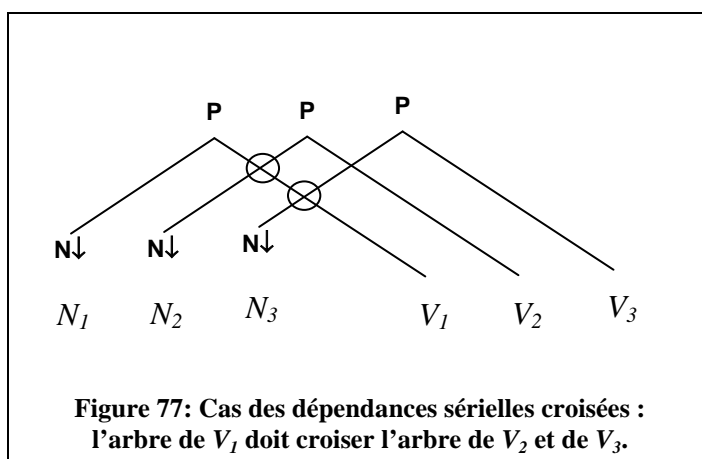
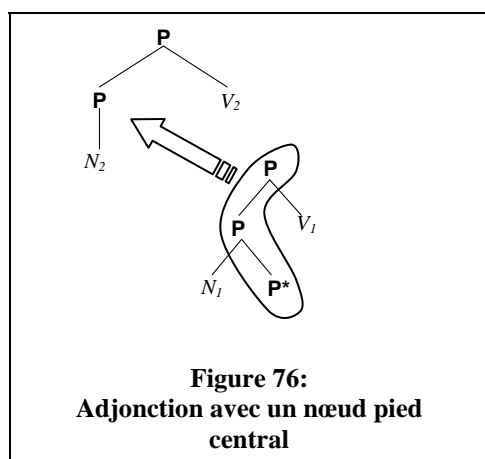
---

<sup>117</sup> Bresnan *et alii* 1982:620 présentent une structure contenant des nœuds SV dominant exclusivement des nœuds nominaux. Cette structure, pour le moindre étonnant, laisse soupçonner que leur structure était au moins influencée par la possibilité de sa génération en LFG. Mais tant que la justification linguistique de cette structure est acceptable, cette influence ne pose pas de problème. Il faudra, bien entendu, définir des critères clairs permettant de comparer la qualité descriptive des structures syntagmatiques. Voir la section II.6 pour une discussion de cette problématique.



de la structure dérivée qui correspond à cette structure de dérivation étant renvoyée aux calendes grecques<sup>118</sup>.

En fait, la description proposée par Joshi pour la grammaire TAG obéit au principe de sous-catégorisation, ce principe ne nous donne pas beaucoup d'alternative à la solution proposée par Joshi. Dans une analyse TAG des dépendances sérielles croisées, il est possible de supposer que chaque niveau supplémentaire d'enchâssement est réalisé par un arbre auxiliaire<sup>119</sup>  $A_i$  contenant un nom  $N_i$  qui doit être placé devant tous les noms qui sont déjà présents, et d'autre part un verbe  $V_i$  qui, de la même manière, doit être placée devant la suite de verbes déjà présente. La seule façon de réaliser le double croisement (Cf. Figure 77) est de faire en sorte que l'arbre auxiliaire  $A_i$  ait un nœud dans sa colonne permettant l'adjonction  $A_{i+1}$ , de manière à ce que le nom  $N_{i+1}$  puisse avoir accès à la place devant le nom  $N_i$ . Le placement du verbe enchâssé doit s'effectuer de la même manière. On n'a donc pas de choix : l'arbre auxiliaire ressemble à la structure que



l'on représente dans la Figure 76.

Pour qu'un arbre auxiliaire puisse servir dans une analyse des dépendances sérielles croisées obéissant au principe de sous-catégorisation, il faut qu'il ait un nœud  $P$  sur la colonne qui domine aussi bien le nom que le nom pied (qui se trouve à droite du nom), mais qui ne domine pas le verbe (qui se trouve à droite de ce nœud)<sup>120</sup>. Par conséquent, dans tous les arbres dérivés qui résultent d'une analyse correcte, tous les verbes sont placés plus haut que tous les noms, et cela indépendamment de l'analyse syntaxique propre. Nous voyons donc que le formalisme impose une structure qui peut ou non correspondre à la structure induite par les critères syntaxiques sur la nature des constituants.

En Français, le cas des coordinations avec « respectivement » constitue un cas parallèle. Bien qu'il soit moins clair que dans le cas du néerlandais de quel type est le lien entre les éléments de la phrase (200) commençant par la même lettre, dans la mesure qu'on veuille une grammaire avec des arbres communs à chaque couple de noms, nous n'avons qu'une seule possibilité pour les structures dérivées : les objets syntaxiques (*un ananas, une baguette et une crêpe*) vont se trouver plus haut dans l'arbre que les sujets (*Aude, Bertrand et Cécile*).

<sup>118</sup> « The TAG described below captures the cross-serial dependencies, and the derivations are in the spirit of TAGS. I must emphasize that the TAG given here is primarily for the purpose of illustrating how a TAG can be constructed to capture the cross-serial dependencies in the right manner, keeping the derivations in the spirit of TAGs. I do not wish to claim any detailed linguistic justification for the structure proposed here; however, the dependencies are correctly represented » Joshi 1985:246

<sup>119</sup> Il est évidemment impossible de réaliser de telles structures avec des substitutions enchâssées.

<sup>120</sup> A ceci près qu'on peut changer le nom des nœuds et qu'on peut insérer d'autres nœuds dans la structure.

(200) Aude, Bertrand et Cécile mangent respectivement un ananas, une baguette et une crêpe.

L'augmentation de la capacité générative (faible et donc aussi forte) d'une TAG par rapport à une grammaire hors contexte est donc d'un grand intérêt formel et elle permet en plus d'obtenir des analyses sémantiques qu'une grammaire hors contexte ne fournit pas, mais les principes linguistiques retenus pour les TAG restreignent considérablement les possibilités structurelles et par conséquent la capacité générative forte.

## 2.2 L'arbre de dérivation

La « beauté linguistique » des TAG par rapport à une grammaire hors contexte ne se trouve donc pas dans les structures dérivés qu'une TAG permet de générer. Elle se trouve plutôt dans un phénomène laissé de côté jusqu'ici : l'interprétabilité sémantique (ou au moins syntaxique profonde) de l'arbre de dérivation.

La lexicalisation ainsi que les critères linguistiques de validité des TAG font qu'un arbre de dérivation ressemble à un graphe de dépendance sémantique. Il s'agit d'un arbre non ordonné dont chaque nœud constitue une entité sémantique. Le Principe de sous-catégorisation ne précise pas le type de sous-catégorisation dont il est question. Il semble clair, bien que non rendu explicite, que les développeurs des TAG existantes sont partis de l'idée qu'il y a une coïncidence entre les deux types de sous-catégorisation : un argument syntaxique est aussi un argument sémantique et vice-versa. Ce dernier postulat est clairement faux, mais si nous l'assumons pour l'instant, il nous est possible d'interpréter un arbre de dérivation comme structure sémantique.

En stipulant cette interprétabilité sémantique, on suppose en fait qu'il existe un lien direct entre la structure syntaxique et la sémantique. En assemblant l'un, on construit l'autre. Or, une telle ambition est trop grande ; si l'on se limite à la description du lien qui existe entre structure de constituants et dépendance syntaxique, on donne déjà une mission suffisamment complexe à un module du modèle linguistique : tous les phénomènes d'ordre doivent être traités à ce niveau, ce qui s'avère déjà difficile. C'est dans cette perspective que nous allons proposer dans la section 3 du quatrième chapitre, une grammaire lexicalisée qui tente de ne garder que les avantages du formalisme LTAG, tout en réparant quelques-unes de ses défaillances.

Par la suite, nous proposons d'examiner quelques difficultés de cette ambition sémantique en détail. Nous développerons en premier lieu le problème des verbes à contrôle puis celui de l'adjonction modifiante multiple.

### 2.2.1 Le problème des verbes à contrôle

Par définition, les verbes à contrôle (du sujet ou de l'objet) partagent un actant sémantique avec leur dépendant verbal. Dans la phrase (201), *Pierre* est en même temps l'actant sémantique du verbe *vouloir* et du verbe *dormir*. On obtient donc une structure sémantique comme celle de la Figure 78.

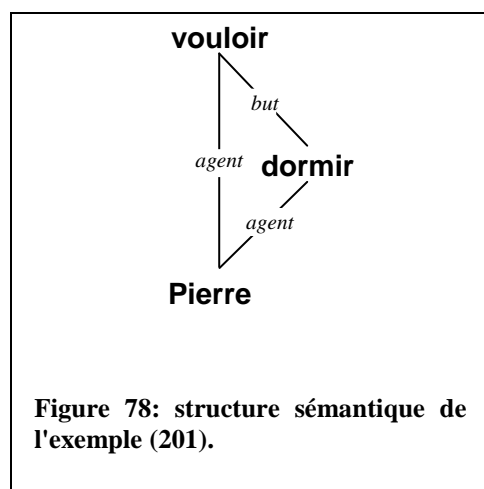


Figure 78: structure sémantique de l'exemple (201).

(201) Pierre veut dormir.

Le partage de l'entité *Pierre* ne peut être exprimé dans l'arbre de dérivation d'une TAG, car le nom *Pierre*, apparaissant une seule fois en surface, ne peut être substitué (ou adjoint) qu'une seule fois, contrairement aux structures sémantiques. Il faut absolument choisir entre l'expression du lien entre le verbe de contrôle et le nom, ou de celui entre le verbe enchâssé et le nom. Un arbre de dérivation est ainsi toujours acyclique. Depuis Abeillé 1991, il est devenu courant dans la tradition TAG de privilégier le lien avec le verbe gouvernant, c'est-à-dire avec le verbe à contrôle. On considère la récupération de ce lien manquant (pour un éventuel calcul sémantique basé sur l'arbre de dérivation) plus facile dans cette direction.

Le verbe à contrôle porte ainsi un nœud de substitution pour son sujet et un nœud pied afin de s'adjoindre à l'arbre élémentaire de son dépendant verbal. Dans la Figure 79, l'adjonction prédicative de *vouloir* dans *dormir* est indiquée avec une ligne pointillée dans l'arbre de dérivation. Pour une discussion de la problématique du verbe à contrôle en TAG voir aussi Candito and Kahane 1998a.

Le choix de l'adjonction de l'arbre élémentaire du verbe à contrôle s'impose par la possibilité d'analyser de cette manière des dépendances de longue distance, comme c'est le cas dans la phrase (202), où *écrire* sous-catégorise le syntagme prépositionnel avec *à* qui se trouve disloqué à gauche en début de phrase. Les arbres élémentaires de la Figure 80 permettent une analyse correcte de la phrase (202). Notons au passage les nœuds unaires *V* dans l'arbre élémentaire de

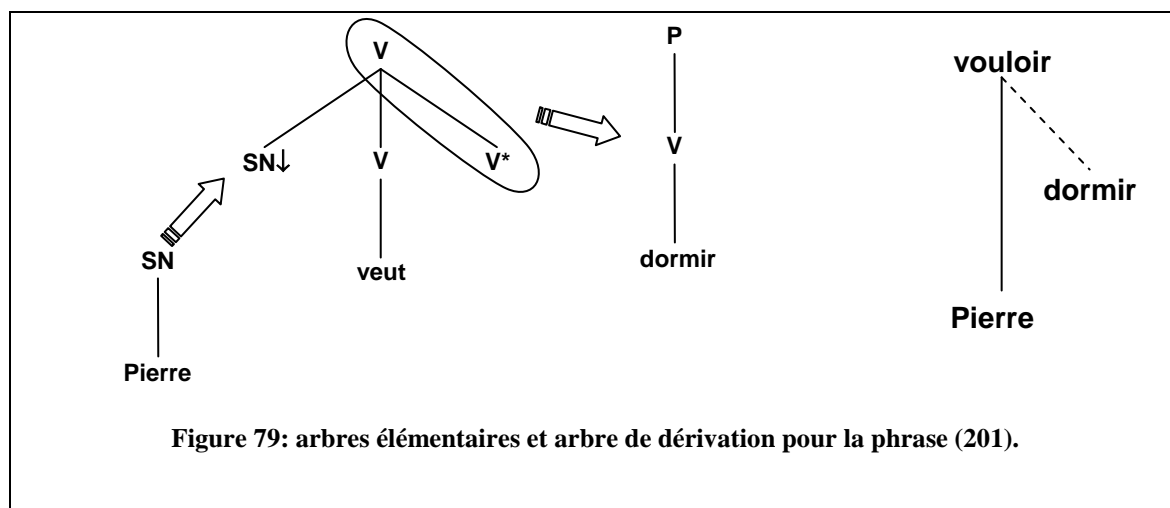
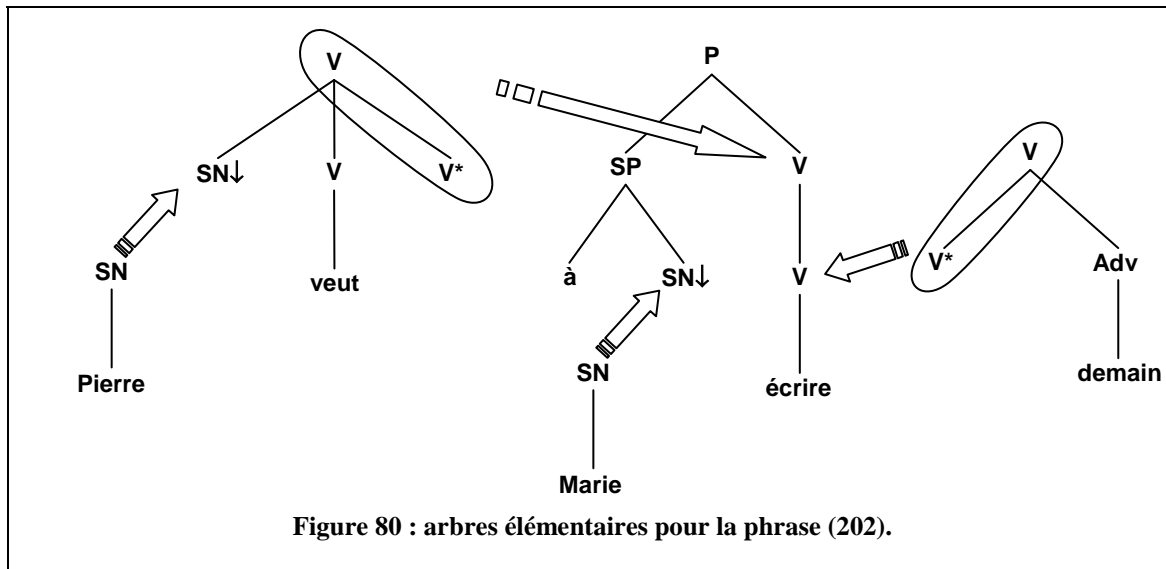


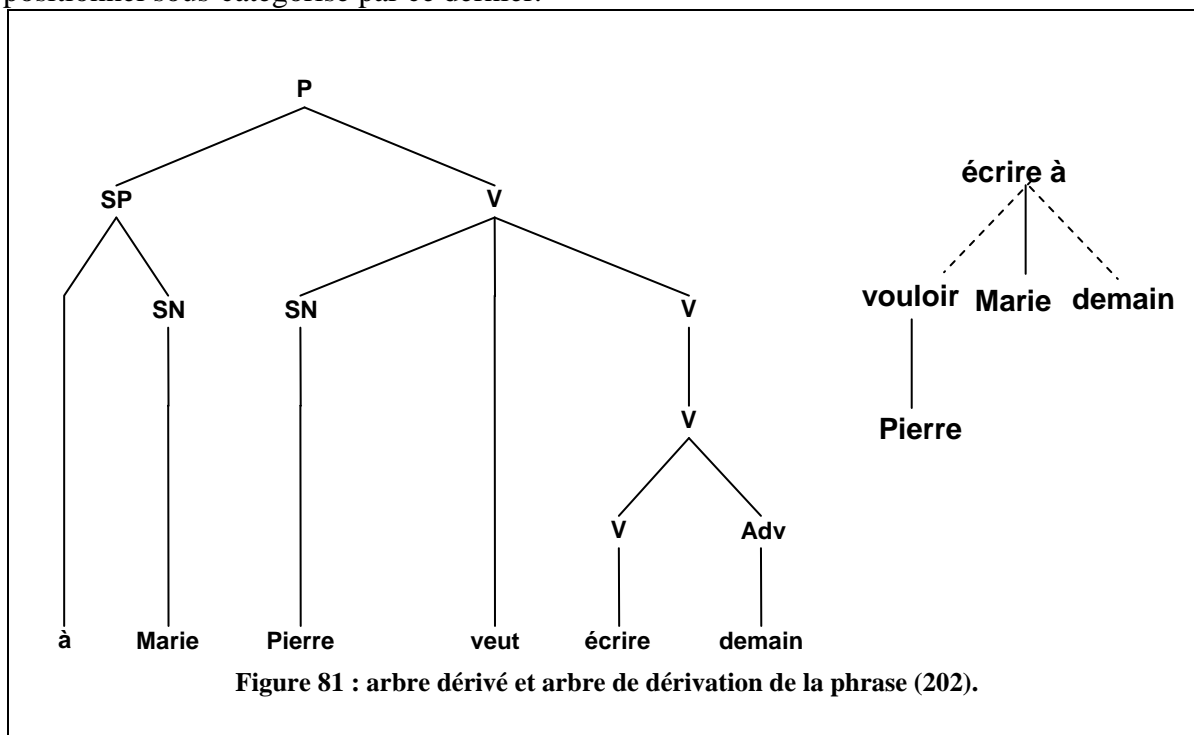
Figure 79: arbres élémentaires et arbre de dérivation pour la phrase (201).

*écrire*. Ces derniers sont nécessaires pour permettre à *vouloir* et à *demain* de s'adjoindre à *écrire*. Avec un seul nœud, il n'est possible d'obtenir que la 2<sup>e</sup> analyse de cette phrase avec l'adjonction de l'adverbe au verbe à contrôle, puisque l'adjonction à un nœud pied est interdite. Nous renvoyons le lecteur à la section 2.2.2 pour une explication de cette règle et des problématiques qui en découlent.

(202) A Marie, Pierre veut écrire demain.



Les structures résultantes de cette analyse TAG sont montrées dans la Figure 81. On ne pourra analyser cette phrase si la sous-catégorisation verbale est exprimée par une substitution, car le verbe gouvernant interviendrait en surface entre le dépendant verbal et le syntagme prépositionnel sous-catégorisé par ce dernier.



L'opération d'adjonction remplit donc deux rôles distincts dans la description linguistique : la modification (comme par exemple l'adverbe *demain* ci-dessus) et la prédication (comme le verbe à contrôle *vouloir*).

(203) l'homme qui veut dormir

Pourtant, il est possible qu'un arbre élémentaire doive remplir les deux rôles à la fois (mais le formalisme ne le permettra pas). Par exemple dans la phrase (203), la proposition relative, gouvernée et donc ancrée par le verbe à contrôle *vouloir*, modifie le nom *homme*, tandis qu'il devrait, en même temps, être adjoindé dans l'arbre élémentaire de *dormir*, son dépendant verbal. On aimerait ainsi donner à l'arbre élémentaire de *vouloir* deux nœuds pieds, l'un pour s'adjoindre sur l'antécédent et l'autre pour gouverner *dormir*. Or un tel arbre élémentaire avec deux nœuds pieds est interdit pour des raisons algorithmiques : pour le faire, il faudrait complètement redéfinir l'adjonction, car le remplacement habituel d'un nœud par la colonne des nœuds entre racine et nœud pied de l'arbre auxiliaire ne serait plus possible. On ne sait pas quel arbre est adjoindé pendant la 2<sup>e</sup> adjonction et on ne saurait choisir la catégorie de la racine de l'arbre auxiliaire, car l'une des adjonctions exige qu'elle soit *N*, tandis que l'autre exige qu'elle soit *V*.<sup>121</sup>

Pour les mêmes raisons, l'algorithme d'analyse de TAG doit traiter les adjonctions avant les substitutions, car dans le cas contraire, on pourrait aboutir à un arbre dérivé intermédiaire avec deux nœuds pied comme l'on peut le voir dans l'exemple de la Figure 82 :

La difficulté de la phrase (203) réside dans le traitement des verbes à contrôle avec argument relativisé qui est différent : au lieu de s'adjoindre au verbe gouverné, il substitue son dépendant

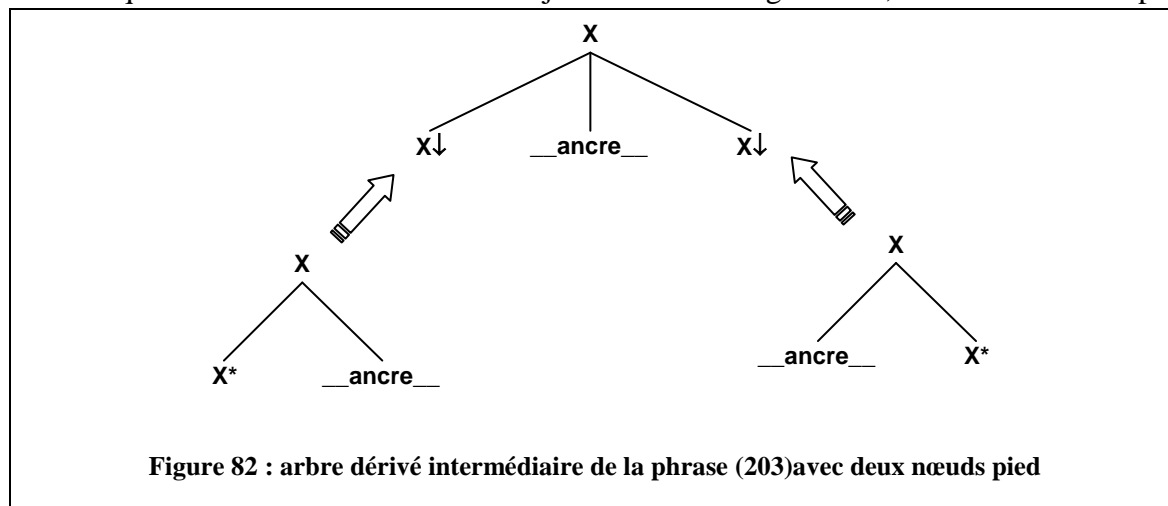
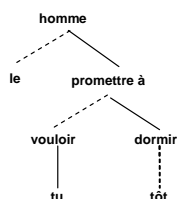


Figure 82 : arbre dérivé intermédiaire de la phrase (203) avec deux nœuds pied

verbal à un nœud de substitution. Cette analyse est possible, car dans les cas des relatives, on ne peut pas avoir une dépendance de longue distance, du type de la phrase (202) (contrainte d'îlot). Le problème est qu'un éventuel post-traitement de l'arbre de dérivation avec comme but d'en faire une dépendance sémantique devra prendre en compte ces exceptions<sup>122</sup>.

<sup>121</sup> Kahane *et alii* 2000 propose une analyse TAG avec une catégorie *amont* et une catégorie *aval*, ce qui résoudrait ce dernier problème. Cf. aussi Candito et Kahane 1998 pour une variante de TAG permettant la double adjonction (GAG).

<sup>122</sup> L'analyse d'un syntagme comme (i) doit adjoindre *vouloir* et substituer *dormir* dans *promettre*, donnant un arbre de dérivation comme suit.



(i) L'homme à qui tu veux promettre de dormir tôt.

Le calcul de la « vraie » structure sémantique, contenant par exemple un lien *agent* entre *dormir* et *tu*, est non trivial.

## 2.2.2 Grammaire d'arbres adjoints – mais adjoints comment ?

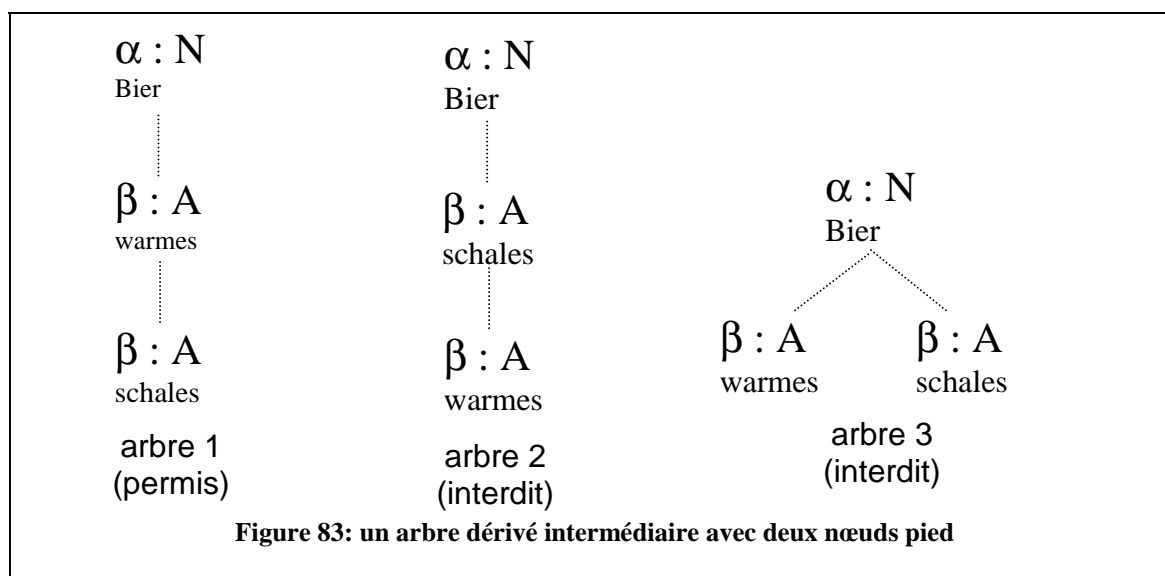
Les problèmes de l'adjonction et les remèdes proposés par Schieber et Shabes 1994 constituent une bonne illustration de la problématique du lien direct supposé en TAG, entre la construction de la structure syntagmatique et la structure sémantique.

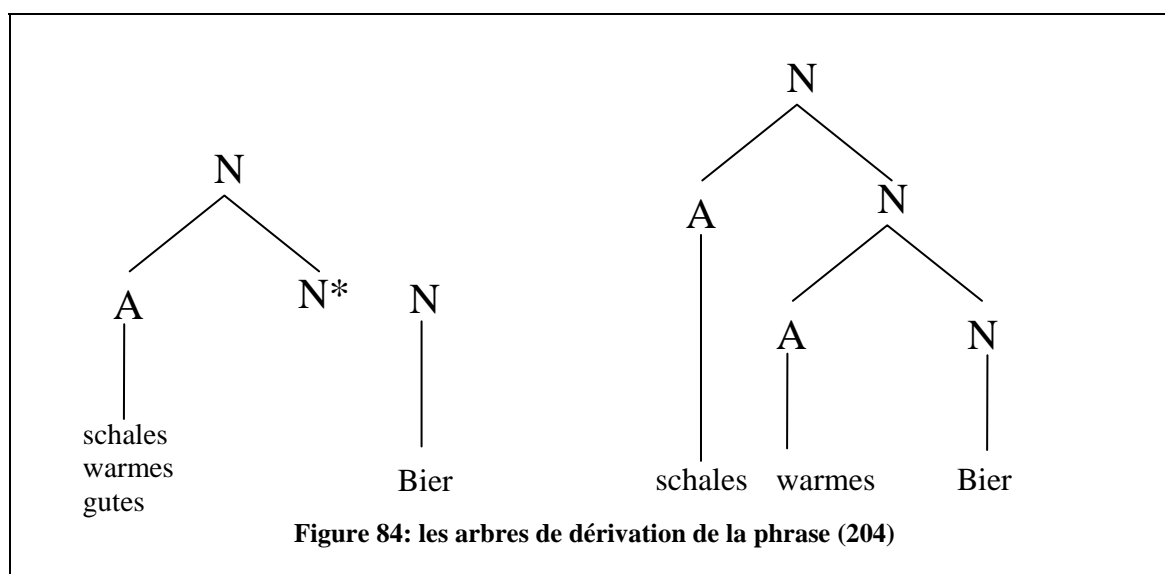
Rappelons brièvement le fonctionnement de l'adjonction : dans la définition classique proposée par Vijay-Shanker 1987, l'adjonction est considérée comme un processus qui consiste en trois étapes principales. Si l'on considère que l'arbre  $\gamma_2$  est adjoint sur l'arbre  $\gamma_1$  au nœud  $X$  :

- remplacer le nœud  $X$  sur lequel on adjoint avec tout l'arbre  $\gamma_2$ .
- recopier le sous-arbre  $\gamma'$  de  $\gamma_1$  qui était dominé par  $X$  sous le nœud pied de  $\gamma_2$ .
- unifier les traits amont de  $X$  avec les traits amont du nœud racine de l'arbre auxiliaire  $\gamma_2$  et unifier les traits aval de ce nœud avec les traits aval du nœud pied de  $\gamma_2$ .

S'il n'y a pas d'autres restrictions, on peut avoir, pour un arbre dérivé donné, plusieurs dérivations, et donc plusieurs arbres de dérivation (le contraire est toujours possible et souhaité). En effet, pour la phrase (204) et les arbres élémentaires de la Figure 82 (nous n'indiquons qu'une structure pour les arbres élémentaires identiques des trois adjectifs), on obtient les trois dérivations de la Figure 84 (les arbres auxiliaires sont marqués avec la lettre  $\beta$ , les autres avec la lettre  $\alpha$ ).

(204) schales warme Bier  
de la bière éventée et chaude





Afin d'établir une bijectivité entre dérivation et arbre de dérivation, il faut donc éliminer deux de ces possibilités. La définition classique de TAG proposée par Vijay-Shanker 1987 prévoit deux restrictions importantes :

1. L'adjonction à un nœud pied est interdite (même s'il a déjà été adjoint et a donc perdu son marquage d'étoile).

Sans cette restriction, chaque enchaînement de deux adjonctions d'arbres auxiliaires correspondrait à deux arbres de dérivation différents. Mais cette restriction fait que l'arbre de dérivation 2 est interdit parce qu'il suppose une adjonction au nœud pied de *schales*.

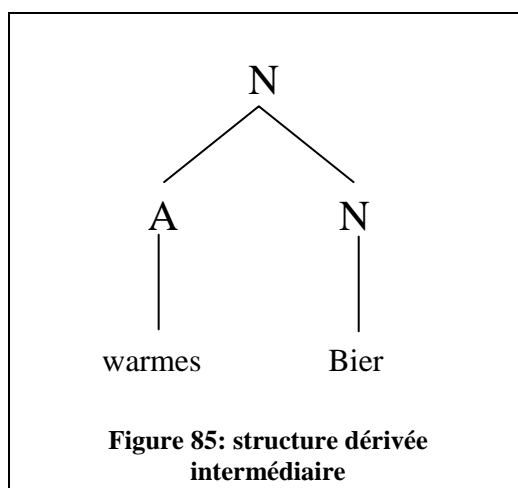
2. Une seule adjonction est permise par nœud.

C'est ainsi qu'on élimine l'arbre 3 où les deux arbres auxiliaires s'adjoignent au même nœud *N* de *Bier*.

La contribution de Schieber et Shabes 1994 intervient sur cette question : dans la mesure que pour quelques arbres auxiliaires, la seule dérivation qui survit aux restrictions (arbre 1) est bel et bien celle qui est souhaitable<sup>123</sup>, Schieber et Shabes constatent que dans le cas d'une modification, l'arbre 3 s'approche plus de la structure sémantique désirée. En effet, on voit mal comment on pourrait justifier dans l'arbre 1 la modification de *warmes* par *schales*. Schieber et Shabes veulent, pour résoudre ce problème, permettre aux arbres auxiliaires de s'accrocher de manière classique (comme dans le cas de l'arbre 1) ou comme dans l'arbre 3. On appelle *adjonction prédicative*, dépendante ou enchâssée<sup>124</sup> l'adjonction classique, où une seule adjonction est possible par nœud. L'adjonction est dite *modifiante* ou indépendante lorsqu'elle permet d'autres adjonctions au même nœud.

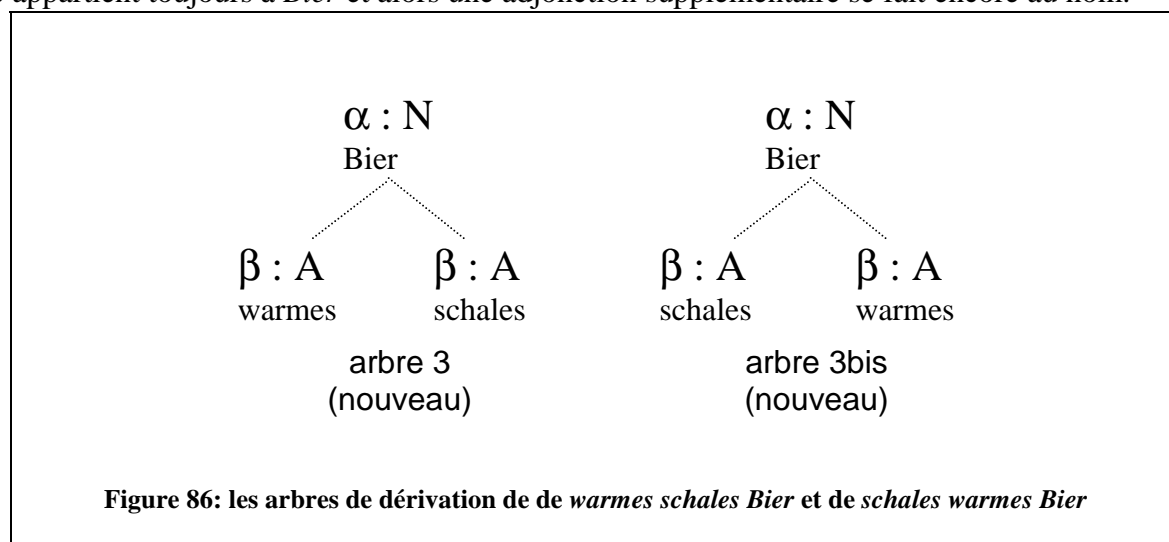
<sup>123</sup> Il s'agit en effet de l'utilisation des arbres auxiliaires pour les verbes à complément phrastique, comme par exemple les arbres des verbes à contrôle, présentés dans la section précédente.

<sup>124</sup> Pour une discussion de ces termes, voir Candito 1999:26.



Malheureusement, l'ajout d'un deuxième type d'adjonction n'est pas sans conséquence : l'arbre 3 représente ainsi aussi bien notre syntagme *schales warmes Bier* que le syntagme *warmes schales Bier* (adjectifs échangés). On perdrait donc la bijection entre les arbres de dérivation et les analyses qui en découlent. C'est la raison pour laquelle Schieber et Shabes proposent d'ordonner partiellement des arbres de dérivation de manière à distinguer les arbres comme 3 et 3bis (Figure 86). Pour cela, ils introduisent une relation d'équivalence sur les *arbres de dérivation ordonnés*<sup>125</sup> : deux arbres de dérivation sont équivalents (et représentent par conséquent la même dérivation) s'ils sont égaux à l'ordre de leur branche prêt, sauf si les deux branches représentent des adjonctions au même nœud. La première interdiction qui défend l'adjonction à un nœud pied reste en place.

Pour comprendre le fonctionnement algorithmique de ces deux sortes de dérivation, il suffit de s'imaginer que dans l'adjonction prédicative, l'arbre d'adjonction remplace le nœud sur lequel on l'adjoint, tandis que dans l'adjonction modifiante, ce nœud remplace le nœud racine de l'arbre auxiliaire. Autrement dit, le nœud racine *N* de l'arbre partiellement dérivé de la Figure 78 appartient toujours à *Bier* et alors une adjonction supplémentaire se fait encore au nom.

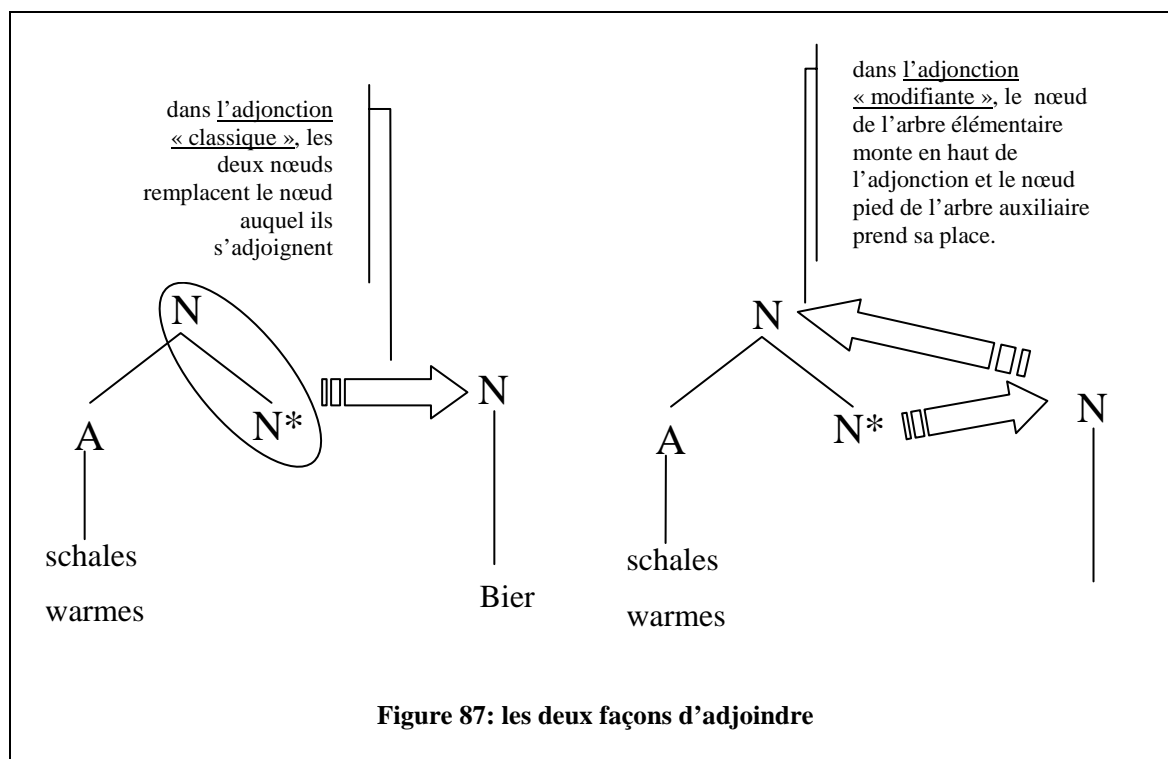


La définition que donnent Schieber et Shabes 1994 pose plusieurs problèmes, aussi bien linguistiques que formels. En effet, leur analyse des modifieurs montre un écart structurel entre l'arbre dérivé et l'arbre de dérivation. Leur but est de représenter cet écart, sans pour autant chercher à augmenter la complexité de l'algorithme liant les deux structures.

<sup>125</sup> Dans l'ancienne définition, les arbres de dérivation étaient considérés comme non ordonnés, autrement dit, les arbres 3 et 3bis étaient deux représentations équivalentes de la même dérivation.



Schieber et Shabes 1994 supposent que l'arbre de dérivation dans le cas de l'adjonction modifiante (comme par exemple dans le cas d'un nom modifié par des adjectifs) a une structure plate, c'est-à-dire que le nom gouverne directement chaque modifieur ; dans notre exemple, chaque adjectif modifie donc le nom de la même manière, et l'ordre linéaire des adjectifs dans la structure dérivée n'est pas censé jouer un rôle sémantique particulier<sup>126</sup>.

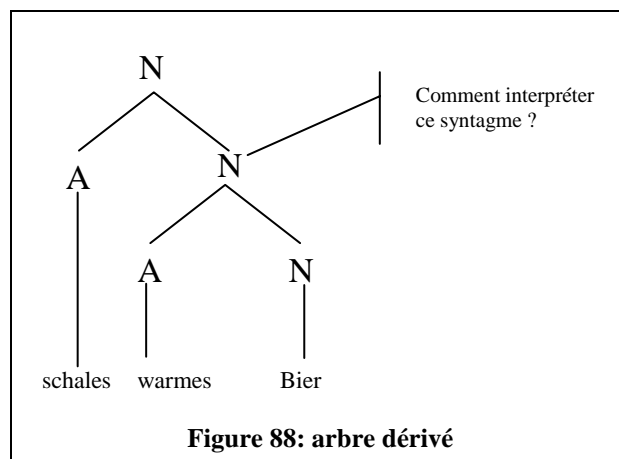


Par contre, l'arbre dérivé a une structure enchâssée, car chaque adjectif ajoute un niveau supplémentaire dans le syntagme nominal. – Or la définition de l'adjonction ne laisse pas d'alternative : l'adjonction d'un arbre non trivial<sup>127</sup> ajoute toujours un niveau supplémentaire dans la structure dérivée (voir la discussion sur la structure dérivée dans la section 2.3) et une véritable *adjonction fraternelle* (angl. *sister adjunction*), c'est-à-dire l'ajout d'une branche à un nœud existant, reste impossible. Or, si une plus grande proximité sémantique entre *warmes* et *Bier* est explicitement exclue (c'est le but de toute cette opération proposée par Schieber et Shabes 1994), on se demande comment justifier l'existence du syntagme contenant ces deux mots<sup>128</sup> (voir Figure 88).

<sup>126</sup> Pour les adjectifs intersectifs, une telle analyse est acceptable, mais pour des adjectifs encapsulants (comme *seul*, *ancien*, *présumé*, ...) il faudra trouver un autre encodage sémantique.

<sup>127</sup> Voir Joshi & Kallmeyer 1999 pour l'utilisation de l'adjonction d'un arbre élémentaire consistant seulement d'un seul nœud pied.

<sup>128</sup> D'autant plus qu'une analyse centrée autour du déterminant (analyse DP) défendrait plutôt un regroupement contraire. Dans mon mémoire de DEA, Gerdes 1998, j'ai montré que l'ordre des formes fortes et faibles (des adjectifs et des déterminants) fait que l'analyse TAG avec le déterminant comme tête est plus simple.



Trois problèmes formels dérivent de cette nouvelle définition de l'adjonction :

1. Il n'est pas clair dans quelle mesure on peut interdire l'ancienne dérivation (l'arbre n° 1 de la Figure 84). Avec quelle règle va-t-on éviter que *schales* s'adjoigne d'abord à *warmes* créant ainsi un autre arbre auxiliaire qui, à son tour, modifie *Bier* ? Va-t-on exclure l'adjonction modifiante d'un arbre auxiliaire à un autre arbre auxiliaire modifiant ? Va-t-on tout simplement dire qu'une analyse moins profonde (au niveau de l'arbre de dérivation) est toujours préférable ? Tout semble ad hoc.
2. Schieber et Shabes 1994 ne mentionnent pas les structures de traits dans leur définition. Si l'on maintient la distribution classique des traits (voir début de cette section), on peut adjoindre un arbre auxiliaire  $\gamma_1$  à un arbre élémentaire  $\gamma_2$ , sans que ces derniers unifient des traits pour autant. En effet, dans le cas d'un nom modifié par trois adjectifs, le deuxième adjectif ne partage ni la structure de traits amont ni la structure de traits aval avec son gouverneur (le nom).<sup>129</sup> Au niveau des traits, l'adjonction se fait toujours d'un arbre auxiliaire à un autre, contrairement à ce qui se passe dans la structure d'arbre de dérivation. Pour résumer, on peut dire que dans le cas de plusieurs adjonctions (modifiantes) à un même nœud, seulement la première et la dernière partagent directement des traits avec le syntagme qui les domine.

Afin de permettre le passage de traits de la tête à tous ses modificateurs, un arbre modifiant devrait toujours porter tous les traits qui peuvent éventuellement apparaître à un nœud auquel il peut se connecter, et il doit recopier tous les traits du nœud pied à sa racine. On peut ainsi compenser le manque de contact direct entre ces arbres au niveau de l'arbre dérivé. Pourtant, un tel procédé va à l'encontre de la conception de TAG, car, pour reprendre les mots de Schieber et Shabes 1994:96 « un tel encodage d'une solution exige de prendre en considération des contraintes qui passent à travers un grand nombre d'arbres, uniquement

<sup>129</sup> Un exemple concret constitue le passage du trait *genre* dans le syntagme nominal de l'allemand : si l'on considère ce trait comme un trait aval (en allemand, contrairement au français, le trait *genre* ne fait pas partie de la sous-catégorisation verbale, et les traits aval sont habituellement utilisés pour l'accord à l'intérieur du syntagme en question), alors, on est obligé de faire passer par les traits aval même des nœuds qui ne portent pas naturellement ce trait. Les adjectifs faibles de l'allemand ont des formes égales pour tous les genres, mais dans l'exemple donné, *schale* devra transmettre le trait *genre* à son nœud racine pour empêcher la phrase (i).

(i) \* *der* [masculin] *schale* [tous les genres : masculin, féminin, neutre] *Bier* [neutre]  
le bière éventé(e)

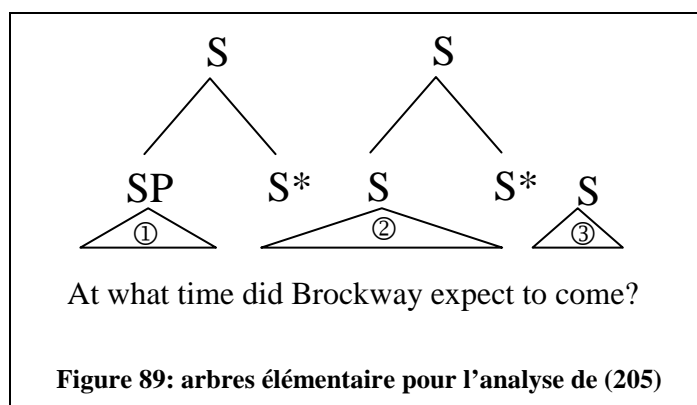
(ii) \* *ein* [masculin ou neutre] *schaler* [masculin] *Bier* [neutre]  
une bière éventé\_

La mise 'en amont' du trait *genre* de *Bier*, différente de la distribution habituelle sans être absolument contre-intuitive, ne résout pas le problème : il faudra de toute manière le mettre aussi 'en aval' pour empêcher (ii) (ou alors *ein* devra recevoir le trait *neutre* de la bière, et le faire passer à son tour à l'adjectif). Pour une discussion détaillée voir Gerdes 1998.

pour empêcher la cooccurrence d'une paire d'arbres. Il correspond pourtant mieux à l'esprit de TAG d'exprimer localement de telles limitations de caractère local<sup>130</sup>.»

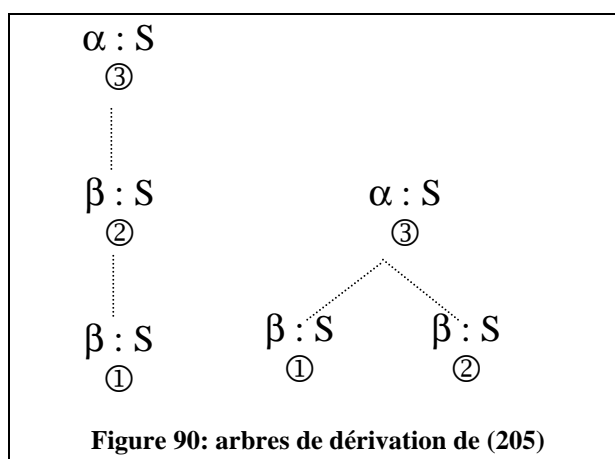
3. Même en dehors de la question sur les traits, la définition de l'adjonction modifiante de Schieber et Shabes 1994 est incomplète : pour s'en rendre compte, il suffit de regarder de plus près les exemples de restrictions d'ajout donnés par ces mêmes auteurs. Pour illustrer les difficultés, nous proposons une phrase concoctée à partir d'exemples donnés dans l'article de Schieber et Shabes 1994, munie de leurs structures arborescentes habituelles :

(205) At what time did Brockway expect to come?



La phrase est ambiguë et présente deux dérivations possibles (voir les arbres correspondants dans la Figure 90) et nous sommes intéressés par la 2<sup>e</sup> (celle qui correspond à la question sur quand arrivera *Brockway*).

L'arbre auxiliaire ① est un arbre modifiant, c'est-à-dire qu'il interdit l'adjonction au nœud pied<sup>131</sup>. L'arbre ② est prédicatif, il se comporte donc « normalement ». Shieber et Schabes



proposent pour l'anglais d'ordonner l'application des différents arbres auxiliaires de la manière suivante : arbre modifiant « non-Qu » > arbre prédicatif > arbre modifiant « Qu ».

Mais comment peut-on appliquer concrètement la proposition de Shieber et Schabes – Après l'adjonction de l'arbre ② dans ③, le nœud racine de l'arbre dérivé partiel ainsi obtenu doit

<sup>130</sup> Texte original : « Such an encoding of a solution requires consideration of constraints that pass among many trees just to limit the co-occurrence of a pair of trees. However, it more closely follows the spirit of TAGs to state such intuitively local limitations locally » (ma propre traduction)

<sup>131</sup> Sans cette interdiction on aurait une ambiguïté dans le cas de deux adjonctions modifiantes, par exemple « At what time today did Brockway... ».

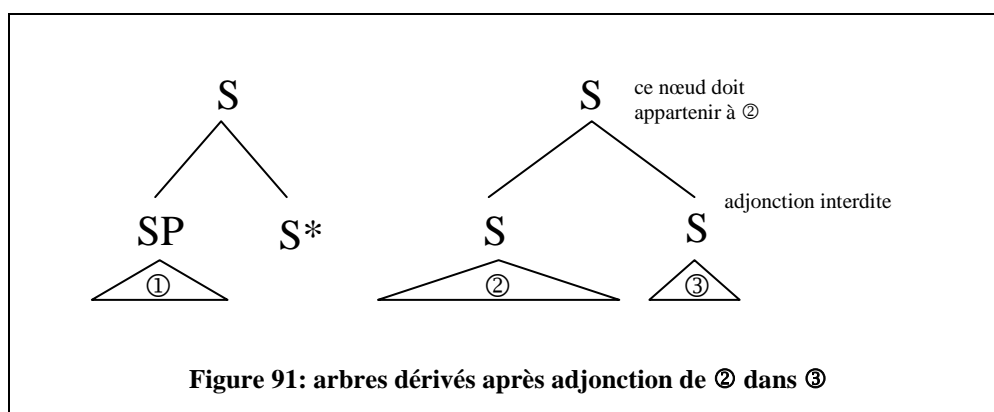
être contrôlé par ②, parce qu'il est justement prédicatif, et qu'il décide du mode, du temps, etc., c'est-à-dire qu'il gouverne cette structure partielle.

Si ce n'était pas ② qui décidait de la possibilité d'adjonction au nœud racine, on ne pourrait pas empêcher une phrase comme (206) parce que l'adjonction sans ② est possible (Cf. (207)).

(206) \* He hopes did Brockway expect to come.

(207) He hopes to come.

Nous ne doutons pas qu'on puisse définir une opération permettant cette sorte d'adjonction, mais il est sûr que l'ajout de l'arbre auxiliaire ① à l'arbre ③ dans la structure de ② et ③ ci-dessus, ne relèvera pas d'une opération très simple.



La tentative de Schieber et Shabes 1994 d'améliorer l'interprétabilité sémantique de l'arbre de dérivation nous apprend que l'hypothèse de base de l'utilisation linguistique des LTAG est fautive : la construction de la structure syntagmatique – une structure linéairement ordonnée, dont l'interprétation sera le sujet de la section suivante – est sûrement liée à la construction de la structure sémantique, mais ce lien est beaucoup moins direct que laissent entendre les analyses sémantiques de TAG. Pour obtenir une structure sémantique utilisable, trois choix se présentent.

Premier choix : On continue dans la voie de Schieber et Shabes 1994, c'est-à-dire qu'on complexifie les opérations de combinaisons des arbres élémentaires afin que l'arbre de dérivation s'approche d'une structure sémantique. L'analyse de Schieber et Shabes des structures sémantiques qui sont très différentes de la structure syntagmatique indique que ce choix n'est pas très prometteur, ou, du moins, qu'on peut faire plus simple en introduisant un formalisme en (au moins) deux modules : une première étape entre linéarisation et sous-catégorisation syntaxique et une deuxième étape entre sous-catégorisation syntaxique et structure sémantique.

Deuxième choix : On ajoute un tout autre mécanisme intervenant pendant l'analyse avec une TAG qui construit la structure sémantique, ce que fait par exemple la proposition des TAG parallèles (Schieber & Schabes 1990, Abeillé 1994). On peut ainsi complètement abandonner l'arbre de dérivation simple ou bien lui donner une autre interprétation, qui serait par exemple celle d'une dépendance purement syntaxique. L'inconvénient de cette dernière proposition est qu'elle ne correspond pas à notre vision « linéaire » de la langue, constituants – dépendance syntaxe – sémantique. L'utilité de l'arbre de dérivation (autrement dit, de la structure syntaxique) nous semble peu claire quand elle ne sert pas elle-même au calcul de la structure sémantique.

Troisième choix : On définit un post-traitement qui traduit l'arbre de dérivation en une structure sémantique. Cette voie nous semble la plus prometteuse tant que ce post-traitement est bien formalisé et que la structure initiale (l'arbre de dérivation de TAG) et la structure finale de ce processus obtiennent des interprétations linguistiques. Il serait probablement possible d'écrire une TAG dont les arbres de dérivation sont interprétables en tant qu'arbres de dépendance syntaxique. Nous ne poursuivons pas cette voie, car nous préférons une grammaire d'arbres lexicalisés dont l'arbre dérivé possède également une interprétation linguistique, que nous développons sous la forme d'une structure topologique. Pour cette raison, nous définirons dans le quatrième chapitre (troisième section) un formalisme d'arbres lexicalisés différent, (bien que voisin de TAG).

Un certain nombre de travaux ont déjà mentionné les défaillances de l'arbre de dérivation par rapport à la structure sémantique souhaitée. Par exemple, en ce qui concerne l'anglais, il est problématique d'obtenir la bonne structure pour des constructions à double contrôle (Rambow, Vijay-Shanker, Weir, 1995). Un autre problème similaire constitue le 'scrambling' de l'allemand (Becker *et alii*, 1991, 1992), nous y reviendrons dans la section 2.4.1.

Pour décrire la capacité d'une grammaire à obtenir la structure de dérivation sémantiquement bonne, Becker *et alii* 1992 ont même introduit le terme de *capacité générative dérivationnelle*.<sup>132</sup> Souvent, le manque de capacité générative dérivationnelle sert de justification pour introduire d'autres formalismes en extension des TAG : les TAG multi-composants de Weir 1988, les VTAG et DTG de Rambow (Rambow 1994, Rambow *et alii* 1995), les GAG de Candito et Kahane, 1998b et les TDG de Kallmeyer 1999.

Une remarque générale qu'on peut faire sur ces travaux est qu'ils ne tentent pas de donner un sens clair à la structure dérivée qu'ils construisent. Cette problématique sera au cœur de la sous-section suivante.

### 2.3 Interprétation de la structure dérivée

La structure dérivée de TAG est considérée comme un arbre syntagmatique X-barre. Il n'est pourtant pas nécessaire de considérer des mouvements pour l'utilisation d'une structure X-barre, comme le montrent par exemple quelques utilisations du formalisme LFG (voir section IV.4) comme de différentes analyses syntaxique réalisées à l'aide du formalisme TAG.

La S-structure des approches basées sur la Théorie du Gouvernement et du Liage (GB) est une structure qui encode des liens de types très variés entre des entités linguistiques. Elle garde par exemple la trace des mouvements d'éléments de la structure profonde (D-structure) et encode ainsi, à l'aide de la coindexation et de nœuds vides, aussi bien la sous-catégorisation d'un prédicat que la place linéaire finale des mots. Cette surcharge en information sera le sujet d'une discussion approfondie dans la section II.7.

Ce qui importe dans le cadre de notre discussion est le fait que la S-structure soit le résultat d'un processus complexe : les nœuds et syntagmes qui apparaissent, peuvent provenir de différentes étapes dans la dérivation de la structure finale (le *spell-out*) et peuvent porter différents types d'information résultant de ce processus. On peut donc dire qu'il n'y a pas de S-structure sans théorie GB.

Depuis les travaux de Frank 1992, la plupart des grammairiens TAG ont comme but d'approximer la S-structure à l'aide de la structure dérivée. Le mouvement, facteur essentiel d'une analyse GB, n'a pas lieu dans la D-structure de la phrase, qui n'est jamais calculée, mais dans les D-structures limitées à un domaine local qui est l'arbre élémentaire ancré par le

---

<sup>132</sup> Le problème de ce terme vient du fait que la dérivation fait partie intégrante de l'analyse TAG. Il serait préférable de l'inclure dans la capacité générative forte.

gouverneur de l'élément déplacé. L'arbre élémentaire d'un gouverneur est considéré comme le résultat d'un mouvement : ainsi, si un argument est déplacé, il laisse une trace à chacune des positions non finales par où il est passé. Ces positions peuvent se trouver dans l'arbre élémentaire à des positions supérieures à celle du syntagme du prédicat même, car toutes les projections font partie de l'arbre élémentaire du prédicat (grâce au domaine étendu de localité des TAG).

Une approche d'analyse GB faite à l'aide des TAG implique donc deux hypothèses très fortes : l'arbre de la D-structure et l'arbre de la S-structure peuvent être découpés de manière à ce que,

- pour un argument A d'un prédicat P donné, toutes les places, initiales, intermédiaires et finales de A se trouvent dans l'arbre élémentaire ancré par P
- et que l'adjonction et la substitution permettent d'assembler ces structures pour en faire une S-structure correcte.

Or, ces conditions représentent des objectifs difficiles à atteindre. Pour les grammairiens qui développent des TAG sous ces prémisses, le fait que la S-structure proposée par différents courant dans le cadre GB soit très variable est d'une grande aide. Par ailleurs, les restrictions imposées par le passage en TAG peuvent même aider à exclure certaines alternatives. Il est donc difficile d'évaluer le succès de ces analyses à aboutir à une S-structure correcte. On peut néanmoins affirmer l'évidence suivante : ce découpage des D-structures et S-structures en arbres élémentaires ne peut fonctionner pour des structures quelconques, car les opérations de combinaison sur les arbres élémentaires sont d'une variabilité très limitée<sup>133</sup>. Si le découpage fonctionnait pour la grammaire d'une langue entière, les restrictions qu'il impose devraient être liées à un principe important de la syntaxe de cette langue.

### 2.3.1 La structure dérivée de FTAG

À notre connaissance, le seul travail qui tente de donner à la structure dérivée une interprétation linguistique hors du cadre GB est la grammaire TAG du français (FTAG) développée par Abeillé 1991, 1993 et Candito 1999. Néanmoins, les structures dérivées proposées restent assez proches d'une approche GB en ce qu'elles encodent dans l'arbre dérivé toute une série d'informations diversifiées : par exemple la portée des adjectifs et adverbes ou la place « initiale », marquée d'un nœud vide, d'un argument relativisé.

Sur d'autres points, l'analyse FTAG s'émancipe clairement d'une approche GB ; elle écarte par exemple la notion de syntagme verbal en faveur d'une structure plate sous le nœud initial *P*. Cependant, un effet secondaire non désiré de cette proposition est qu'une phrase comme (208) n'est pas analysable, car dans cette phrase, un adverbe a été introduit entre deux branches du nœud initial *P* sans qu'il y ait un nœud *V* ou *SV* sur lequel l'arbre auxiliaire de l'adverbe pourrait s'adjoindre. La problématique ressemble à celle rencontrée ci-dessus concernant les adjectifs de l'allemand. Un analyse TAG donne aux nœuds un rôle interne au formalisme : ils servent de place d'adjonction, et le niveau supplémentaire qu'apporte chaque adjonction est en fait nécessaire pour fixer la place qu'a l'élément lexical adjoint. En ce sens, on n'est pas libre d'exiger que l'arbre dérivé soit une structure syntagmatique quelconque, basée sur des critères indépendants du formalisme.

(208) Marie écrit à Pierre souvent des cartes postales.

### 2.3.2 La structure profonde sous-jacente

Qu'un encodage du mouvement dans la structure dérivée soit réalisable ou non, l'utilité d'un tel encodage reste à démontrer. En effet, supposer une place initiale pour chaque argument dans la D-structure est une façon de coder la sous-catégorisation d'un prédicat. Mais le but de la

---

<sup>133</sup> À cette difficulté s'ajoutent les critères linguistiques sur les arbres élémentaires ainsi que le souhait d'obtenir une structure de dérivation sémantiquement acceptable (voir section précédente), ce qui impose encore d'autres restrictions sur le découpage qui sont difficiles à atteindre.

structure de dérivation est justement de capturer cette sous-catégorisation. Il serait envisageable d'essayer de maintenir ces deux sous-catégorisations. Une raison possible pour expliquer ce choix est qu'on donne une interprétation plutôt syntaxique à la sous-catégorisation qui apparaît dans l'arbre dérivé, tandis que celle qui apparaît dans l'arbre de dérivation reçoit une interprétation plutôt sémantique. Pourtant, l'approche habituelle de TAG (et de V-TAG, Rambow 1994:149) est de voir les nœuds de substitution et d'adjonction en bijection avec des théta-rôles qu'assigne un prédicat, et le formalisme TAG ne permet pas d'avoir des sous-catégorisations différentes au niveau de la structure dérivée et de la structure de dérivation.

Pour d'autres approches dans le cadre des TAG, la forme et l'interprétation de la structure dérivée semblent être d'un moindre souci. L'approche de Rambow 1994 est intéressante de ce point de vue. Rambow prône « une approche où l'on peut complètement renoncer à l'étape transformationnelle »<sup>134</sup>, mais il essaie de construire un arbre dérivé qui s'insère dans un modèle GB : en plus d'une notation X-barre, il introduit entre autres des nœuds vides et des syntagmes inflexionnels (IP) dans ses arbres. Or, ce sont justement des notations dont le statut est peu clair en dehors d'un modèle transformationnel.

Le point problématique vient surtout du fait que la validité de la structure dérivée finale ne soit jamais discutée, et qu'on ne dispose pas de critères pour la tester. La même objection est valable pour des extensions du formalisme TAG comme V-TAG (Rambow 1994) ou DTG (Rambow *et alii* 1995). Notons qu'il est difficile d'évaluer une grammaire écrite dans ces formalismes sans des critères linguistiques qui valident la structure dérivée finale.

Parallèlement aux travaux de Rambow, qui justifie ses structures dérivées en rapport au modèle GB, on trouve un autre élargissement de TAG, où cette fois-ci, une justification des structures dérivées semble très difficile : il s'agit des grammaires de descriptions d'arbres (TDG, angl. *tree description grammars*) de Kallmeyer 1999. Le point de vue habituel que l'on adopte sur les arbres, où un nœud non terminal crée une sorte d'entité (syntaxique, sémantique, prosodique etc.) à partir de ses enfants, n'a plus sa place dans les structures intermédiaires des TDG, car il ne s'agit plus d'arbres mais de descriptions d'arbres (des ensembles de contraintes qui peuvent être interprétés comme des ensembles d'arbres, voir section 3.2 pour les détails). Les descriptions d'arbres peuvent être présentées sous forme de graphes (directionnels) où un nœud peut dominer des parties de la phrase linéairement non connexes. Dans les TDG, c'est justement cette représentation graphique qui est choisie comme structure dérivée. Ainsi, l'interprétation habituelle d'un nœud non terminal comme constituant (qu'il soit syntaxique, sémantique, prosodique, ...) est exclue.

Cela dit, il faut reconnaître que ces formalismes ont la capacité générative dérivationnelle nécessaire pour analyser des phénomènes complexes d'ordre des mots de l'allemand. On pourrait se contenter de cette capacité générative comme justification de la structure dérivée, et considérer que l'arbre dérivé doit être bon parce que les bons arbres de dérivation ont été trouvés.

Néanmoins, une telle attitude serait critiquable pour deux raisons :

- Du point de vue de l'insertion du formalisme dans un modèle de la langue qui n'inclut pas seulement l'écrit, mais qui connaît encore plusieurs étapes avant la réalisation de la parole (comme la plupart des formalismes linguistiques, par exemple le modèle Sens-Texte, mais aussi GB et d'autres), il est nécessaire d'avoir accès à d'autres structures que la seule chaîne des mots entre la dépendance et la surface.
- Même si on se limitait à l'écrit, justifier les structures intermédiaires (apparaissant dans le module entre chaîne des mots et structure de dépendance) seulement par le fait que le module fonctionne, ramènerait à justifier dans les mêmes termes chaque notation qui apparaît dans une quelconque boîte noire d'analyse syntaxique. Si on avait un module

---

<sup>134</sup> texte original : « an approach in which we can dispense with the transformational step altogether », ma propre traduction, Rambow 1994:133

statistique qui donne les mêmes résultats que notre modèle TAG, il serait possible de considérer une série de chiffres apparaissant dans ce module statistique comme la structure syntagmatique. Une telle interprétation est évidemment non souhaitable et on a donc intérêt à définir des critères linguistiques indépendants du formalisme. Nous sommes ici en face d'une problématique presque philosophique, car il n'est pas clair qu'une notation puisse être considérée comme explicative juste parce qu'elle nous aide à décrire un processus. Savoir si l'on peut se contenter ou non d'une structure avec une justification purement procédurale, dépendra donc du modèle linguistique choisi et de la place que ce dernier donne (ou non) à la syntaxe.

Pour accepter une structure intermédiaire entre sémantique et surface, il faut que celle-ci soit placée dans un modèle linguistique global ; le choix de cette structure intermédiaire doit être justifiée par

- la simplicité des règles que le choix de la structure entraîne (ce qui est le critère procédural dont la plupart des formalismes se contentent)
- des critères linguistiques qui sont indépendants du processus (qui transforme une chaîne de mot ou une structure en une autre structure) et qui donne une raison d'être à chaque partie de la structure.

Il est sûrement possible de développer des critères linguistiques même pour le formalisme de grammaire de description d'arbres. Mais sans l'insertion du formalisme dans un modèle linguistique et sans le développement de ces critères, il est impossible d'évaluer la portée linguistique du formalisme. Le problème est le même que celui rencontré avec les dépendances sérielles croisées (section 2.1 ci-dessus) : si on ne fixe pas les structures linguistiques, on n'a même plus de raison de chercher autre chose que les grammaires hors-contexte.

## **2.4 TAG et l'allemand**

Dans la section précédente, nous avons vu que le formalisme TAG pose problème aussi bien au niveau de la structure dérivée qu'au niveau de la structure de dérivation. L'expressivité du formalisme est limitée, et souvent, les critères linguistiques qui ont été fixés ne peuvent pas être remplis. Pourquoi alors tenter d'écrire une grammaire de l'allemand dans un formalisme qui pose tant de problèmes ?

Deux raisons m'ont amené à essayer de construire une LTAG de l'allemand (que je nomme « DTAG »).

1. La première raison concerne la valeur que l'on donne aux exemples de phrase qui ont une analyse défailante dans un formalisme donné. En effet, il est important d'évaluer un formalisme sur l'ensemble de la langue, ou au moins sur une grande partie des phénomènes que l'on voudrait traiter. L'introduction d'un nouveau formalisme ne peut être justifié que si l'on peut montrer que ses analyses sont globalement meilleures que les analyses fournies par les formalismes déjà connus. Il ne suffit pas de donner quelques exemples de phrases dont l'analyse est insatisfaisante dans les autres formalismes mais que le nouveau formalisme peut expliquer. C'est pourtant souvent la justification principale de l'introduction d'un nouveau formalisme. Il ne faut pas oublier que le choix d'un formalisme correspond à un modèle qu'on a d'une langue. Or un modèle peut être utile même si des contre-exemples vont à l'encontre de ce modèle, c'est-à-dire même en présence de cas concrets où le modèle fait de mauvaises prédictions. Dans le domaine de la physique par exemple, ce phénomène est bien illustré par le modèle atomique de Bohr : ce modèle qui décrit les électrons qui tournent en orbite autour d'un noyau, continue à être enseigné aux étudiants de physique, bien qu'on sache depuis au moins 50 ans que ce dernier ne correspond pas parfaitement à la réalité. La raison pour laquelle on continue à insister sur ce modèle est qu'il donne de bonnes prédictions permettant d'expliquer de nombreux phénomènes physiques, et les alternatives sont de nature



beaucoup plus complexes et ne servent que pour un nombre très limité de phénomènes. Ainsi, il est essentiel d'examiner aussi bien les avantages que les limites d'une TAG de l'allemand avant de s'avancer dans le développement d'un nouveau formalisme. Ce dernier doit donc faire ses preuves en montrant qu'il propose de meilleures analyses sur l'ensemble des phénomènes à traiter. Il n'est pour autant pas nécessaire qu'il soit sans défaillances.

2. Néanmoins, d'un point de vue de l'application concrète, un nouveau formalisme peut être moins utile que l'ancien à cause du manque de ressources (lexiques, grammaires, analyseurs, etc.) et de systèmes qui utilisent des grammaires TAG, même si ce nouveau formalisme donne de meilleures analyses d'un point de vue théorique. En un sens, un formalisme qui est intégré dans un modèle implémenté dans son intégralité serait meilleur qu'un formalisme qui reste sur papier ou qui est implémenté à moitié (d'une couverture grammaticale très limitée, qui souvent ne qui ne dépasse pas « Jean aime Marie »). En ce qui concerne le formalisme LTAG, il existe tout un ensemble d'outils informatiques : on trouve d'une part des outils de développement et de maintenance d'une grammaire et d'autre part, des outils d'application de la grammaire pour l'analyse ou pour la génération. Il a été démontré qu'en théorie, pour quelques exemples d'ordre des mots, il est possible de mieux faire que LTAG<sup>135</sup>, mais pour le moment, il n'est pas possible de redévelopper ou adapter tout de suite tous les outils existants. Vive donc DTAG !!

Concrètement, il existe des TAG importants pour l'anglais (XTAG, 1995) et pour le français (FTAG, Abeillé 1991, 1993, 2002, Candito 1999), et des grammaires moins développées pour d'autres langues (le coréen et le japonais, Han *et alii* 2000 et l'italien, Candito 1999). Quant à l'allemand, il n'y a pas de grammaire de grande envergure. On peut mentionner deux travaux allant dans cette direction : Poller 1994 et Poller & Becker 1998 proposent une grammaire dans un formalisme très proche de TAG appelé LD/LP-TAG (grammaire d'arbres adjoints de dominance locale et de précedence linéaire, de l'anglais « Local Dominance/Linear Precedence-TAG », voir ci-dessous. Cette grammaire permet de laisser sous-spécifié l'ordre linéaire de quelques nœuds dans l'arbre élémentaire, et il peut ainsi regrouper plusieurs arbres TAG de l'allemand sous une seule entrée. Elle était utilisée dans des versions préliminaires de Verbmobil (un projet de traduction automatique de la parole, Wahlster 2000) mais elle avait comme défaut de traiter très peu de phénomènes grammaticaux. C'est ainsi que le groupe de génération au DFKI autour de Tilman Becker a été amené à entreprendre une compilation automatique d'une grammaire TAG à partir d'une plus grande grammaire conçue dans le formalisme HPSG qui avait été utilisée dans la direction de l'analyse dans le même projet. Cependant, produits sous de tout autres critères, les arbres élémentaires qui résultent de cette grammaire (et les arbres dérivés qu'on peut en construire) n'obéissent pas à des critères linguistiques formulés pour les TAG. La thèse de doctorat de Rambow 1994 représente le travail le plus important sur les TAG de l'allemand. En prenant comme point de départ le fait que les possibilités du scrambling de l'allemand dépassent la capacité générative dérivationnelle de TAG (Becker *et alii* 1992) Rambow 1994 propose comme solution une alternative aux LTAG classiques : la V-TAG. Il s'agit d'un formalisme dans la famille des *TAG multi-composants*<sup>136</sup>, qui permet d'exprimer en

---

<sup>135</sup> Voir VTAG, DTG, TDG, GAG. Par ailleurs, nous présentons dans la section IV.3 un formalisme d'arbres lexicalisés basé sur le modèle topologique. Néanmoins, nous sommes conscient du danger qu'une telle grammaire ne dépassera jamais le stade expérimental.

<sup>136</sup> Dans une TAG multi-composants (angl. *multi-component TAG*), une entrée lexicale ne consiste pas en un seul arbre ancré par le mot, comme c'est le cas de LTAG, mais en un ensemble d'arbres élémentaires, partiellement non lexicalisés. On peut imposer différentes restrictions sur où et comment les arbres d'une entrée se combinent, ce qui donne différents formalismes de complexité très variée (Weir 1988)

plus, des liens de dominance entre des nœuds de différents arbres d'une même entrée. Nous reviendrons sur les défauts de son approche par la suite.

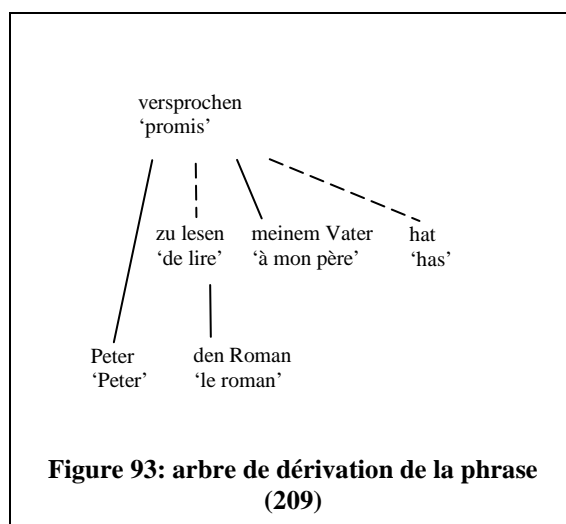
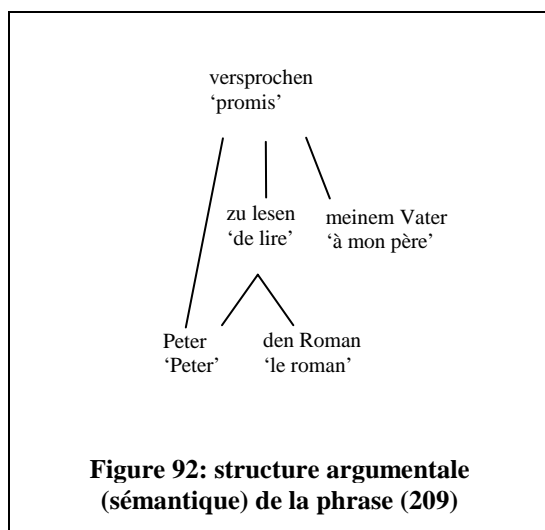
Dans les sous-sections suivantes, nous allons exposer trois exemples qui sont très problématiques pour l'expression de l'ordre des mots en allemand dans une LTAG, et discuter des conséquences de cette incompatibilité. Plus spécifiquement, nous verrons le problème du brouillage (scrambling) des arguments et des modificateurs ainsi que le problème des relatives.

### 2.4.1 Brouillage d'arguments d'un verbe

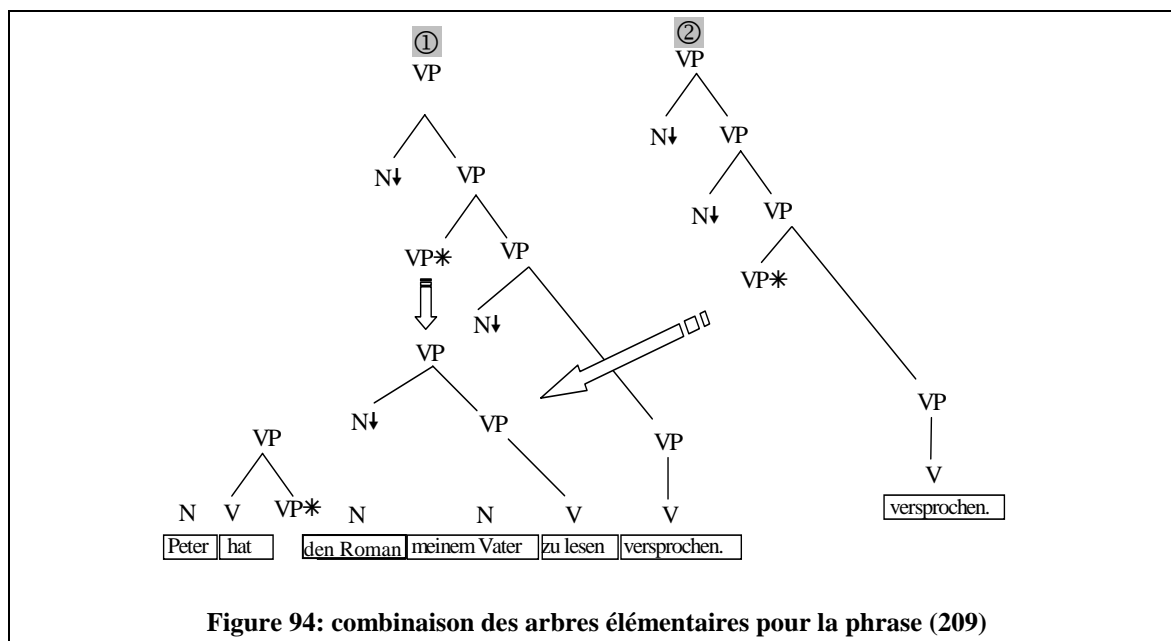
La phrase (209).a est l'exemple standard que l'on trouve pour l'illustration du brouillage (voir troisième section du deuxième chapitre). Dans une approche transformationnelle, on dirait que les deux constituants dans le Mittelfeld ont « échangé » leurs positions. Notons que cet ordre est très naturel et que c'est même l'ordre standard dans le cas de la pronominalisation des arguments (comme en (209).b).

- (209) a. Peter hat den Roman meinem Vater zu lesen versprochen.  
 Peter a le roman à mon père à lire promis.  
 'Peter a promis à mon père de lire le roman.'
- b. Peter hat ihn ihm zu lesen versprochen  
 Peter a le lui à lire promis.  
 'Peter lui a promis de le lire.'

Par définition, le verbe à contrôle *versprechen* 'promettre' assigne une théta-rôle au sujet, exactement comme le fait le verbe enchâssé (voir discussion détaillée dans la section 2.2.1 ci-dessus). Dans la Figure 92, nous esquissons la structure argumentale (sémantique). Nous considérons pour l'instant que l'auxiliaire de temps s'adjoit sur le participe comme il est de coutume en TAG, mais nous reviendrons sur cette question ci-dessous. Pour les mêmes raisons, nous ne discutons pas ici la structure interne du syntagme nominal (sans importance pour la présente discussion). Comme nous l'avons remarqué *supra*, nous devons restreindre le résultat de l'analyse TAG d'une construction à contrôle à l'un des deux liens prédicat-argument, et nous privilégions le lien entre le verbe à contrôle et son sujet. Nous montrons dans la Figure 93 l'arbre de dérivation sous ces prémisses, que nous souhaitons obtenir avec notre analyse TAG.



Nous partons de l'arbre élémentaire simple pour *zu lesen* 'lire', qui porte un nœud de substitution pour son objet. Dans cet arbre élémentaire, tel qu'il apparaît dans la Figure 94, l'agent n'est pas exprimé (voir discussion dans la section 2.2.1). L'appellation *VP* des nœuds non-terminaux n'a pas d'importance dans ce contexte et sera discutée *infra* dans la section 4.3.1.



L'arbre élémentaire ① de *versprochen* 'promis' peut s'adjoindre au nœud racine de l'infinitive *zu lesen* et par conséquent, le nœud de substitution pour le sujet apparaît au bon endroit, devant l'objet *den Roman* 'le roman'. Par contre, le nœud de substitution de l'objet indirect, prévu pour *meinem Vater* 'à mon père', demeure à l'extérieur de l'arbre élémentaire de l'infinitif. Nous n'obtenons donc pas l'ordre de la phrase (209) : avec ces arbres élémentaires, le datif apparaîtra à la droite de l'infinitif, donnant comme résultat la structure *d'intraposition* de la phrase (210).

- (210) Peter hat den Roman zu lesen meinem Vater versprochen.  
 Peter a le livre à lire à mon père promis.  
 'Peter a promis à mon père de lire le roman.'

Inversement, l'arbre élémentaire ② de *versprochen* 'promis' place l'argument au datif dans la bonne position linéaire. Or, le nœud du sujet apparaît entre l'infinitif *lesen* et son argument *den Roman*, et il ne peut être substitué dans sa position de *Vorfeld* (au début de la phrase).

Pourtant, si on veut obtenir cette structure de dérivation, nous ne pouvons pas distribuer autrement les arguments : l'ordre entre les arguments nominaux étant fixe dans l'arbre de *versprochen*, la place du nœud pied permet deux ordres des mots possibles. Soit les arguments entrent tous les deux entre l'infinitif *zu lesen* et son argument *den Roman*, soit ils restent en dehors, mais dans tous les cas, il n'est pas possible qu'un argument reste dehors tandis que l'autre y entre, comme il devrait être le cas pour expliquer l'ordre linéaire de la phrase (209).

Il est facile de vérifier que les résultats seraient identiques si on décidait de traiter *versprochen* comme un verbe à montée, c'est-à-dire si on laissait le nœud de substitution du sujet au verbe *zu lesen*. Dans ce cas, on ne pourrait pas adjoindre l'auxiliaire à son gouverneur *versprochen*. Bien entendu, il est envisageable de ne pas adjoindre l'auxiliaire au participe passé *versprochen*, mais par exemple à l'infinitif *zu lesen*. En fait, il faudrait faire en sorte que l'information de l'adjonction de l'auxiliaire passe par l'arbre de *zu lesen* (avec l'information d'accord etc.). Sans être impossible, un tel procédé fait passer l'information lexicale par des traits, ce qui va à l'encontre des principes linguistiques de TAG. Il faut donc retenir que ce n'est pas la capacité générative de TAG qui fait défaut, mais que c'est notre analyse linguistique, qui exige des structures, que TAG est incapable de produire<sup>137</sup>.

<sup>137</sup> Le formalisme des 'TAG multi composants locales aux arbres', un formalisme, décrit dans Weir 1988 est capable de traiter le cas de la phrase (209) avec de bons résultats au niveau de la dérivation, sans augmenter pour autant la capacité générative faible de TAG. Néanmoins, il suffit d'ajouter un verbe à la

Le seul choix possible pour régler le cas particulier des phrases comme (209) est de faire porter le sujet par l'auxiliaire *hat* 'a', donc de changer la structure de dérivation souhaitée. Ce choix pourrait être utile dans le but de construire une grammaire d'analyse utilisable, car les cas où le sujet n'est pas voisin de la position V2 du verbe sont plutôt rares en allemand écrit<sup>138</sup>. Cependant, trois raisons font qu'il est plus prudent de ne pas choisir cette solution :

1. La première raison est théorique : un tel choix nous éloignerait encore plus d'une interprétabilité sémantique de l'arbre de dérivation.
2. D'un point de vue plus pratique, le fait que les cas de l'éloignement du sujet de la position V2 soient plus rares ne résout pas le problème théorique sous-jacent, car ces cas, bien que rares, existent et sont complètement grammaticaux.
3. La troisième raison est encore plus pratique : l'un des objectifs de cette grammaire TAG de l'allemand est de l'inclure dans des projets existants, en particulier pour la génération. En effet, la structure de dérivation sert comme entrée du module où intervient la grammaire TAG (utilisé pour la génération). Or une modification de la structure de dérivation rend incompatible la grammaire avec le système existant qui compte l'utiliser. Nous serions donc obligés d'écrire une interface pour rendre compatible les deux versions.

## 2.4.2 Débrouillages des arguments

Plusieurs extensions du formalisme TAG standard ont été proposées afin de remédier au problème de brouillage d'arguments en allemand. En se basant sur l'extension LD/LP-TAG de Joshi 1987, Becker & Rambow 1990 et Becker 1991 ont proposé l'extension FO-TAG. L'idée est de représenter un arbre élémentaire comme d'un côté des relations de dominance linéaires entre les nœuds (angl. *linear dominance*, LD), et de l'autre côté des contraintes de précédence linéaire (angl. *linear precedence*, LP). Dans une FO-TAG, il est possible d'exprimer ces deux conditions d'une manière afin d'obtenir une plus grande liberté d'ordre des mots. Il est cependant possible que la combinaison des règles LP et LD d'un arbre soit incompatible avec une structure d'arbres. Dans ce cas, on obtient une structure arborescente ayant des branches croisées, montrant ainsi une structure de brouillage d'arguments. Pour éviter certains croisements de branches, les FO-TAG permettent d'ajouter des « contraintes d'intégrité » à des nœuds, indiquant qu'aucun croisement ne peut intervenir de l'extérieur du sous-arbre ancré par le nœud qui est marqué par la contrainte. Rambow 1994 montre cependant que de telles contraintes d'intégrité ne sont pas suffisantes pour un traitement linguistiquement adéquat du brouillage d'arguments en allemand.

C'est ainsi que Rambow est amené à proposer sous le nom de *V-TAG* une TAG multi-composante, munie de liens de dominance entre les nœuds et des contraintes d'intégrité. Ce nouveau formalisme a une capacité générative dérivationnelle adéquate dans le sens qu'il peut lier les phrases brouillées de l'allemand à un arbre de dérivation interprétable comme arbre de dépendance sémantique, c'est-à-dire que chaque prédicat est lié dans l'arbre de dérivation à ses arguments (sauf, bien entendu, dans les cas où la structure de dérivation n'est pas un arbre, comme par exemple dans le cas des constructions de verbes à contrôle). Par ailleurs, la topicalisation d'un constituant (c'est-à-dire son placement dans le Vorfeld) en V-TAG, ne peut être réalisé avec le même processus que le brouillage, et il est donc nécessaire de préserver l'adjonction habituelle des TAG comme la combinaison d'arbres élémentaires.

Rambow propose une solution très élégante en introduisant le formalisme DTG (Rambow et al. 1995), les grammaires d'arbres de dominance. Comme pour V-TAG, nous avons comme arbres

---

chaîne subordinative, atteignant ainsi le nombre de 4 verbes, pour dépasser la capacité générative dérivationnelle de ce formalisme.

<sup>138</sup> Les rares sujets où cet 'éloignement' est souvent préférable sont les pronoms négatifs comme *niemand* 'personne', car il est difficile de leur faire porter une valeur communicative de *thème*. Je renvoie à la discussion de la section IV.5.

élémentaires des ensembles d'arbres élémentaires habituels (dans le sens que les liens entre deux nœuds restent fixes – un tel lien est appelé « i-edge », arête de dominance immédiate). Ces arbres élémentaires d'une entrée lexicale sont liés entre eux par des liens de dominance (d-edges) formant ainsi une seule structure arborescente élémentaire par entrée lexicale. De tels arbres peuvent être combinés de deux manières : par *subsertion*, une généralisation de la substitution, et par *adjonction sœur*. Les liens de dominance peuvent être munie d'une restriction de subsertion ; seuls les composants étant marqués compatibles peuvent s'insérer dans un lien de dominance. De même, chaque nœud porte l'information sur les types d'adjonction sœur (gauche ou droite) qui peuvent avoir lieu au nœud. Le formalisme DTG était encore élargi sous le nom de GAG (Grammaires d'adjonction guidées par les graphes, angl. Graph-Driven Adjunction Grammars, Candito et Kahane 1998b) afin d'obtenir des graphes sémantiques souhaités et non seulement des arbres de dérivation.

Du côté de la structure de dérivation, les résultats de DTG et de GAG sont donc tout à fait satisfaisants. Pourtant, pour obtenir la bonne structure de dérivation, on reste coincé face aux structures dérivées : pendant la composition par subsertion de deux structures lexicalisées, l'ordre entre deux composants appartenant à deux structures différentes ne peut être contrôlé qu'à l'aide d'un niveau d'enchâssement supplémentaire dans l'arbre dérivé. C'est ainsi que les exemples d'arbres dérivés donnés par Rambow et al. 1995 sont tous des arbres ayant un très grand nombre de niveaux d'enchâssement intermédiaire, ce qui dépasse souvent le nombre de mots de la phrase analysée. Afin de pouvoir imposer des restrictions sur la forme des arbres dérivés, il faudrait pouvoir contrôler l'ordre entre deux constituants d'une autre manière que par des niveaux d'enchâssement. Nous reviendrons sur cette problématique lorsque nous tenterons de définir à notre tour, une extension des TAG multi-composantes, basée sur le modèle topologique de l'allemand.

### 2.4.3 Brouillage de modifieurs

Passons maintenant aux modifieurs. En allemand, les modifieurs peuvent poser problème car même avec tous les mots dans 'l'ordre standard' (voir discussion dans la section I.6), la position linéaire d'un modifieur peut empêcher une analyse correcte de la phrase.

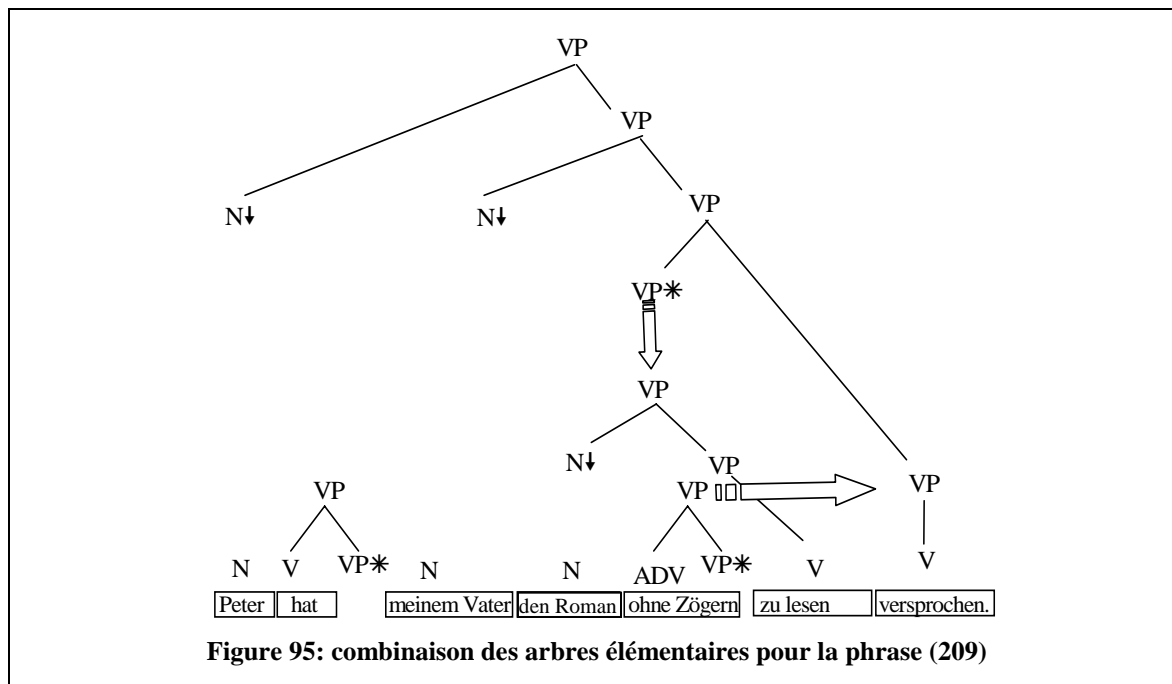
(211) Peter hat meinem Vater das Buch ohne Zögern zu lesen versprochen.

Peter a à\_mon père le livre sans hésitation à lire promis.

'Sans hésitation, Peter a promis à mon père de lire le roman.'

Dans l'interprétation préférée de la phrase (211), le syntagme prépositionnel modifie le verbe à contrôle *versprochen* 'promis', bien que ce modifieur se retrouve entre l'infinitif *zu lesen* 'à lire' et l'argument de l'infinitif *den Roman* 'le roman'.

La Figure 45 montre les arbres élémentaires correspondants. L'arbre élémentaire de l'adverbial *ohne Zögern* 'sans hésitation' ne peut se connecter avec aucun nœud de l'arbre élémentaire de son gouverneur. La seule analyse TAG possible portera donc le mauvais arbre de dérivation sémantiquement avec une interprétation d'une lecture faite avec hésitation.



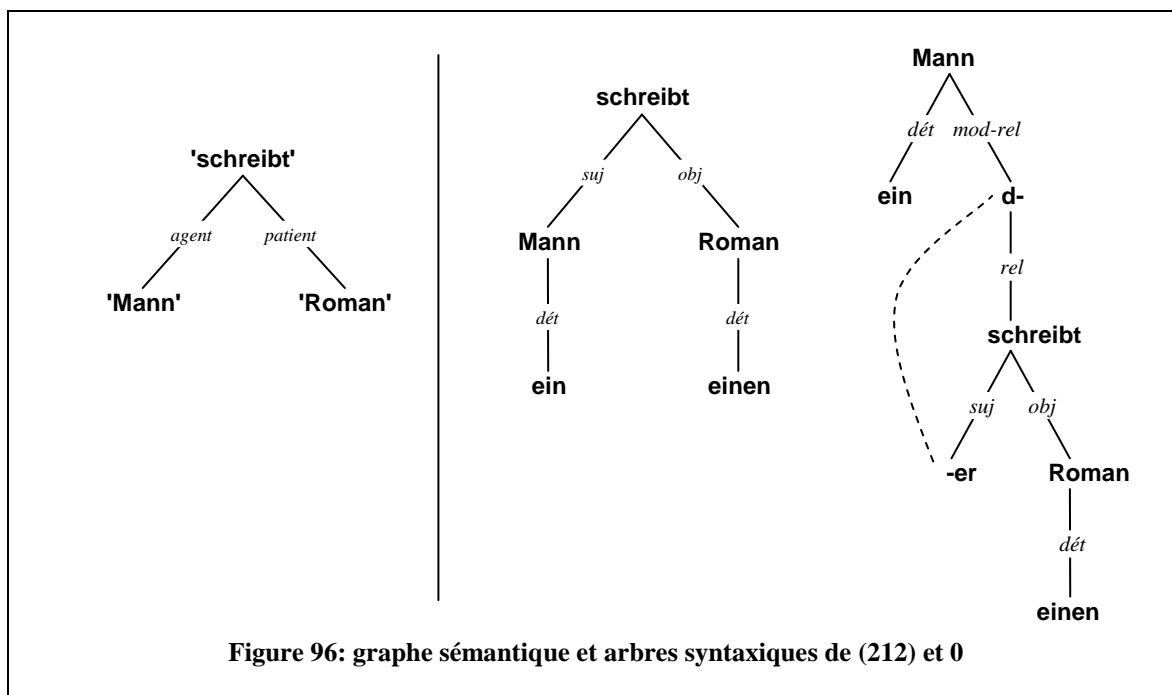
### 2.4.4 Relativement difficile

Les propositions relatives de l'allemand représentent un obstacle pour TAG : d'une part à cause de la structure interne des relatives, et d'autre part à cause de leur placement à l'intérieur de la phrase principale.

Rappelons d'abord l'analyse TAG habituelle pour les phrases relatives avec un argument relativisé :

Comme nous l'avons présenté dans la section 6 du premier chapitre, la structure prédicat-argument des phrases (212) et 0 sont identiques. Par contre, d'un point de vue syntaxique, la structure de la relative est beaucoup plus complexe que celle de la phrase simple (Cf. Figure 45).

- (212) der Mann, der einen Roman schreibt  
 l'homme qui un roman écrit  
 'l'homme qui écrit un roman'



- (213) Der Mann schreibt einen Roman  
 l'homme écrit un roman  
 'L'homme écrit un roman'

Pour ce genre de propositions relatives, l'analyse classique en TAG (Kroch & Joshi 1985 et Abeillé 1991:158) prévoit un seul arbre avec co-têtes (c'est-à-dire deux feuilles lexicalisées, l'une pour le verbe et l'autre pour le pronom relatif). Afin de pouvoir modifier le nom, l'arbre élémentaire de la relative (Figure 46, gauche) possède un nœud pied de catégorie  $N^{139}$ .

Le résultat au niveau de l'arbre de dérivation ne contient donc pas d'éléments à part pour le pronom relatif (Figure 46, droite). Si l'on fait abstraction du problème de la représentation sémantique des déterminants, ce qui est sans importance dans ce contexte, cette structure de dérivation se rapproche en fait assez de la structure sémantique proposée (Figure 45). La différence apparente qui consiste en une racine *schreibt* 'écrit' dans un cas et *Mann* 'homme' dans l'autre cas n'est pas essentielle ; ce qui importe, ce sont les liens sémantiques de la structure prédicat-argument, car il est difficile de les récupérer dans un post-calcul de la structure dérivée. Ces liens sont tous présents dans la structure de dérivation résultante.

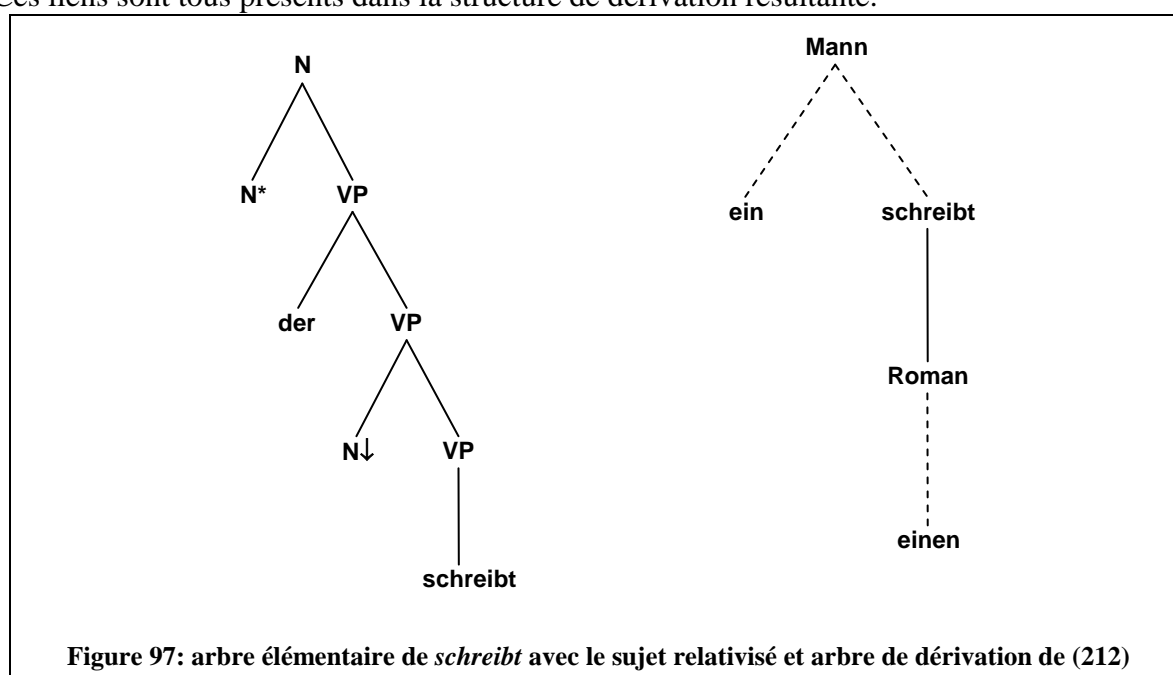


Figure 97: arbre élémentaire de *schreibt* avec le sujet relativisé et arbre de dérivation de (212)

#### 2.4.5 Modifieurs relatifs en TAG

Dans cette section, nous verrons que l'analyse TAG se détériore nettement quand ce ne sont plus les arguments du verbe qui sont relativisés mais des modifieurs. Tant que la modification a lieu au niveau de la phrase, on peut obtenir des résultats plus ou moins satisfaisants ; par contre, pour une modification d'un élément qui est lui-même substitué (ou adjoint), les résultats sont linguistiquement peu appropriés. Regardons d'abord le cas où la modification a lieu au niveau de la proposition relative (phrase (214)).

- (214) der Balkon auf dem er singt  
 'le balcon sur lequel il chante'

<sup>139</sup> La structure de l'arbre élémentaire de la relative diffère légèrement de celle proposée pour le français par Abeillé 1991. Nous justifierons ce point dans la section 4.3.3. Ce qui est essentiel pour la présente discussion est exclusivement le fait que la structure de verbe contienne directement un nœud pied et une co-tête pour le pronom relatif. Dans la version finale que nous retiendrons, le pronom relatif sera substitué comme un argument dans l'arbre du verbe. Ceci est important pour réaliser l'accord en nombre et en genre avec l'antécédent, porté par ce pronom.

Dans la section 4 du deuxième chapitre, nous avons présenté des structures sémantiques et syntaxiques contenant le syntagme présenté ci-dessus. Nous reproduisons sa structure dans la Figure 98. Il est clair que *auf dem Balkon* 'sur le balcon' n'est pas un argument du verbe *singen*. Une analyse qui inclut le syntagme prépositionnel dans l'arbre élémentaire du verbe, c'est-à-dire avec un seul arbre pour *auf etwas singen* 'chanter sur quelque chose', sera donc fautive, car elle enfreint clairement la non-compositionalité des arbres élémentaires découlant du principe sémantique.

Si on veut garder le même traitement que pour les arguments relativisés et analyser la relative en tant qu'arbre auxiliaire qui s'adjoint au nom, trois choix se présentent pour l'analyse TAG : on peut introduire la relative soit par la préposition, soit par le pronom relatif soit par le verbe, mais aucun des trois choix n'est parfait<sup>140</sup> :

1<sup>er</sup> choix : On reste le plus fidèle que possible à la structure sémantique (Figure 98) en introduisant un arbre auxiliaire pour la préposition *auf* 'sur'. La préposition propose un nœud de substitution pour le pronom relatif (voir la note de bas de page ci-dessus) et un deuxième nœud de substitution pour une phrase complète. La Figure 99 illustre un arbre élémentaire auxiliaire de ce genre. En allemand, cette phrase se distingue d'une phrase principale en ce qui concerne la position (finale) du verbe. Il faudrait donc introduire une entrée spécifique pour les phrases relatives avec modifieur relativisé. De plus, il est linguistiquement discutable d'introduire une structure spécifique pour les prépositions qui prennent comme argument un pronom relatif.

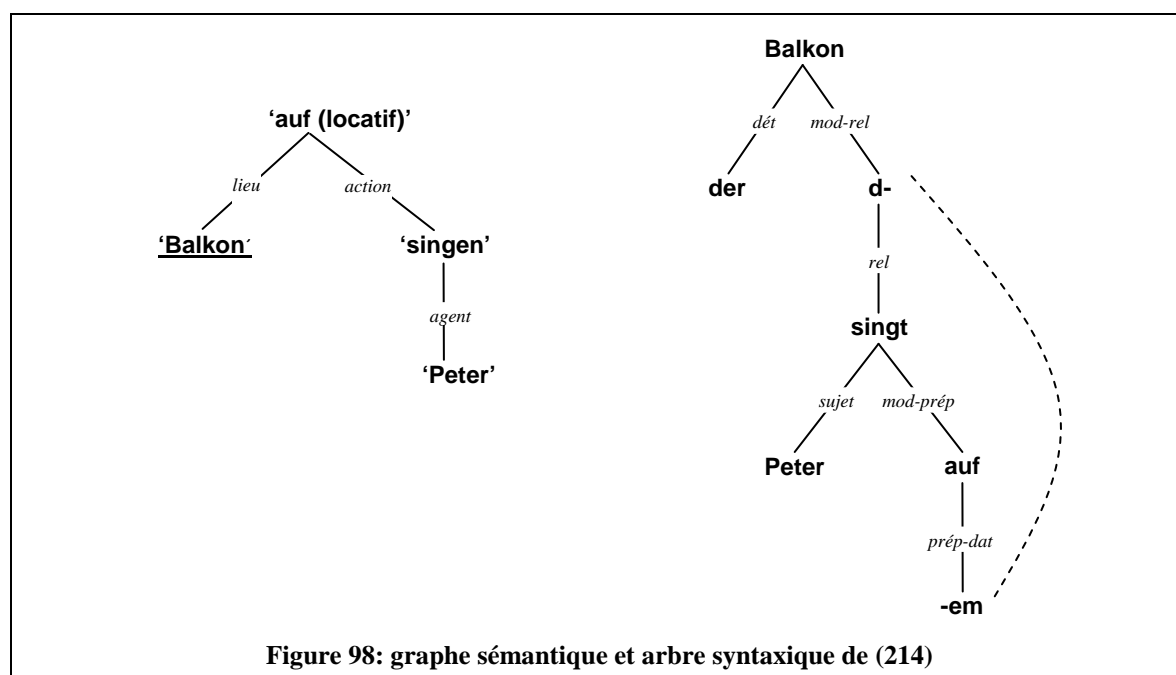
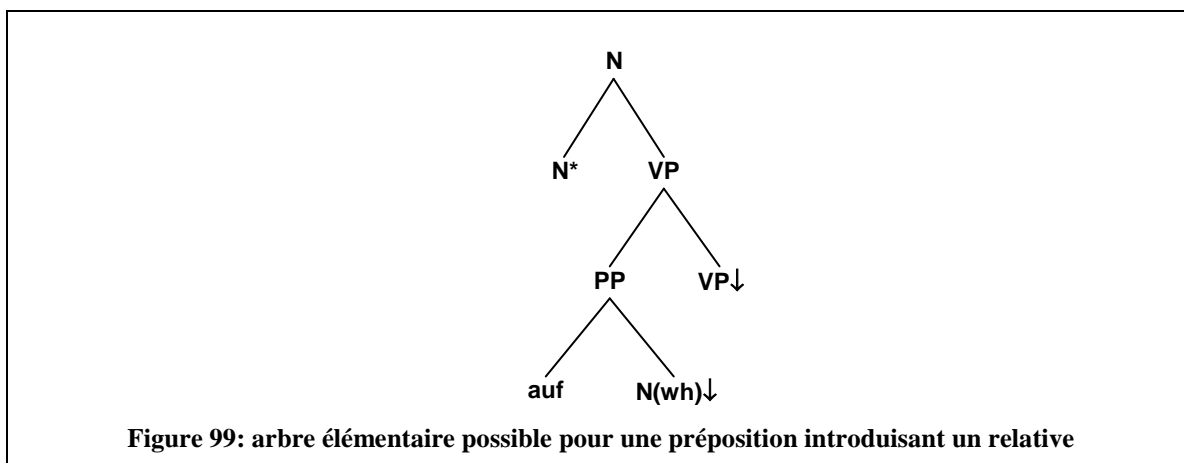


Figure 98: graphe sémantique et arbre syntaxique de (214)

<sup>140</sup> Suivant Kahane *et alii* 2000, on peut améliorer les résultats des analyses TAG des modifieurs relativisés en permettant des arbres auxiliaires avec des nœuds racine et pied de catégorie différente, qui s'adjoignent sur des nœuds avec une double catégorie, i.e. des nœuds avec des catégories amont et aval différentes. Cette approche ne dépasse pas le cadre formel des TAG classiques, car on peut considérer que tous les nœuds de la grammaire ont la même catégorie (par exemple *X*) et ils possèdent tous un trait *cat* qui peut, comme tous les traits en TAG, avoir des valeurs amont et aval différentes. Une telle approche peut être avantageuse aussi pour l'analyse d'autres phénomènes d'extraction. Pourtant, nous avons maintenu l'analyse classique (Kroch & Joshi 1985 et Abeillé 1991), car d'une part des doubles catégories nécessitent une réinterprétation des arbres élémentaires, qui ne constituent plus simplement des parties « découpées » de l'arbre dérivé final, et d'autre part, cette approche ne résout pas le problème qui est à la base de la difficulté de l'analyse TAG : le double rôle syntaxique du pronom relatif.



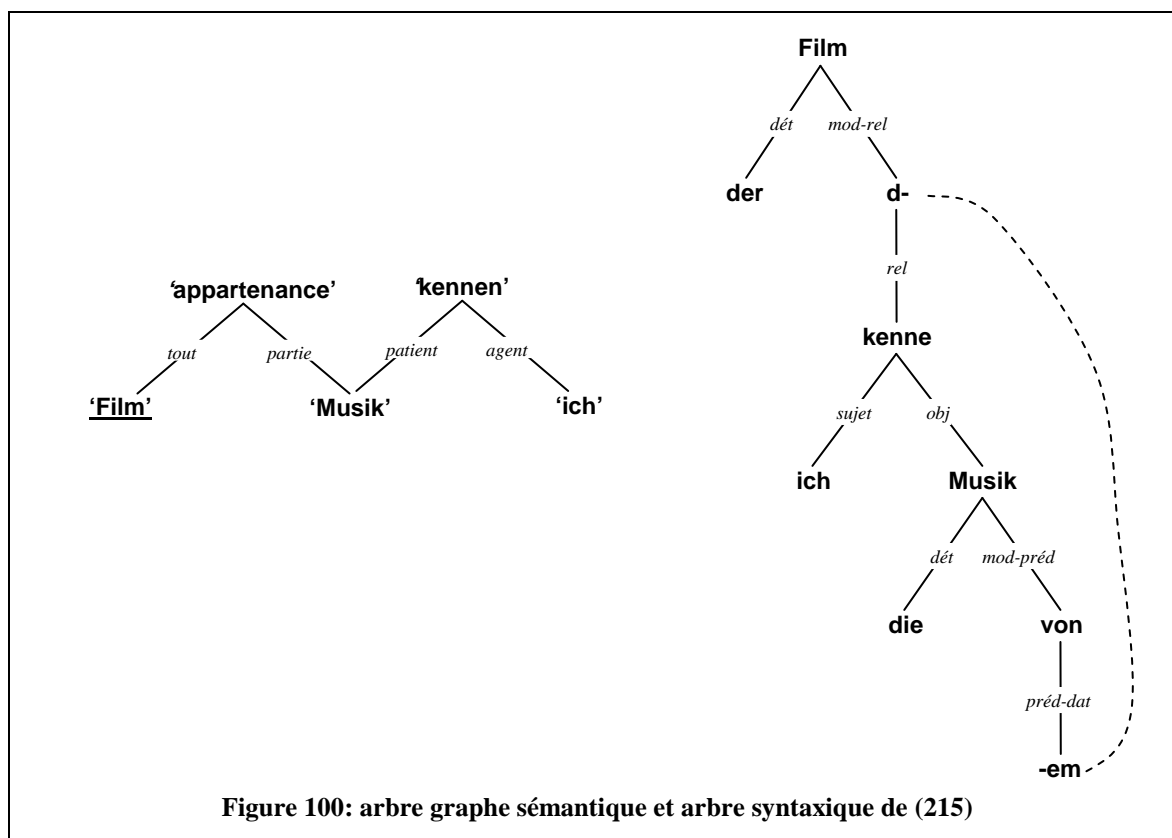


2° choix : C'est le pronom relatif qui a la tâche introduire la relative, ce qui nous rapproche de la structure de dépendance syntaxique. Cependant, cette analyse est sémantiquement moins bonne que le premier choix et est très différente de l'analyse des arguments relativisés. De plus, tout comme c'était le cas pour le 1<sup>er</sup> choix, nous avons besoin d'un arbre spécifique pour les prépositions dans des contextes relatifs, car la préposition doit être adjointe (ou substitué) à l'arbre du pronom relatif.

3° choix : Le verbe a une forme spécifique pour les cas où l'un de ses modificateurs est relativisé. Cette possibilité, que nous retenons pour la grammaire TAG de l'allemand, permet une description de la relativisation uniforme, pour les arguments et pour les modificateurs, car le verbe reste toujours la tête de la relative. Tout comme c'était le cas pour le traitement des arguments relativisés, le verbe modifie directement le nom, sans que le pronom relatif en tant qu'élément « complémentant » (ou « transférant ») apparaisse entre le nom modifié et le verbe de la relative.

Malheureusement, aucune de ces approches ne permet un traitement uniforme des prépositions ou des verbes apparaissant dans une relative, car la préposition ou le verbe en question doit avoir recours à des arbres élémentaires spécifiques s'il fait partie d'une relative. Cela nous amène à avoir des structures différentes pour par exemple une même préposition, selon la nature de l'argument qu'elle prend. Puisque aucun de ces trois choix est linguistiquement clairement préférable, il faudra choisir entre ces différentes approches en fonction des besoins concrets des structures dérivées pour l'application envisagée de la grammaire.

La situation devient pire dans le cas de la relativisation des modificateurs enchâssés. Prenons comme exemple le syntagme nominal (215) et ses structures sémantique et syntaxique représentées dans la Figure 100.



- (215) der Film von dem ich nur die Musik kenne  
 le film de qui je seulement la musique connais  
 'le film dont je ne connais que la musique'

La structure argumentale semble claire : *Musik* 'musique' est l'objet de *kenne* 'connais' et *von dem Film* 'du film' modifie *Musik*. Pourtant aucune des solutions proposées pour la relativisation des modifieurs de phrase ne marche ici : le verbe *kenne* a deux arguments nominaux, *ich* 'je' et *die Musik* 'la musique', et les critères linguistiques, formulés en vue d'une bonne interprétabilité sémantique de l'arbre de dérivation, nous obligent à faire apparaître dans l'arbre élémentaire verbal des nœuds de substitution pour ces deux syntagmes nominaux. Un quelconque arbre élémentaire doit maintenant choisir s'il veut un contact direct avec l'argument *die Musik* ou bien avec le nom modifié par la relative *der Film*.

La seule issue pour ce cas est d'introduire des arbres à part pour les verbes avec une modification relativisée d'un de ses arguments. Mais de la même manière que dans le cas précédent, cela enfonce le principe sémantique, car un tel arbre est clairement compositionnel. On obtient des arbres de dérivation fortement embrouillés et difficilement interprétables quant à leur contenu sémantique.

De manière générale on peut noter que, dès qu'on veut déplacer un élément qui est dans une position substituée, nous dépassons la capacité générative (dérivationnelle) des TAG.

#### 2.4.6 La position de la relative dans la phrase

Passons maintenant à la position des relatives dans la phrase principale, ce qui pose des problèmes qui peuvent être comparés aux problèmes de la structure interne des relatives.

La place préférée des propositions relatives en allemand n'est pas directement derrière le nom modifié, mais dans le Nachfeld du domaine principal de la phrase. Cette préférence est indépendante du placement de l'antécédant dans le Vorfeld ou le Mittelfeld, et elle est particulièrement prononcée dans deux circonstances : premièrement, la place de la relative est presque toujours le Nachfeld quand la parenthèse droite est occupée par un élément court et sémantiquement de faible contenu, comme par exemple une particule verbale (préfixe séparable) sémantiquement vide. Deuxièmement, la préférence pour un placement dans le Nachfeld est

prononcée quand la relative est plutôt longue. Dans l'exemple (216), la particule *ab* du verbe *abschließen* 'fermer à clef' doit aller dans la parenthèse droite du domaine principal (voir l'analyse dans la section II.4). Le positionnement de la relative dans le Nachfeld est clairement préférable au placement derrière le nom modifié comme dans (216).a.

(216) a. ?? Peter schließt die Tür, die ich gestern bemalt habe, ab.  
Peter ferme la porte que je hier peint ai PARTICULE

b. Peter schließt die Tür ab, die ich gestern bemalt habe.  
Peter ferme la porte PARTICULE que je hier peint ai.  
'Peter ferme la porte à clef que j'ai peinte hier'

Indépendamment des choix faits pour les arbres de la relative, nous sommes confrontés au même problème que pour les modifieurs brouillés : la modification doit avoir accès à un élément qui est substitué à l'intérieur de deux éléments lexicaux qui ont un lien direct avec le gouverneur. Dans l'exemple (216), ce sont le verbe et la particule qui entourent le nom modifié 'porte'. Comme c'était le cas pour les modifieurs, la seule issue à ce problème serait de faire des compromis au niveau des critères linguistiques des TAG et donc sur l'interprétabilité sémantique de l'arbre dérivé. Nous présenterons une analyse comprenant un tel compromis dans la section suivante (4.3).

#### 2.4.7 XTAG – FTAG – DTAG ?

La question qu'il faut se poser maintenant est de savoir si les cas problématiques de l'allemand sont vraiment plus graves que la situation anglo-romane. L'anglais et le français ne connaissent pas de brouillage, mais l'analyse avec une LTAG donne aussi pour ces langues, souvent de mauvaises structures (qu'elles soient dérivées ou de dérivation) nous obligeant ainsi à relâcher nos critères linguistiques et nos exigences sémantiques. De la même manière, le cas des relatives n'est pas si différent en allemand qu'en anglais ou en français : le syntagme (217) et son équivalent anglais<sup>141</sup> (218) posent exactement le même problème que l'exemple allemand (214), et le syntagme (219) et son pendant anglais (220) n'ont pas de bonne analyse TAG non plus. On trouve également des cas d'extraposition des relatives dans les deux langues (voir par exemple la traduction du syntagme (216)<sup>142</sup>).

(217) le balcon sur lequel il chante

(218) the balcony on which he sings  
'le balcon sur lequel il chante'

(219) le film dont je ne connais que la musique

(220) the movie I only know the music of  
'le film dont je ne connais que la musique'

Et pourtant, malgré toutes ces imperfections, les deux grandes TAG de l'anglais et du français sont utiles et constituent une partie essentielle de différents projets de recherche. Il existe une version de la grammaire XTAG de l'anglais qui possède un grand nombre d'arbres extraits automatiquement des corpus. En fait, seule cette grammaire est vraiment capable d'être utilisée

<sup>141</sup> L'exemple est correct mais la version avec la préposition en position finale ('stranded') est préférable.

<sup>142</sup> Dans le cas de l'anglais, on trouve même des « adverbiaux résultatifs » qui se comportent très similairement aux particules allemands, en ce qu'ils se placent de préférence avant la relative. Un exemple est la phrase (i).

(i) She threw the book away that I wanted to read.  
Elle jetait le livre 'away' que je voulais de lire  
'Elle a jeté le livre que je voulais lire.'

pour le *parsing*, les versions créées à l'aide de règles (ou à la main), ne donnant pas de résultats très satisfaisants sur du texte tout venant. Pour l'allemand nous avons encore moins d'espoir qu'une grammaire TAG pourrait être utile pour l'analyse, à cause de la plus grande liberté d'ordre des mots. De manière générale, il semblerait que l'allemand présente encore plus de phénomènes difficiles pour TAG<sup>143</sup>, mais l'anglais ou le français ne sont pas exempts de ces problèmes.

### 3 Les métagrammaires pour les TAG

Comme tous les formalismes lexicalisés, les grammaires TAG n'ont pas la possibilité d'exprimer des généralisations syntaxiques qui dépasse le cadre d'une structure lexicalisée. Plusieurs idées ont été proposées par différents chercheurs afin de remédier à ces lacunes.

#### 3.1 Les méta-règles

Une première proposition allant dans cette direction a été avancée par Becker 1989, 1994, 2000. Becker n'introduit pas encore un véritable méta-niveau de représentation, mais propose de coder nos connaissances syntaxiques générales dans des règles, dites des méta-règles, qui agiraient directement sur les arbres élémentaires de base. Ces arbres élémentaires "basiques" sont par exemple des arbres élémentaires où V apparaît dans un contexte "simple", avec une diathèse active, avec des arguments nominaux etc.

Une méta-règle prend comme entrée un arbre simple de ce genre et elle produit un autre arbre qui est moins « basique » : elle construit par exemple, l'arbre du passif à partir de l'arbre du verbe à l'actif. Des méta-règles de ce type sont utilisées pendant la création d'une TAG et permet de rendre la création de l'ensemble des arbres élémentaires plus simple et plus consistante. Cependant, les régularités qui sont exprimées par ces règles n'apparaissent pas en tant que tel dans les arbres générés et ne jouent aucun rôle pendant l'utilisation de la TAG pour l'analyse ou la génération. Une méta-règle capte de cette manière, notre intuition linguistique sur le lien qui existe entre deux arbres (entre l'arbre du passif et celui de l'actif par exemple). Comme les règles peuvent s'appliquer dans les deux sens, décider quels arbres sont "basiques" par rapport aux autres est un choix qui est purement linguistique. De plus, les arbres "basiques" doivent toujours être générés à la main et leurs parties ne sont donc pas le produit d'un processus computationnel. Ainsi, il s'avère qu'il est difficile d'exprimer des généralités (comme par exemple la généralité sur le sujet) qui seront partagées entre l'arbre "basique" et les arbres qui ont été produits par cet arbre "basique". Néanmoins, cette approche qui est utilisée et implémentée en tant que partie intégrante du système XTAG, permet de réduire de manière considérable le nombre d'arbres à créer à la main, et elle participe ainsi à la consistance et la maniabilité du système. Prolo 2002 a repris le flambeau du développement de XTAG à l'aide des méta-règles. Il propose d'organiser le grand nombre de méta-règles et l'ordre de leur application dans une structure hiérarchique, ce qui permet des méta-règles plus fines, comme par exemple « extraction de sujet », qui interviennent dans des changements de structures importants, comme par exemple « structure interrogative ».

---

<sup>143</sup> Il sera difficile de prouver effectivement ce point, car une preuve nécessiterait qu'on dispose d'une grammaire de l'anglais ou du français et une autre de l'allemand, qui sont plus ou moins équivalentes d'un point de vue du lexique, de la couverture grammaticale, et de la violation de critères linguistiques.

### 3.1.1 Organisation hiérarchique de l'information contenu dans les arbres élémentaires

Une deuxième proposition cherchant à aboutir à des généralisations a été l'introduction de l'idée d'hierarchie. L'idée de base d'une hiérarchie est qu'une classe porte l'information d'une certaine généralisation linguistique, présentée sous forme de restrictions portant sur des nœuds d'arbres finaux. En fonction de l'analyse linguistique adoptée, on peut regrouper l'information de deux (ou plusieurs) classes dans une nouvelle classe qui "hériterait" l'information de ses « super-classes ». L'application de cette idée fondamentale à la génération des arbres élémentaires TAG a été proposée pour la première fois par Vijay-Shanker 1992 et Vijay-Shanker et Schabes 1992. Ils proposent que l'héritage soit complètement monotone, c'est-à-dire qu'une sous-classe porte toujours toute l'information de ses "super-classes". La sous-classe peut éventuellement porter en plus, de l'information héritée d'autres super-classes ou encore de l'information propre à la classe, mais elle ne porte jamais moins d'information que ses "super-classes": l'information n'est jamais perdue ni modifiée pendant l'héritage.

Contrairement à Vijay-Shanker, Evans *et alii* 1995 proposent (dans un formalisme du nom de DATR) une approche de représentation d'une LTAG où la hiérarchie est non monotone. L'avantage d'un héritage non monotone est qu'il permet l'application d'autres règles pendant le passage de l'information d'une super-classe à sa sous-classe. Cependant, l'implémentation actuelle ne génère pas de LTAG mais une DTG (Rambow *et alii* 1999), un formalisme voisin de TAG qui a une plus grande capacité générative dérivationnelle.

Il faut noter que Vijay-Shanker et Schabes 1992 sont principalement concernés de la description de la structure d'un seul arbre, et leur proposition ne s'occupe pas de l'applicabilité d'une telle description à grande échelle. En effet, dans des descriptions de toute une grammaire, on a besoin d'exprimer des contraintes sur les croisements des classes elles-mêmes, ce qui crée une hiérarchie qui, dans un premier temps, n'est pas complètement stricte puisque d'autres outils sont nécessaires afin de simplifier l'expression des généralités linguistiques. Par exemple, on voudrait forcer (ou interdire) le croisement d'un certain nombre de classes avec d'autres classes, ce qui permettrait de ne pas exprimer tous les croisements d'une hiérarchie mais de généraliser sur les croisements nécessaires. Dans un deuxième temps, l'hierarchie complète est calculée automatiquement d'après ces règles générales de croisement.

### 3.1.2 La métagrammaire de FTAG

Dans ce même esprit, Candito 1996, 1999 propose de structurer la sous-catégorisation verbale<sup>144</sup>. Ainsi, elle présente pour l'organisation modulaire des schèmes verbaux d'une TAG du français et de l'italien, une structuration de l'information syntaxique qui constitue en un arbre TAG divisé en trois « dimensions », chaque dimension constituant une hiérarchie stricte de cette information syntaxique. Voyons en détail chaque dimension.

La première dimension correspond à une classification des arbres en famille suivant la proposition de Abeillé 1991. C'est dans cette dimension que la structure argumentale du verbe est décodée, donnant ainsi la *sous-catégorisation initiale*.

La deuxième dimension représente une étape intermédiaire où les différentes « redistributions » de la structure argumentale sont décrites ; c'est à ce niveau qu'apparaîtra par exemple, la diathèse du passif. Une redistribution peut modifier une information de la première dimension, et c'est ainsi qu'un sujet devient un syntagme prépositionnel introduit par « par » dans le cas du passif. Dans ce sens, la description n'est pas monotone. De ce fait, Candito envisage même l'application successive de plusieurs redistributions en théorie (comme par exemple le « passif impersonnel »), mais cette idée ne peut se retrouver au niveau de l'implémentation où l'on doit

---

<sup>144</sup> Les arbres des verbes constituent la plus grande partie des arbres élémentaires TAG car, comme la phrase complète est vue comme une projection du verbe, toutes les variations structurelles au niveau de la phrase se reflètent dans différents arbres élémentaires du verbe. Candito propose aussi, de manière moins détaillée, l'extension de son approche à d'autres parties du discours.

prédéfinir une multiple redistribution de ce type dans une seule classe car l'application des classes portant l'information non monotone complexifie considérablement l'algorithme sous-jacent.

La redistribution donne comme résultat l'information sur les *sous-catégorisations finales* qui sont portées par les différents arguments. Par ailleurs, elle peut causer des conséquences au niveau de la structure verbale, et faire en sorte que dans le cas du passif par exemple, l'auxiliaire de la construction passive soit introduite comme co-tête.

Enfin la troisième dimension comporte les différentes possibilités de la *réalisation grammaticale* des fonctions grammaticales finales. Ce que Candito entend par *réalisation grammaticale*, c'est par exemple la réalisation phrastique d'un sujet ou encore son extraposition. Notons que dans sa description, on ne trouve que des arguments qui ont une réalisation grammaticale. Par ailleurs, même la tête (le verbe) doit déjà contenir toute l'information nécessaire et reste inchangée grâce à la réalisation grammaticale. De plus, la troisième dimension est décalée par rapport aux deux premières dans le sens qu'il ne s'agit pas d'une hiérarchie homogène mais plutôt une collection de différentes dimensions, correspondant à une hiérarchie par fonction grammaticale à réaliser.

Passons maintenant à la manière comment le compilateur de Candito 1999 calcule les schèmes finaux. La classe qui comporte l'information d'un schème doit hériter exactement d'une classe terminale de la première, de la deuxième et de la troisième dimension pour chaque fonction de la sous-catégorisation finale. De plus, doivent être prises en compte les *contraintes de compatibilité*, c'est-à-dire les obligations et les interdictions de croisement qui peuvent être exprimées pour chaque classe. Les contraintes de compatibilité peuvent forcer le compilateur à ajouter à la description finale d'un schème, l'information contenue dans d'autres classes. Ainsi, la classe croisée obtenue de cette manière comporte exactement l'information héritée de chaque classe. Le graphe final produit par le compilateur représente donc une hiérarchie stricte dont les classes finales sont les classes qui ont été croisées automatiquement. En fait, chacune de ces classes croisées décrit des arbres dans le sens proposé par Roger et Vijay-Shanker à l'exception près que Candito 1999 ajoute la possibilité d'exprimer la catégorie finale de chaque nœud. Cependant, cet ajout, qui est pourtant nécessaire pour l'utilité concrète d'une description, a des conséquences théoriques importantes sur la description partagée de l'information dans une métagrammaire. Nous reviendrons en détail sur cette question dans la section 3.3.

### 3.1.3 Vers une métagrammaire de l'allemand

En ce qui concerne la description des arbres TAG de l'allemand (Cf. section 3.3 infra), elle repose pour une grande partie sur les descriptions proposées par Candito pour le français et l'italien. Néanmoins, nous avons rencontré un certain nombre de problèmes, ce qui explique notre choix d'implémentation d'une métagrammaire de l'allemand au moyen du compilateur de Gaiffe 2002 et non celui de Candito. En effet, on constate de manière générale que l'implémentation proposée par Candito est directement guidée par des besoins propres à la description du français et de l'italien, deux langues qui sont très proches, et on considère qu'elle n'est pas adéquate pour une description de l'allemand. Par ailleurs, on ne trouve aucune discussion sur l'applicabilité de son approche à des langues avec une structure qui est très différente<sup>145</sup>. Dans le cas d'une description d'une métagrammaire, on aimerait disposer d'outils

---

<sup>145</sup> Encore une fois, on a affaire à un formalisme qui est directement guidé par les besoins concrets en espérant que par la suite, les outils de descriptions qui ont été proposés peuvent être appliqués à d'autres langues. A mon avis, il serait préférable de proposer des descriptions dans des formalismes « sur-puissants », où toute la puissance expressive n'est pas exploitée pour la descriptions proposée et où ce n'est que dans un deuxième temps, une fois qu'un nombre important d'analyses ont été faites dans ce formalisme, qu'on peut décider qu'un certain outil n'est pas nécessaire. Et c'est seulement sur cet ensemble d'outils déjà triés comme obligatoires qu'il est possible d'avoir des considérations sur la complexité d'une implémentation.

plus généraux et plus puissants que les trois dimensions ou les restrictions sur le « croisement du type<sup>146</sup> » proposés par Candito pour son compilateur de métagrammaire. Par ailleurs, d'un point de vue théorique, il est préférable de disposer d'un formalisme plus général qui ne contient pas seulement trois dimensions, mais un nombre non défini de dimensions, et qui impose des restrictions de croisement qui sont plus générales.

Revenons à la description de l'allemand. Sa réalisation nécessite une hiérarchie qui est d'une grande complexité afin de pouvoir fixer la structure topologique de la phrase, mais cette structure est directement liée à une information qui provient d'autres dimensions, comme la dimension de la structure argumentale du verbe et celle de la réalisation des arguments. Nous avons besoin donc d'introduire une sorte de quatrième dimension pour la réalisation structurelle de la tête. En fait, comme la troisième dimension de Candito représente déjà une série d'hiérarchies qui ne sont pas obligatoirement connexes, nous aurions pu utiliser son implémentation qui permet d'ajouter une autre hiérarchie pour la réalisation du verbe, et qui peut même obliger un croisement de chaque arbre avec une telle classe. Cependant, il s'agit là d'un bricolage peu satisfaisant car l'information sur la forme générale de l'arbre, qui est contenu dans la classe de la réalisation verbale et l'information sur la réalisation d'un argument sont deux types d'informations très différents.

En fait, les raisons qui nous ont amené à chercher un autre compilateur (bien que beaucoup de concepts linguistiques de base aient été préservés) reposent surtout sur des problèmes techniques d'implémentation qui sont difficiles à résoudre. Par exemple, utiliser le compilateur de Candito impliquerait qu'on doive introduire le concept de *description partielle*, une notion qui est à la base du partage de l'information des arbres élémentaires dans différentes classes et que nous avons délibérément gardée comme abstraite jusqu'à présent.

Dans les sous-sections suivantes, nous discutons certaines notions théoriques qui sont, à notre avis, essentielles pour la conception d'une métagrammaire de l'allemand. Dans un premier temps, nous introduirons la notion de description d'arbre, qui par rapport à la définition donnée par Roger & Vijay-Shanker 1994, est augmentée d'une fonction *cat*, qui assigne la catégorie du nœud. A partir de cette nouvelle description d'arbre, il devient nécessaire de reformuler plus précisément le terme de « référent minimal » ; nous essayerons par la suite de justifier ce terme à partir d'exemples linguistiques. Par ailleurs, il faudra encore discuter de la notion de « traits polaires », des traits proposés par Gaiffe *et alii* 2002 permettant de contrôler le croisement des classes, qui sont d'une nature plus générale si on les compare aux dimensions de Candito 1999. En fait, il se trouve que malgré la généralisation rendue possible grâce aux traits polaires, l'implémentation de Gaiffe *et alii* 2002 reste défectueuse car l'ordre dans lequel les traits polaires étaient introduits, jouait un rôle dans la réalisation (ou la non réalisation) des croisements. Ainsi, pour arriver à une métagrammaire plus ou moins correcte de l'allemand, j'ai été amené à corriger l'implémentation de Gaiffe et à redéfinir la notion même des traits polaires comme des *multi-sets*. Cette nouvelle définition sera à la base de la métagrammaire de l'allemand que je présente dans les sections 4.

## 3.2 Les descriptions et leurs arbres

### 3.2.1 Notions de base et la fonction d'assignation de catégories

Un arbre peut être décrit comme un ensemble fini  $K$  de constantes muni des contraintes de dominance et de précédence (dans le cas des arbres ordonnés).

Nous appelons *description*, un ensemble fini de *contraintes* portant sur un ensemble fini de constantes  $K$ . Les contraintes sont des prédicats binaires portant sur les constantes. Roger et Vijay-Shanker 1994 proposent d'utiliser quatre prédicats binaires pour la description de ces contraintes :

---

<sup>146</sup> Si les classes A et B sont croisées, alors il faut croiser aussi avec la classe C.

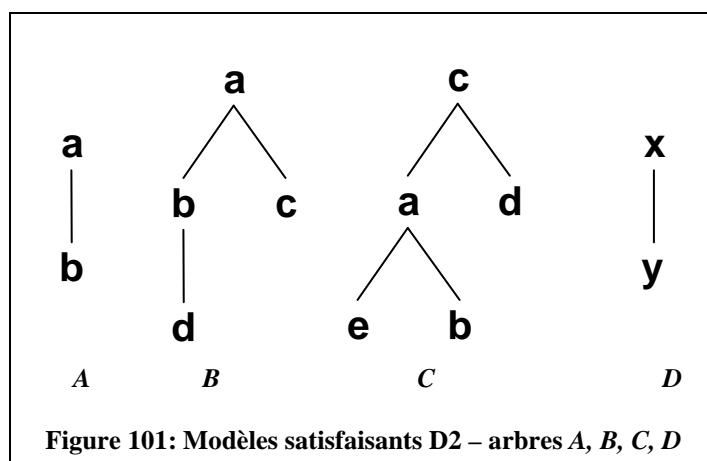
- précedence linéaire stricte (notée  $<$ ),
- parenté (directe) (notée  $\rightarrow$ )<sup>147</sup>,
- égalité (notée  $=$ )
- et dominance (notée  $\Rightarrow$  et interprétée comme chemin de parenté de longueur finie).

Roger et Vijay-Shanker 1994 s'intéressent essentiellement à la forme des arbres satisfaisant de telles contraintes. Pourtant, dans les applications réelles, nous avons besoin de créer des arbres complets avec des nœuds nommés par des catégories. C'est ainsi que Candito 1999 a décidé de donner une interprétation spéciale au trait *cat* (catégorie) attaché aux constants : ce trait, quand il est présent, assigne le nom du nœud. S'il n'est pas spécifié, on assigne un nom par défaut, composé des noms des constantes. Pourtant cette assignation par le trait *cat* n'est jamais discutée en détail, et dans la suite de cette section nous la formaliserons à l'aide d'une fonction *cat* d'assignation de catégorie des constantes, qui fait correspondre les constantes à des nœuds d'un arbre. Cette fonction *cat* n'est ni injective ni surjective, c'est-à-dire que deux constantes peuvent être envoyées au même nœud et il peut y avoir des nœuds sur lesquels ne pointe aucune constante.

Soient  $D$  une fonction d'assignation de catégorie et  $A$  un arbre dont les nœuds sont l'image de *cat*. Si chaque couple  $(n, m)$  de nœuds de  $A$  vérifie toutes les contraintes exprimées sur leur constants d'origine  $cat^{-1}(n)$  et  $cat^{-1}(m)$ , alors l'arbre  $A$  avec sa fonction *cat* est dit un *modèle* de la description  $D$ . L'ensemble des modèles d'une description  $D$  est noté  $Mod(D)$ .

Une description peut avoir soit zéro soit un nombre infini de modèles. (Elle n'a jamais un nombre fini de modèles, comme nous le verrons dans la suite.) La description est donc respectivement *non-satisfaisable* ou *satisfaisable*.

Il est possible de spécifier les propriétés qu'une description doit satisfaire pour qu'elle soit satisfaisable. Une collection possible de ces propriétés est proposée par Roger et Vijay-Shanker 1994<sup>148</sup>, où une telle description est appelée *description partielle d'un ensemble d'arbres*.



<sup>147</sup> Dans la mesure que les fils d'un nœud sont totalement ordonnés, il n'est pas nécessaire d'inclure la parenté (directe) dans l'ensemble des contraintes, car elle peut être exprimée par les autres contraintes, par exemple :  $a \rightarrow b \Leftrightarrow a \neq b \wedge a \Rightarrow b \wedge (\forall n \text{ tel que } a \Rightarrow n, n < b \vee b < n \vee b \Rightarrow n)$

Autrement dit,  $a$  et  $b$  sont différents,  $a$  domine  $b$  et aucun autre nœud ne peut s'insérer entre eux. Cette contrainte est donc juste une abréviation qui simplifie les expressions.

<sup>148</sup> cf. pp. 4-5 où ils définissent les propriétés  $T_1$  à  $T_{11}$  ainsi que des restrictions sur la finitude des ensembles en jeu, incluant par exemple l'unicité de la parenté et la non-circularité et la réflexivité de la dominance.



Par exemple, la description  $D1 = \{a \rightarrow b, a = b\}$ <sup>149</sup> n'a pas de modèle qui la satisfait, car un nœud ne peut pas être à la fois égal à un autre et son parent, et donc  $\text{Mod}(D1)$  est vide. Par contre, la description  $D2 = \{a \rightarrow b\}$  a l'arbre  $A$  de la Figure 101 parmi ses modèles. En effet, on peut définir une fonction  $cat$  allant de l'ensemble des constantes de  $D2$ , noté  $C(D2)$  à l'ensemble des nœuds de l'arbre de  $A$ , noté  $\mathbf{N}(A)$  :  $cat(a) = \mathbf{a}$  et  $cat(b) = \mathbf{b}$  et les contraintes de la description sont vérifiées car  $\mathbf{a}$  est parent de  $\mathbf{b}$ .

De la même manière, les arbres  $B$  et  $C$  de la de la Figure 101 sont des modèles de  $D2$  ayant la même fonction  $cat$  :  $cat(a) = \mathbf{a}$  et  $cat(b) = \mathbf{b}$ . De plus, l'arbre  $D$  est un modèle de  $D2$  avec la fonction  $cat(a) = \mathbf{x}$  et  $cat(b) = \mathbf{y}$ . On a aussi comme modèle tout autre arbre qui possède un nœud  $cat(a)$  parent d'un nœud  $cat(b)$ , donc forcément un nombre infini d'arbres.

On peut alors définir un *modèle référent* (Roger et Vijay-Shanker 1994:6, Candito 1999:61) comme un modèle où chaque nœud correspond à au moins une constante<sup>150</sup>. On n'a donc jamais plus de nœuds dans un modèle référent qu'il y a de constantes dans la description. Dans le cas de  $D2$  décrite ci-dessus, il ne reste que les arbres ayant deux nœuds avec le premier nœud  $cat(a)$  parent d'un deuxième  $cat(b)$ , comme par exemple les arbres  $A$  et  $D$ .

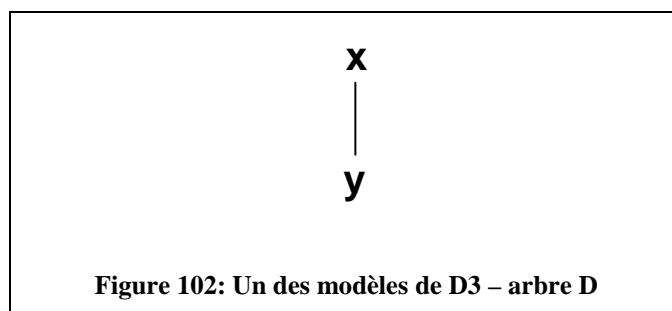
Les arbres  $A$  et  $D$  et leurs fonctions  $cat$  sont identiques, et on aimerait faire abstraction des modèles qui ne se distinguent que par le choix du nom de la catégorie. On définit donc une relation d'équivalence entre deux modèles  $A$  et  $B$  d'une représentation  $D$  de la manière suivante :  $A$  et  $B$  sont considérés équivalents s'il existe une bijection entre les nœuds de  $A$  et de  $B$  gardant toutes les contraintes de parenté et de précédence sur toutes les constantes  $cat^{-1}(\mathbf{n})$  de chaque nœud  $\mathbf{n}$  :  $A \approx_D B \Leftrightarrow \exists f : \mathbf{N}(A) \leftrightarrow \mathbf{N}(B)$  avec  $\forall \mathbf{n} \in \mathbf{N}(A) : cat^{-1}(f(\mathbf{n})) = cat^{-1}(\mathbf{n})$

Dorénavant, nous considérons donc les modèles d'une description  $D$  modulo cette équivalence. Nous notons en gras  $\mathbf{Mod}(D) = \text{Mod}(D) / \approx_D$

Il est important pour la suite de noter que deux classes d'équivalence du même modèle peuvent contenir le même arbre (mais avec des fonctions d'assignation différentes). Par exemple dans la description

$$D3 = \{a \Rightarrow b, a \rightarrow c\},$$

la longueur de la dominance entre  $a$  et  $b$  n'est pas spécifiée. Il est donc possible que cette longueur soit 0, 1 ou plus. Parmi les modèles où la longueur est 0, on trouve l'arbre  $D$  de la Figure 102, car il se peut que  $cat(a) = cat(b) = \mathbf{x}$  et  $cat(c) = \mathbf{y}$ . Cependant, parmi les modèles où la longueur est 1, on trouve encore le même arbre, car il est possible que  $cat(a) = \mathbf{x}$  et  $cat(b) = cat(c) = \mathbf{y}$ . La notion de modèle n'inclut pas juste un arbre, mais aussi la fonction  $cat$  identifiant les nœuds, et nous avons donc défini l'équivalence sur les modèles et non pas uniquement sur les arbres.



<sup>149</sup> Nous ne spécifions pas explicitement l'ensemble  $K$  et nous considérons comme ensemble  $K$  l'ensemble des nœuds apparaissant dans nos contraintes ; par exemple, l'ensemble des constantes de  $D1$  est  $\{a, b\}$ .

<sup>150</sup> Roger et Vijay-Shanker 1994 : 6 « The referents of a description  $\Phi$  are those trees ... which satisfy  $\Phi$  and interpret just those constants which occur in  $\Phi$ . » (Notre définition est un peu différente, car elle spécifie explicitement la fonction d'identification  $cat$ , ce qui n'est pas le cas dans Roger et Vijay-Shanker 1994.)

Cette correspondance nous permet de considérer les arbres  $A$  et  $D$  comme équivalents. Pour représenter la classe d'équivalence ainsi obtenue, nous utilisons le modèle dont la fonction  $cat$  envoie une constante à un nœud portant la même lettre (en gras et sans-sérifs :  $cat(n) = \mathbf{n}$ ) et, au cas où un nœud provient de constantes égales, nous prenons comme nom de catégorie les noms de toutes les constantes en jeu avec un signe d'égalité entre eux : par exemple ( $cat(a) = cat(b) = cat(c) = \mathbf{a=b=c}$ ).<sup>151</sup> Cette notation permet de toujours distinguer les classes d'équivalence par leur représentant, deux classes différentes ayant toujours un représentant différent. Par abus de langage, nous continuerons d'utiliser le terme *modèle* pour le représentant de la classe d'équivalence.

L'ensemble  $\mathbf{Mod}(D)$  des modèles ne contient donc que des arbres ayant comme nœuds des lettres provenant des constantes de  $D$ , seules où avec des signes d'égalité entre elles, ce qui nous donne un ensemble fini de nœuds possibles pour un modèle dans  $\mathbf{Mod}(D)$ . Et comme il y a un nombre fini de combinaisons en arbres pour un ensemble fini de nœuds, on a un nombre fini de modèles référents par description. Suivant Roger et Vijay-Shanker 1994, nous notons l'ensemble de modèles référents  $L_D$ .

Seul l'arbre  $A$  de la Figure 101 est un modèle référent de  $D2$ . La description

$$D4 = \{a \Rightarrow b\}$$

possède deux modèles référents (Figure 103), l'un où  $a$  et  $b$  nomment le même nœud (le chemin de parenté entre  $a$  et  $b$  est de longueur 0), et l'autre où  $a$  est parent de  $b$ .

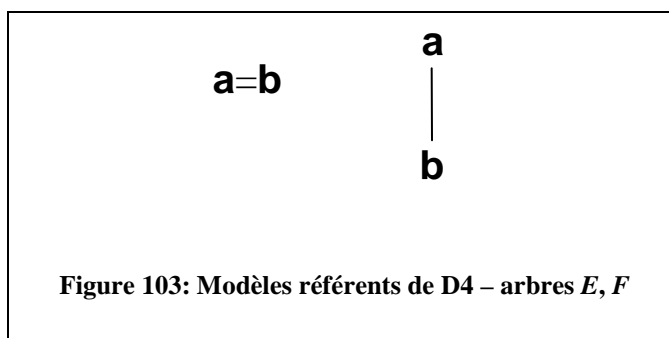


Figure 103: Modèles référents de  $D4$  – arbres  $E, F$

Les modèles référents ne possèdent ainsi pas toujours un nombre minimal de nœuds, ce qui revient à dire qu'on ne peut pas forcer deux constantes à désigner le même nœud seulement avec la notion de modèle référent.

On peut maintenant faire l'inverse et donner une description  $D$  qui décrit complètement un arbre  $A$ , ou, autrement dit, dont le seul modèle référent est  $A$  : il s'agit d'une description avec une constante pour chaque nœud de l'arbre ; on stipule que chaque constante est différente de toutes les autres, et on trouve la contrainte de dominance ou de précédence existant entre chaque paire de nœuds de l'arbre correspondant. Pour  $n$  constantes, on a donc  $n^2$  contraintes, incluant la réflexivité de la dominance qui fait que chaque constante se domine elle-même :  $a \Rightarrow a$ . Nous appelons une telle description, une *description complète d'un arbre*.

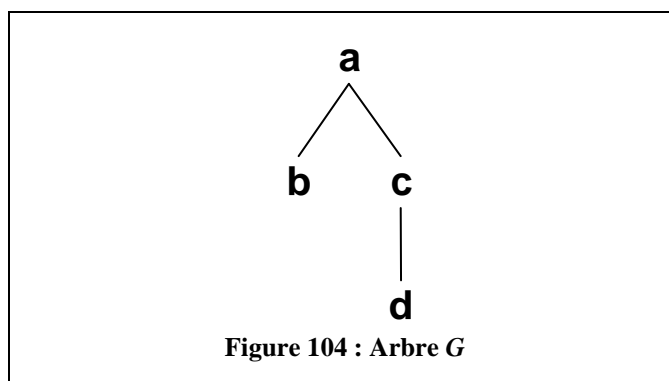


Figure 104 : Arbre  $G$

<sup>151</sup> Cette notation est équivalente à l'utilisation du underscore ( $\mathbf{a\_b\_c}$ ) dans l'implémentation du

L'arbre  $G$  de la Figure 104, par exemple, possède 4 nœuds et on a donc  $4^2 = 16$  paires de nœuds, incluant les paires réflexives comme  $(\mathbf{a}, \mathbf{a})$ . La description complète de l'arbre  $G$  est l'ensemble des constantes  $\{a, b, c, d\}$  avec les contraintes  $\{a \neq b, a \Rightarrow b, a \neq c, a \Rightarrow c, a \neq d, a \Rightarrow d, b \neq c, b < c, b \neq d, b < d, c \neq d, c \Rightarrow d, a \Rightarrow a, b \Rightarrow b, c \Rightarrow c, d \Rightarrow d\}$ . On voit une certaine redondance dans le fait qu'on exprime la différence entre deux constantes même si l'une précède l'autre. Cela provient de l'asymétrie des contraintes de Roger et Vijay-Shanker 1994 : la dominance permet l'égalité tandis que la précédence l'exclut. Il serait préférable de permettre, de manière symétrique, à la précédence d'être aussi réflexive ( $\leq$ ), ce qui faciliterait également la description linguistique comme nous verrons par la suite dans la section 3.3. Nous restons pourtant avec ce choix en conformité avec les implémentations existantes.

Pour passer d'une description satisfaisable à une description complète d'un arbre, il faut donc

1. introduire l'auto-dominance de chaque constante ( $a \Rightarrow a, \dots$ ) –  $n$  contraintes.
2. introduire l'égalité ou l'inégalité entre chaque paire de nœuds –  $\binom{n}{2}$  contraintes.
3. introduire la précédence ou la dominance entre chaque paire de nœuds –  $\binom{n}{2}$  contraintes.

Ce qui nous donne les  $\binom{n}{2} + \binom{n}{2} + n = n^2$  contraintes

Pour passer de la description  $D3 = \{a \Rightarrow b, a \neq c\}$  à une description complète, nous avons plusieurs possibilités. – Il faut d'abord spécifier l'égalité entre les nœuds : on sait déjà que  $a$  est différent de  $c$ , mais  $b$  peut être ou non égal à  $a$ , et  $b$  peut être ou non égal à  $c$ . Ensuite, en cas d'inégalité entre  $b$  et  $c$ , il manque encore la spécification de la contrainte les liant, qui peut être la précédence dans les deux sens ou la dominance de  $b$  par  $c$  (l'inverse étant exclu par le lien de parenté entre  $a$  et  $c$ ).

On a donc cinq modèles référents pour  $D3$ , représentés dans la Figure 105 :

H :  $\{a=b, a \Rightarrow c, b \Rightarrow c, a \neq c, b \neq c, a \Rightarrow b, a \Rightarrow a, b \Rightarrow b, c \Rightarrow c\}$

I :  $\{a \Rightarrow b, a \Rightarrow c, b \Rightarrow c, a \neq c, b=c, a \Rightarrow b, a \Rightarrow a, b \Rightarrow b, c \Rightarrow c\}$

J :  $\{a \Rightarrow b, a \Rightarrow c, b < c, a \neq c, b \neq c, a \Rightarrow b, a \Rightarrow a, b \Rightarrow b, c \Rightarrow c\}$

K :  $\{a \Rightarrow b, a \Rightarrow c, c < b, a \neq c, b \neq c, a \Rightarrow b, a \Rightarrow a, b \Rightarrow b, c \Rightarrow c\}$

L :  $\{a \Rightarrow b, a \Rightarrow c, c \Rightarrow b, a \neq c, b \neq c, a \Rightarrow b, a \Rightarrow a, b \Rightarrow b, c \Rightarrow c\}$

Bien que les modèles référents ne possèdent pas toujours un nombre minimal de nœuds, ils sont minimaux dans le sens où nous avons spécifié une contrainte entre chaque paire de constantes, l'égalité (cas H et I), la précédence (cas J et K) ou la dominance (cas L) selon les cas.

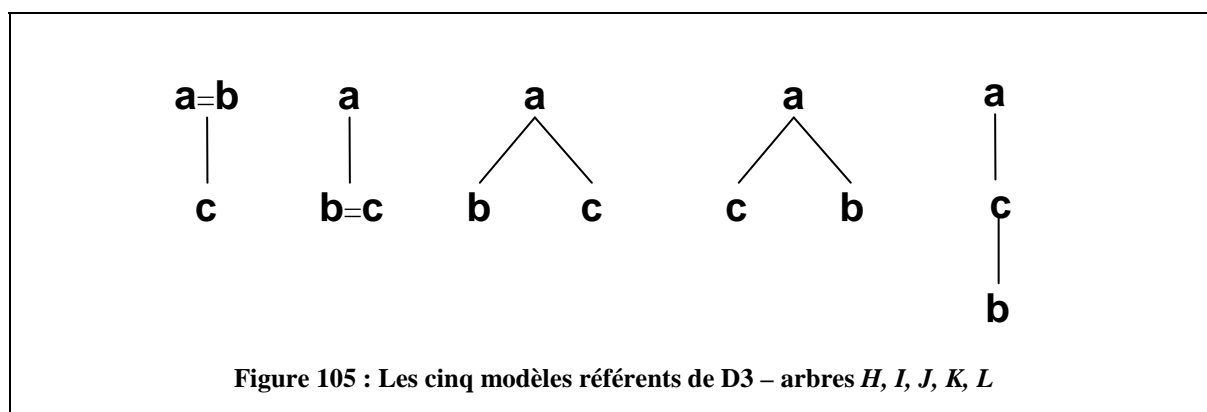


Figure 105 : Les cinq modèles référents de  $D3$  – arbres  $H, I, J, K, L$

On peut maintenant introduire le terme de *modèle minimal* défini comme modèle d'une description avec un nombre minimal de nœuds.

L'ensemble des modèles minimaux est toujours un sous-ensemble des modèles référents. Par conséquent, pour la description D2, son seul modèle référent est aussi son seul modèle minimal ; pour la description D3, seuls les arbres *H* et *I* sont minimaux ; pour la description D4, seul l'arbre *E* est minimal.

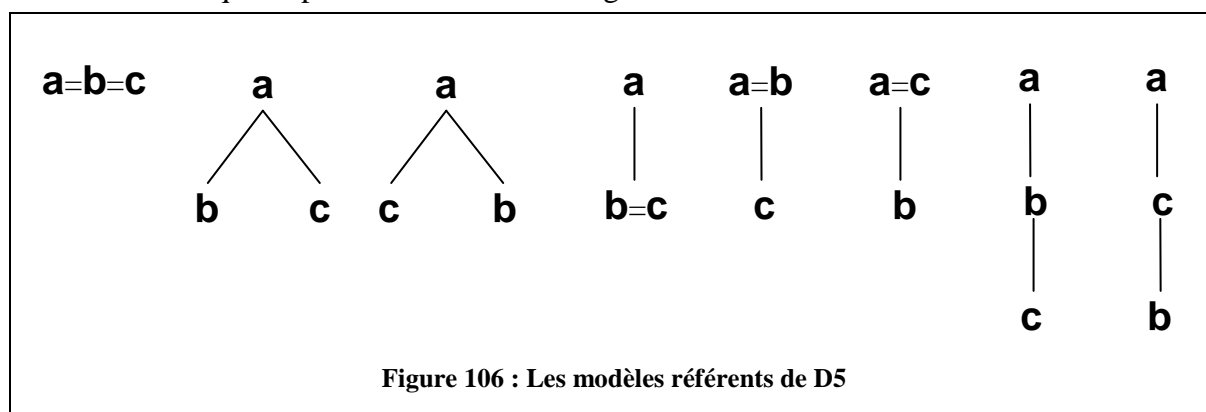
Roger et Vijay-Shanker 1994 introduisent le terme de *référent minimal* pour désigner un arbre dans lequel chaque dominance entre deux constantes, dont la distance n'est pas spécifiée, est prise comme dominance de longueur zéro, et donc comme égalité entre les nœuds. Cette définition est différente de la mienne, car un modèle minimal n'est pas forcément un référent minimal, l'inverse étant vrai. Par exemple, dans la description D4, la longueur de dominance entre *a* et *b* n'est pas spécifiée. Le référent minimal doit donc avoir l'égalité entre *a* et *b*. Le seul référent minimal est donc l'arbre *H* de la Figure 105. De même pour D3, *H* est le seul référent minimal.

Nous verrons dans la section suivante que la notion de référent minimal n'est pas compatible avec l'ajout de traits et en particulier avec l'idée d'une fonction d'identification *cat* libre, c'est-à-dire une fonction d'identification dont la valeur peut être spécifiée sur le nœud dans la description.

Notons encore que Candito 1999:75 définit différemment ce même terme de *référent minimal* : il renvoie aux « arbres qui mettent en jeu le moins de nœuds possibles, une fois fixés le type (précédence ou dominance) et le sens de la relation existant entre chaque couple de nœuds<sup>152</sup>. » Cette définition est redondante car elle est clairement équivalente à la définition du modèle référent, qu'elle utilise par ailleurs.

La version originale du générateur de schèmes de Bertrand Gaiffe (Gaiffe *et al.* 2002) avait encore un autre concept du référent minimal qui n'est jamais explicitement introduite :

Pour la description D3 il ne donnait que l'unique modèle minimal, l'arbre *E*, tandis que pour la description D3, il ne trouvait ni les deux modèles minimaux *H* et *I*, ni tous les modèles référents, mais il trouvait les quatre premiers arbres de la Figure 105.



Pour la description

$$D5 = \{a \Rightarrow c, a \Rightarrow b\},$$

le générateur ne donne ni les huit modèles référents de la Figure 106, ni le seul modèle minimal où **a=b=c**, mais seulement les trois premiers arbres des huit modèles référents.

Nous étions donc obligé à revoir l'implémentation de manière à ce que le générateur produise les modèles minimaux suivant notre définition<sup>153</sup>.

<sup>152</sup> M.-H. Candito veut probablement dire *constantes* et non *nœuds*.

<sup>153</sup> Nous avons donc été amené à ajouter un filtre optionnel au compilateur, qui a permis d'obtenir uniquement les modèles minimaux. Les changements que nous avons effectués sont basés sur le principe d'un filtre et ne sont donc pas les plus efficaces, car nous calculons d'abord les modèles référents avant de

### 3.2.2 Les traits

Les constantes dans les descriptions d'arbres peuvent être augmentées d'un système de traits, compliquant considérablement le calcul des modèles pour une description. Nous notons de manière abrégée les traits associés aux constantes directement entre crochets derrière la constante (dans une des contraintes qui porte sur la constante), par exemple  $a \rightarrow b[\text{bidule}=\text{B}]$ <sup>154</sup>.

Pour que deux constantes fassent référence au même nœud, il faut que leurs structures de traits s'unifient. Autrement dit, les traits peuvent interdire des arbres parmi les modèles référents, mais ils ne peuvent clairement pas en rajouter. Les modèles référents  $D'$  d'une description d'arbres avec traits forment un sous-ensemble des modèles référents de la même description  $D$  sans trait<sup>155</sup>. Par contre, ceci n'est pas vrai pour les modèles minimaux. Si les modèles minimaux de  $D$  sont interdits par les traits, d'autres modèles référents de  $D$  seront considérés comme minimaux.

On voit maintenant que la définition de référent minimal de Roger et Vijay-Shanker 1994 ne s'applique pas dans une description augmentée de traits, car deux constantes dont la distance de dominance n'est pas spécifiée, ne peuvent pas identifier le même nœud, si les deux constantes ont deux structures de traits qui ne s'unifient pas. Si on augmente par exemple la description  $D3$  d'un trait *bidule* de manière à bloquer l'unification entre  $a$  et  $b$ , alors la description  $D3'$  n'a plus de référent minimal du tout :

$$D3' = \{a[\text{bidule}=+] \Rightarrow b[\text{bidule}=-], a \rightarrow c\}$$

Dans les générateurs de M.-H. Candito et de B. Gaiffe, le trait *cat* joue un rôle particulier, très différent des autres traits, sans que les implications de cet ajout soient prises en compte dans toute leur ampleur théorique.

Pour l'application des descriptions d'arbres à la génération de schèmes de TAG, il est souhaitable de pouvoir exprimer sur les constantes quelle étiquette portera le nœud qu'elle identifie. Le trait *cat* remplit cette fonction. Si la constante  $n$  porte le trait  $[\text{cat}=\text{C}]$ , cela exprime que  $\text{cat}(n) = \mathbf{C}$ , ou, autrement dit, que le nœud identifié par  $n$  dans le représentant de la classe d'équivalence du modèle en question doit être le nœud  $\mathbf{C}$ . Une telle explicitation écrase la fonction *cat* par défaut définie ci-dessus, qui choisit comme nom de nœud l'énumération de toutes les constantes dont le nœud est l'image sous *cat*. Le problème est que cette explicitation

filtrer les modèles qui utilisent un nombre minimal d'arbres. Si on ne cherchait que les modèles minimaux, on pourrait faire mieux déjà pendant l'exploration de différentes possibilités des modèles, tout en excluant celles qui mettent en jeu un nombre plus grand de nœuds. Le filtre opère é en deux étapes : d'abord il compte les nœuds de chaque modèle proposé et établit le nombre minimal de nœuds. Puis, il compare mutuellement tous les arbres qui ont un nombre minimal de nœuds, et il ne garde qu'un seul modèle s'il se trouve face à des modèles identiques.

<sup>154</sup> Notons que dans une description dans le générateur de Bertrand Gaiffe, on peut faire porter une contrainte sur une constante qui n'est pas déclarée explicitement dans la même classe de la métagrammaire. La contrainte sera ignorée tant que la constante n'est pas déclarée dans une des classes qui forment la classe finale. Seulement si on veut ajouter un trait à une constante, alors il faut la déclarer. Les résultats du générateur donnés ici supposent que toutes les constantes sont déclarées. Les résultats sont différents, et parfois meilleurs, sans la déclaration des constantes, bien que formellement, il ne devrait pas y avoir de différence entre une constante non déclarée et une constante sans traits. Quelques problèmes pourraient sans doute être résolus en évitant de déclarer les constantes, sauf dans les classes où on est sûr que la constante sera réalisée comme nœud à part. Pour une approche plus générale, on ne pourrait pas éviter le calcul des modèles minimaux.

<sup>155</sup> Soit  $D$  une description sans traits avec ses modèles référents  $M = \{M_1, M_2, \dots, M_n\}$  (l'ensemble  $M$  peut être vide). Soit  $D'$  la même description augmentée de quelques traits. Les modèles référents de  $D'$  forment un sous-ensemble (éventuellement vide)  $M'$  de  $M$ . En effet, si  $M_1'$  est un modèle référent de  $D'$ , chaque nœud de  $M_1'$  correspond à au moins une constante  $c'$ , et si  $c$  est la constante  $c'$  sans trait,  $c$  fait partie de  $D$ . Ceci est vrai pour tous les nœuds, et il y a donc aussi un modèle  $M_x$  dans  $M$  correspondant à  $M_1'$ .

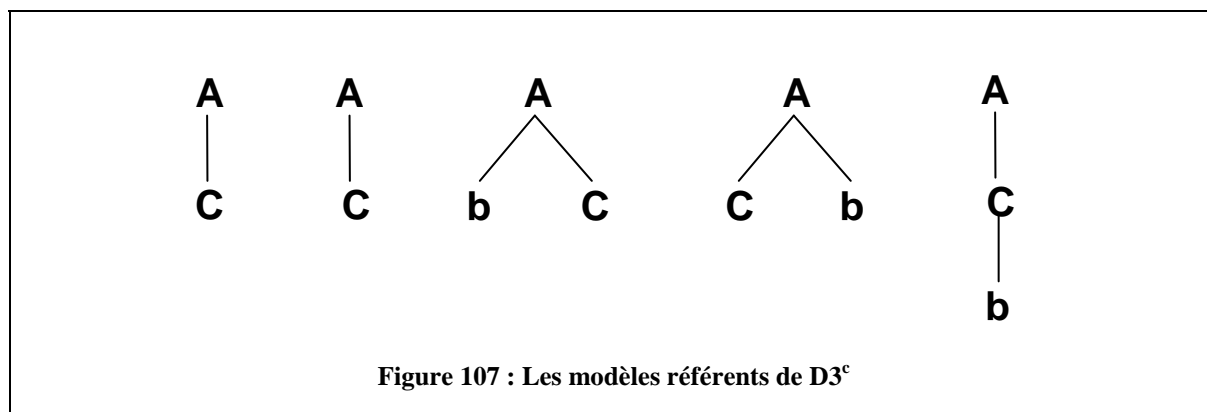
du nœud à utiliser ne donne plus la garantie que deux classes d'équivalence de modèles ne soient représentées par le même arbre.

On peut augmenter la description D3 du trait *cat* de manière à ce que deux classes d'équivalence soient représentées par le même arbre :

$$D3^c = \{a[\text{cat}=A] \Rightarrow b, a \rightarrow c[\text{cat}=C]\}$$

D3<sup>c</sup> a comme modèles référents les arbres présentés dans la Figure 107.

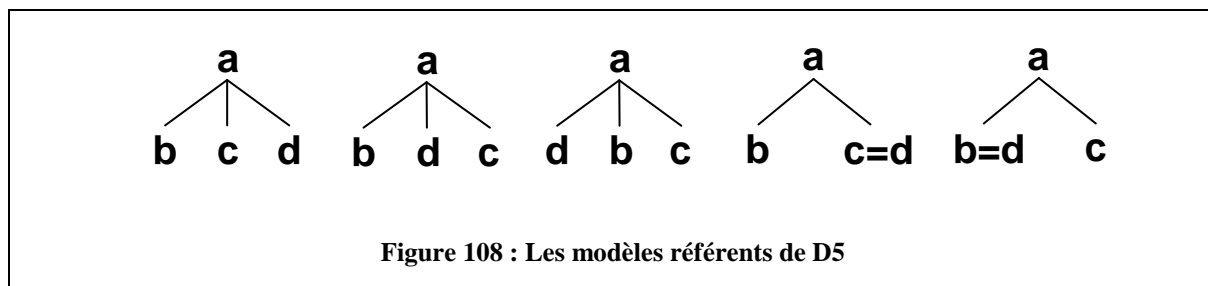
Les modèles *H* et *I* de la description D3, ne se distinguent plus par leurs arbres mais seulement par leur fonction *cat* sous-jacente.



Analysons maintenant un autre exemple montrant la différence entre les traits habituels et le trait *cat* :

$$D5 = \{a \rightarrow b, a \rightarrow c, a \rightarrow d, b < c\}$$

Un nœud *a* a trois enfants, dont on sait seulement que *b* précède *c*. La Figure 108 montre les cinq modèles référents de D5. Seuls les deux derniers modèles, les deux arbres à trois nœuds, sont minimaux.



Supposons qu'on augmente D5 des traits de la manière suivante :

$$D5' = \{a \rightarrow b[\text{bidule}=B], a \rightarrow c[\text{bidule}=C], a \rightarrow d[\text{bidule}=D], b < c\},$$

Dans ce cas, les deux modèles à trois nœuds ne sont plus possibles, car les traits *bidule* de *c* et *d* et de *b* et *d* ne s'unifient pas. Les trois premiers modèles sont donc devenus minimaux.

Si les traits ne gênent pas l'unification de constantes comme dans,

$$D5'' = \{a[\text{bidule}=A] \rightarrow b[\text{bidule}=B], a \rightarrow c[\text{bidule}=D], a \rightarrow d, b < c\},$$

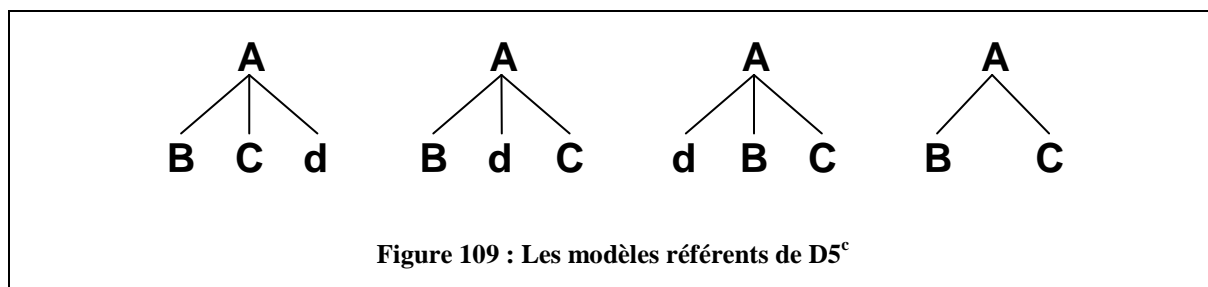
où *d* ne porte pas de traits, alors on obtient les mêmes modèles référents que dans D5, avec les nœuds augmentés des traits.

Augmentons maintenant D5 du trait *cat* :

$$D5^c = \{a[\text{cat}=A] \rightarrow b[\text{cat}=B], a \rightarrow c[\text{cat}=C], a \rightarrow d[], b < c\}$$

Ici, *d* ne porte pas d'information sur sa catégorie. Maintenant, les deux modèles à trois nœuds de D5 sont représentés par des arbres identiques. Il n'est donc plus possible d'identifier une classe d'équivalence des modèles par l'arbre la représentant.

Le but d'une description est de générer un ensemble d'arbres satisfaisant la description. Doubler un arbre identique dans l'ensemble d'arbre généré est inutile. Nous considérons donc deux modèles provenant de différentes classes d'équivalence comme équivalents si leurs classes sont représentées par le même arbre. Nous élargissons donc la définition du modèle référent aux descriptions augmentées d'un trait *cat*.



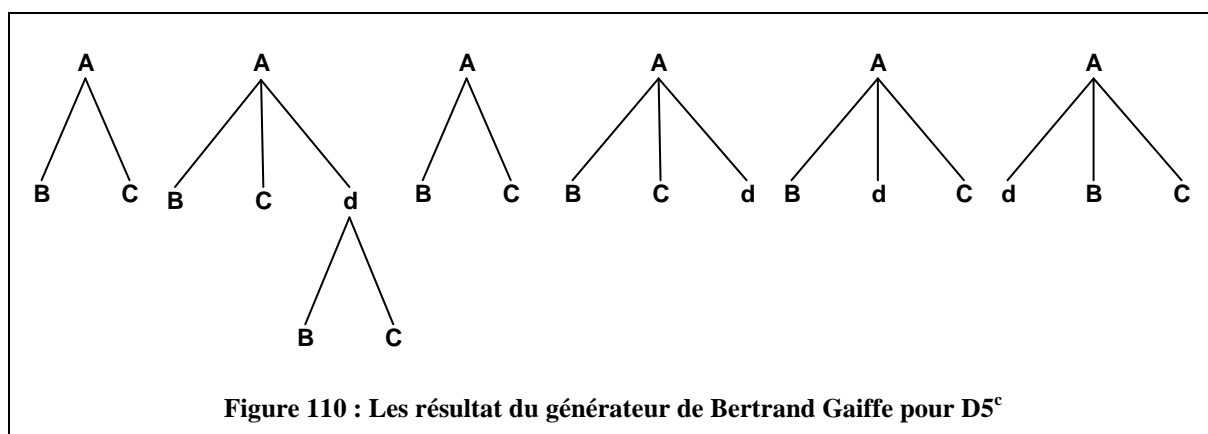
Pour  $D5^c$  on n'a donc que quatre modèles référents dont un minimal. De même, pour la description  $D3^c$  on ne note qu'une seule fois les deux premiers arbres identiques de la Figure 107 ; nous obtenons ainsi quatre modèles référents de  $D3^c$ .

La définition du modèle minimal peut rester la même pour les descriptions augmentées du trait *cat* : les modèles d'une description avec un nombre minimal de nœuds.

### 3.2.3 L'implémentation du trait *cat*

Le générateur de Bertrand Gaiffe et le générateur de M.-H. Candito prennent en compte le trait *cat* comme nom du nœud interprété par la constante qui le porte, sans pour autant changer la définition du modèle référent, ce qui donne des résultats souvent obscurs : pour la description  $D3^c$ , la version originale du générateur de Bertrand Gaiffe donne comme résultat les quatre arbres de la Figure 107, les deux premiers étant identiques aussi dans leurs traits. Il y a donc premièrement le redoublement causé par le fait que la classe d'équivalence où *b* s'est unifié avec *a* et la classe où *b* s'est unifié avec *c* soient représentées par le même arbre. Deuxièmement, il manque, parmi les modèles, le dernier arbre de la Figure 107 excluant la possibilité que la dominance entre *a* et *b* ne soit pas une parenté directe.

Les résultats pour  $D5^c$  sont encore plus énigmatiques. Voici les 6 modèles proposés <sup>156</sup> :



<sup>156</sup> Si on ne déclare pas la constante *d*, la contrainte  $a \rightarrow d$  est tout simplement ignorée, et on obtient qu'un seul arbre (le premier de la Figure 110).

On voit deux fois le modèle à trois nœuds, identique aussi dans leurs structures de traits (le redoublement qui apparaît pour les mêmes raisons : le nœud *B* du premier modèle à trois nœuds et le nœud *C* du 2<sup>e</sup> modèle à trois nœuds proviennent probablement de l'unification avec le constantes *d*). La raison de l'existence du 2<sup>e</sup> arbre de la Figure 110 parmi les résultats est encore plus mystérieux : ce modèle n'est pas un modèle référent.

On voit clairement apparaître la nécessité d'introduire la nouvelle notion de référent minimal que nous proposons chaque fois qu'on a affaire à une fonction de assignation. Ajouter cette fonction sans pour autant revoir les notions liées, comme c'est le cas dans les implémentations des Candito 1999 et de Gaiffe *et alii* 2002, conduise à des implémentations peu utilisables si on s'éloigne de l'objectif concret pour lequel elles ont été conçues.

### 3.3 Le modèle minimal dans la métagrammaire

Pour la génération des schèmes d'une TAG, Candito 1999 a jugé utile, exclusivement l'idée de modèle référent/référent minimal. Elle explique ainsi, que « pour représenter les phrases *Paul donne un livre à Marie / Paul donne à Marie un livre*, on utilise la même classe croisée, qui place l'objet indirect à droite du verbe. C'est la compilation en TAG qui génère les schèmes correspondants (comme ils portent des fonctions différentes, ils ne peuvent pas être unifiés) » (Candito 1999:76). Dans sa grammaire n'apparaissent aucun cas, où une constante n'est unifiée avec une autre que dans l'étape de la réalisation de la fonction grammaticale (dimension 3). On trouve même un mécanisme spécifique, documenté dans son guide d'utilisation du compilateur de métagrammaire, mais jamais justifié : une constante restée sans fonction après le croisement avec la dimension 2 est par défaut considérée comme nœud de substitution, ce qui rend impossible l'unification, même forcée, de cette constante avec son parent dans l'étape de la réalisation (dimension 3). Le fait même qu'elle considère les modèles référents au lieu des modèles minimaux est à l'origine de ce problème : Si des nœuds peuvent rester « traîner », bien qu'ils puissent aussi être unifiés avec d'autres nœuds, la question qui se pose est quel statut obtiennent ces nœuds dans le schème final. Elle a donc décidé de les rendre par défaut, des nœuds de substitution.

Un autre problème lié aux modèles référents concerne la question d'ordre entre deux constantes : Il est souvent problématique dans la formalisation syntaxique, de définir l'ordre fixe entre deux éléments *au cas où ils apparaissent tous les deux*, et la métagrammaire n'en fait pas exception. Prenons comme exemple le cas des verbes à trois arguments en anglais où l'objet direct précède l'objet indirect avec la proposition *to*, discuté par Candito 1999:79.

(221) a. Mary gives a book to Peter.

(222) b. \* Mary gives to Peter a book.

Quand la classe des verbes bitransitifs exige la réalisation d'un objet direct et d'un objet indirect, ces deux classes sont appelées indépendamment et leurs contraintes sur les nœuds sont ajoutées. Bien sûr, la classe de réalisation de l'objet indirect *I* ne possède pas d'information sur la classe de l'objet direct *D*, et l'inverse n'est pas valable non plus. Si la classe *I* voulait porter l'information sur la place de l'objet direct, elle doit mentionner le nœud de l'objet direct, par exemple en postulant  $d < i$ . Même si la constante *d* peut s'unifier avec une autre constante, on trouvera parmi nos modèles référents un schème où la constante *d* a formé un nœud tout seul. Cela arrivera aussi quand la classe *I* est utilisée pour un verbe avec objet indirect et sans objet direct comme par exemple pour le verbe *dance* dans

(223) Mary dances to the music.

On aurait donc des schèmes finaux avec des nœuds vides pour un tel verbe, ce qui n'est pas souhaitable.

Concrètement, considérons la classe



$V = \{p \rightarrow s[\text{cat}=\text{P}] \rightarrow s[\text{cat}=\text{N} ; \text{type}=\text{subst}], p \rightarrow v[\text{cat}=\text{V} ; \text{type}=\text{ancre}], s < v\}$ ,  
 qui a comme seul modèle référent (et seul modèle minimal l'arbre  $M$  de la Figure 111.  
 Supposons qu'il appartient à la classe des verbes bitransitifs, ayant donc encore deux réalisations  
 des fonctions grammaticales à faire : une première réalisation de l'objet direct supposée fournie  
 entre autre par la classe

$$D = \{p \rightarrow d[\text{cat}=\text{N} ; \text{type}=\text{subst}], v < d\},$$

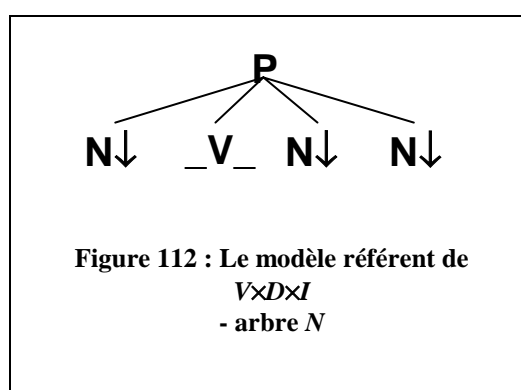
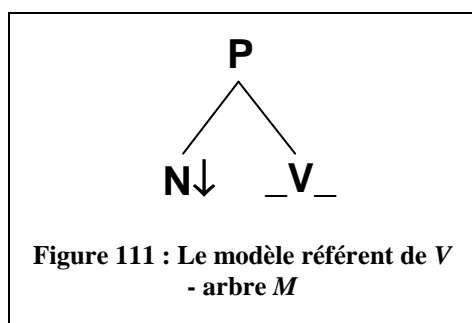
et une deuxième réalisation de l'objet indirect fournie par la classe

$$I = \{p \rightarrow i[\text{cat}=\text{N} ; \text{type}=\text{subst}], v < i, d < i\}.$$

Croisons maintenant les classes  $V$ ,  $D$  et  $I$ . On obtient donc

$$V \times D \times I = \{p \rightarrow s[\text{cat}=\text{N} ; \text{type}=\text{subst}], p \rightarrow v[\text{cat}=\text{N} ; \text{type}=\text{ancre}], p \rightarrow d[\text{cat}=\text{N} ; \text{type}=\text{subst}], p \rightarrow i[\text{cat}=\text{N} ; \text{type}=\text{subst}], s < v, v < d, d < i, v < i\}.$$

Le seul modèle référent (et seul modèle minimal) est l'arbre  $N$  de la Figure 112, ce qui est le schème souhaité, car on a l'ordre *sujet < verbe < objet direct < objet indirect* et on n'a pas obtenu de nœuds vides.<sup>157</sup>



La classe  $I$  porte l'information sur la position de l'objet indirect, qui reste valable indépendamment de la présence de l'objet direct. Pour l'ergonomie de la représentation, il est donc souhaitable de réutiliser la classe dans un autre contexte. Si, par exemple, on veut réutiliser la classe  $I$  pour générer le schème du verbe *to dance to something* de la phrase (223), il faudra assurer que l'information  $d < i$  (c'est-à-dire le fait que l'objet direct précède l'objet indirect) ne fasse pas apparaître un nœud  $d$  pour l'objet direct. On veut pouvoir dire que l'ordre entre l'objet direct et l'objet indirect s'applique quand les deux sont présents. Ici, les résultats sont différents suivant la méthodes du calcul des modèles référents utilisée : la classe

$$V \times I = \{p[\text{cat}=\text{P}] \rightarrow s[\text{cat}=\text{N} ; \text{type}=\text{subst}], p \rightarrow v[\text{cat}=\text{N} ; \text{type}=\text{ancre}], p \rightarrow i[\text{cat}=\text{N} ; \text{type}=\text{subst}], s < v, d < i, v < i\}$$

possède six modèles référents, présentés dans la Figure 113.<sup>158</sup> Ce premier arbre est le seul arbre minimal de la description  $V \times I$ .

Quels sont les autres choix pour représenter élégamment le phénomène d'ordre entre deux éléments potentiellement présents, pourtant simple linguistiquement ?

La procédure inverse, c'est-à-dire mettre l'information sur l'objet direct dans la classe  $I$ , aurait le même effet pour les verbes qui ne prennent qu'un objet indirect.

<sup>157</sup> Les résultats sont identiques pour le générateur de Bertrand Gaiffe.

<sup>158</sup> La version originale du compilateur de Bertrand Gaiffe trouve deux fois le premier arbre (provenant des unifications différentes), sortant donc sept schèmes.

On pourrait mettre l'information sur la précédence linéaire entre objet direct et indirect sur la classe finale du verbe, i.e. la classe NOVN1toN2 indiquant la sous-catégorisation initiale. Cette classe introduit la famille des verbes ayant un sujet, un objet direct et un objet indirect, et on peut donc être sûr de l'apparition d'un objet direct et d'un objet indirect. En contrepartie, on perdait ainsi la généralité de cette classe, car elle devrait aussi être utilisée pour des cas de non réalisation d'un argument (construction de contrôle, passifs, etc.) et de cliticisation des arguments. Dans ce dernier cas, l'ordre des arguments change. En fait, seule la classe de

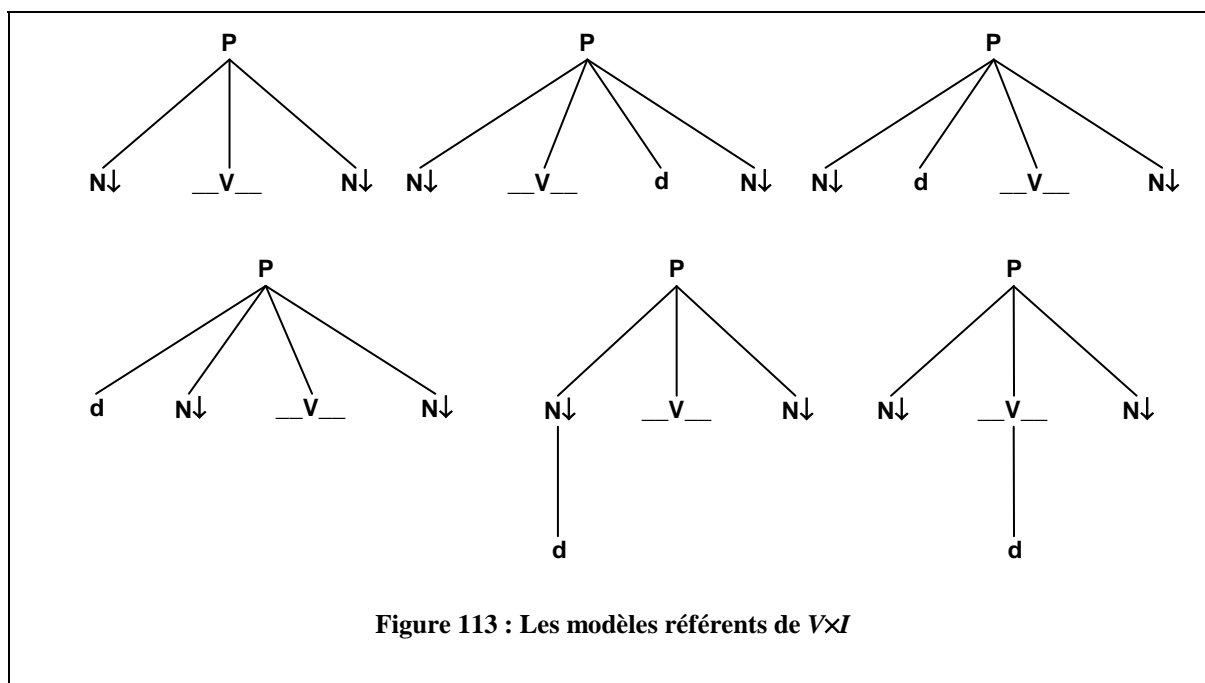


Figure 113 : Les modèles référents de  $V \times I$

réalisation d'un argument sait comment se placer par rapport à d'autres éléments, mais elle ne sait pas si ces éléments sont présents. On a donc besoin de mentionner d'autres éléments de manière optionnelle, la mention seule ne doit pas déclencher la création d'un nœud.

L'introduction des modèles minimaux est un premier pas dans cette direction. Dans l'exemple de la classe  $I$  ci-dessus, le nœud correspondant à la constante  $d$  n'apparaîtra pas en tant que nœud à part dans un modèle minimal tant qu'il y a un modèle dans lequel il a été « avalé » par unification avec une autre constante. Dans le modèle minimal, une constante s'unifie à une voisine si présente, tant que la description de la classe croisée finale ne fournit pas de traits interdisant cette unification. Dans notre exemple, la constante  $d$  peut « disparaître » dans la constante  $s$  et dans la constante  $v$ <sup>159</sup>. Sans cette possibilité d'unification, un nœud correspondant à  $d$  apparaîtrait. Il ne s'agit donc pas vraiment d'une description avec nœuds optionnels, car il faut se soucier du sort de la constante qui pourrait éventuellement disparaître. Par exemple, si on avait mis l'information sur l'ordre linéaire entre objet direct et indirect sur la classe de l'objet direct, comme suit :

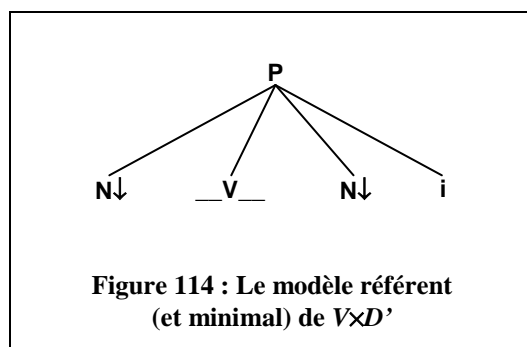
$$D' = \{p \rightarrow d[\text{cat}=\text{N} ; \text{type}=\text{subst}], v < d, \underline{d} < i\}$$

Le nœud pour l'objet indirect apparaîtra aussi dans le modèle minimal de la classe  $V \times D'$ , car à sa droite, la constante  $i$  ne trouve pas d'autres nœuds avec lesquels elle aurait pu s'unifier.<sup>160</sup>

<sup>159</sup> Ces deux possibilités expliquent le redoublement de l'arbre qui résulte du compilateur de Bertrand Gaiffe.

<sup>160</sup> Ce qui est d'ailleurs une autre bonne raison pour l'existence de la parenthèse droite, qui peut être vide, car les nœuds dans le Mittelfeld dispose de ce fait toujours d'un nœud à leur gauche et d'un autre à leur droite. On peut donc librement statuer des relations d'ordre entre les éléments sans se soucier de « la place de refuge » en cas de non réalisation des nœuds ainsi mentionnés.

Bien que la bonne utilisation des descriptions d'arbres reste difficile, l'accès aux modèles minimaux permet une plus grande expressivité et une réutilisation plus facile des classes de la métagrammaire.



### 3.4 Les traits polaires

Candito 1999:68 donne des critères de croisement qui sont complètement et directement liés aux trois dimensions qu'elle propose pour la description des structures verbales : ainsi, une classe croisée doit hériter à la fois d'une classe terminale de la dimension 1, de la dimension 2 et de la dimension 3 par le biais de la fonction de sou-catégorisation finale établie dans la deuxième dimension. Par ailleurs, indépendamment de ces croisements obligatoires, une classe peut provoquer, ou au contraire interdire, un croisement avec une autre classe. Or d'un point de vue formel mais aussi linguistique, la quête de généralisations portant sur cet ensemble d'obligations et d'interdictions de croisement s'impose. Une alternative est proposée par Gaiffe *et alii* 2002. Ces derniers proposent d'exprimer les incompatibilités qui existent entre deux classes (et donc qui interdisent le croisement de ces deux classes) par l'unification simple d'un jeu de traits attaché à chaque classe : autrement dit, si l'unification de la structure échoue, alors les classes sont considérées comme incompatible. Un tel système de traits acquiert également une importance linguistique si on joint ce jeu de traits aux schèmes résultants de la description d'arbre de la classe croisée. Les traits proposés par Gaiffe *et alii* 2002 s'apparentent aux « hypertags » proposés par Kinyon 2000 et s'avèrent être d'une grande utilité dans le processus de la mise en correspondance d'une forme d'un mot avec les arbres de sa famille qui lui correspondent<sup>161</sup>. En fait, la nécessité de disposer de cette information sur les schèmes générés était à l'origine de la réimplémentation du générateur de Candito car ce dernier ne gardait pas l'information des classes croisées sur les schèmes qui en résultent.

Gaiffe *et alii* 2002 proposent d'utiliser les traits polaires pour exprimer les obligations de croisement entre deux classes. Ce choix permet de mettre ensemble les obligations d'unification qui apparaissent dans l'approche de Candito et qui découlent de la séparation des classes en dimensions, avec les « contraintes de cooccurrence positives », c'est-à-dire des obligations de croisement de deux classes qui sont exprimées explicitement sur les classes. Suivant la définition de Gaiffe *et alii* 2002, un trait polaire est un trait qui peut avoir une valeur positive, correspondant à un besoin de ce trait, ou une valeur négative, correspondant à une fourniture de ce trait. Le compilateur croise deux classes quand l'une « fournit » un trait positif à l'arbre qui a « besoin », c'est-à-dire qui contient une classe qui a une valeur de trait négative. Les classes finales qui sont retenues sont les classes « neutres », c'est-à-dire les résultats de croisement qui arrivent à zéro, sans besoins ni fournitures.

Illustrons avec un exemple. Au lieu d'exprimer à l'instar de Candito 1999 qu'une classe de la dimension 1 doit se croiser avec une classe de la dimension 2, la classe exprime directement ce

<sup>161</sup> Par exemple, un infinitif doit choisir un certain sous-ensemble de la famille qui est différente du sous-ensemble de la forme finie, et on a besoin de ce jeu de traits pour établir les compatibilités entre forme et arbre.

dont elle a « besoin » à l'aide d'un trait particulier que les classes correspondantes de la dimension 2 remplissent. L'approche des traits polaires représente donc un élargissement propre du formalisme car il est possible de traduire directement la méta-grammaire de Candito dans une méta-grammaire à traits polaires. Il est en effet possible dans un premier temps d'ajouter à chaque classe de la dimension 1 un trait position du nom de « dim2 », ce qui indiquerait que chaque classe doit trouver une autre classe qui peut neutraliser ce trait. Ensuite on ajoute cette capacité de fournir le trait « dim2 » à chaque classe de la dimension 2, et on traduit l'obligation de croisement entre les classes de la dimension 1 et de la dimension 2 dans un système à traits polaires. De la même manière, chaque fonction de sous-catégorisation finale a besoin d'une réalisation spécifique à la fonction qui peut être remplie par les classes correspondantes de la dimension 3.

A l'aide des traits qui sont unifiés de la façon habituelle et des traits polaires qui doivent se neutraliser, nous arrivons à exprimer toutes les contraintes de cooccurrence imaginables tout en disposant d'un système ayant une généralité formelle.

Cependant, malgré sa simplicité apparente, le système des traits polaires révèle des difficultés importantes dans le calcul de la combinaison de deux jeux de traits polaires. En effet, même si on accepte qu'une classe ne puisse exprimer qu'un simple besoin ou une simple fourniture d'un trait donné, il n'est pas clair comment on peut effectuer la combinaison de deux jeux de traits qui ont tous les deux besoin d'un même trait ou qui fournissent tous les deux le même trait. Pourtant, nous sommes confrontés à de tels cas même si on suppose que l'algorithme ne croise que deux classes dont un trait polaire remplit le besoin de l'autre, pour la simple raison que le croisement peut avoir lieu à cause des valeurs complémentaires du trait  $A$  qui implique la combinaison de deux valeurs identiques d'un autres trait  $B$ . Les classes  $C_1$  et  $C_2$  illustrent un tel cas :

$C_3$
$B : -$

Plusieurs analyses sont possibles pour des cas pareils. D'abord, on peut complètement interdire les croisements de ce genre. Néanmoins, ce choix crée des problèmes au niveau de l'algorithme car il peut y avoir une troisième classe  $C_3$  qui neutralise le trait qui pose problème :

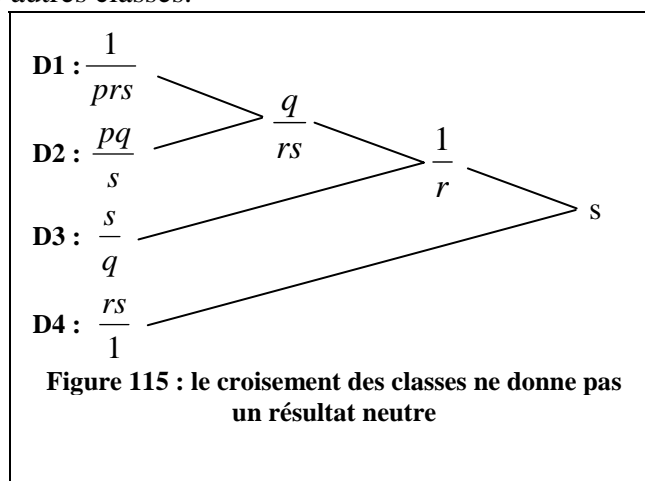
$C_1$	$C_2$
$A : +$	$A : -$
$B : +$	$B : +$

On peut donc croiser  $C_1$  avec  $C_3$  ( $C_1 \times C_3$ ) et  $C_2$  avec  $C_3$  ( $C_2 \times C_3$ ), ce qui neutraliserait le trait problématique  $B$ , puis on continuerait par croiser ces deux ensembles déjà croisés ( $(C_1 \times C_3) \times (C_2 \times C_3)$ ). La possibilité d'arriver ou non à une classe croisée neutre dépendra donc de l'ordre dans lequel on aura essayé les différentes combinaisons de croisement de classes.

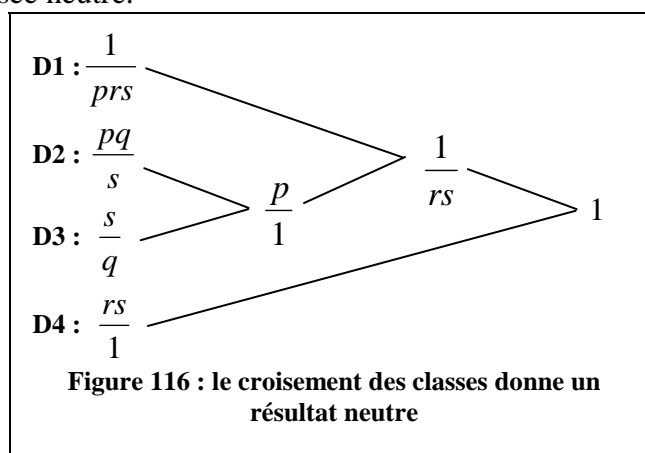
Une autre analyse possible, parallèle à la première, est de proposer un algorithme qui dépend de l'ordre de l'introduction des différents traits polaires. Cependant, cette analyse est à l'origine d'un bogue important dans la version originale du compilateur de Gaiffe *et alii* 2002. En effet, les auteurs partent du principe qu'un trait polaire ne peut avoir que trois valeurs : un trait polaire peut être positif, négatif ou neutre. Par conséquent, la combinaison de deux traits qui ont la même valeur implique que la valeur persiste : la combinaison de deux traits positifs donnent comme résultat un trait positif, la même chose étant possible également pour la combinaison de deux traits négatifs. On peut donc dire qu'une classe, qui est le produit d'un croisement de deux classes qui ont besoin toutes les deux de quelque chose, a toujours doublement besoin de la même chose.

Pourtant, le fait qu'on mette de côté le caractère « double » du besoin et de la fourniture du nouveau croisement peut avoir des conséquences assez graves au niveau de l'algorithme.

Regardons d'abord l'illustration donnée dans la Figure 115<sup>162</sup>. Pour un algorithme qui ne considère pas des besoins et des fournitures « doubles », les quatre classes de la Figure 115 ne donnent pas de classe croisée neutre si on traite les traits polaires dans l'ordre (p, q, r). En effet, le premier croisement combine deux valeurs négatives du trait *s* en un seul trait avec une valeur négative simple, ce qui fait que nous perdons la capacité de neutraliser deux fois la valeur positive du trait *s* dans d'autres classes.

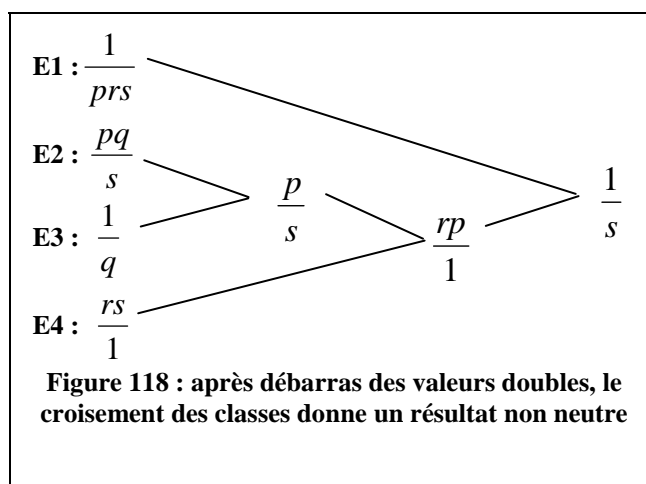
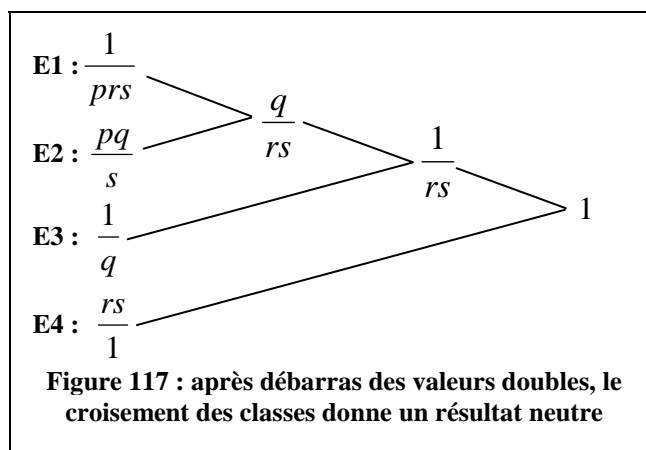


Si on procède maintenant dans un autre arbre que nous reproduisons dans la Figure 116, on peut aboutir à une classe croisée neutre.



En effet, il est parfois possible de trouver une solution, et il se peut même qu'on en trouve trop dans le sens que seul le fait de s'être débarrassé du double trait permet d'aboutir à une classe croisée neutre (Cf. exemples de la Figure 117 et Figure 118). Ainsi, il s'avère que l'algorithme qui ne considère pas des valeurs doublées de traits polaires est non déterministe, car il dépend de l'ordre qui fixe la neutralisation des traits polaires.

<sup>162</sup> Dans la figure, les besoins sont notés en haut et les fournitures en bas. Ainsi,  $\frac{s}{q}$  signifie que *s* a une valeur positive et *q* une valeur négative. Par ailleurs, le symbole 1 signifie respectivement l'absence d'une valeur positive ou négative.



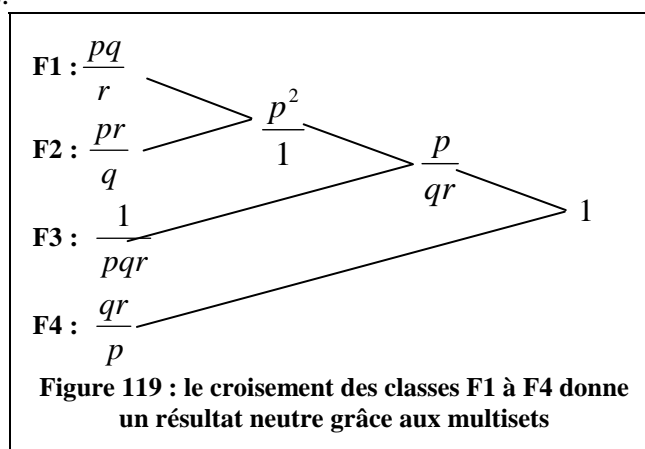
Face à cet algorithme non déterministe, nous avons été amené à proposer deux solutions et à changer l'algorithme. Dans un premier lieu, chaque fois que deux classes doivent être croisées avec un trait polaire ayant la même valeur, on a fait en sorte que ce croisement n'est pas effectué tout de suite mais à la fin, après le croisement des autres traits polaires. En d'autres termes, deux classes avec un trait polaire avec la même valeur sont « mises en attente » jusqu'à que les autres traits polaires aient été croisées. Afin d'éviter une boucle infinie, l'algorithme s'arrête si chaque trait polaire a été mis en attente sans qu'on ait pu neutraliser ce trait. Cette solution permet donc de rendre l'algorithme non déterministe sans pour autant avoir à compter les besoins et les fournitures. Ainsi, dans le cas de l'exemple de la Figure 115, dès que l'algorithme rencontre le problème qui apparaît pendant la neutralisation des traits des classes D1 et D3, il traite la neutralisation du trait  $q$ . La situation qui se produit est donc la même que dans le cas de l'exemple de la Figure 116, permettant ainsi l'aboutissement à une neutralisation des traits.

Par contre, le désavantage de cette « déterminisation » simple de l'algorithme est qu'il existe des cas d'interblocage (angl. 'deadlock') où cet ensemble de classes a le même nombre de traits polaires de valeur positive que de valeur négative, faisant en sorte que la somme est nulle. Et il se trouve que la l'algorithme ne peut pas trouver ces traits. Les classes F1 à F4 illustrent un tel cas :

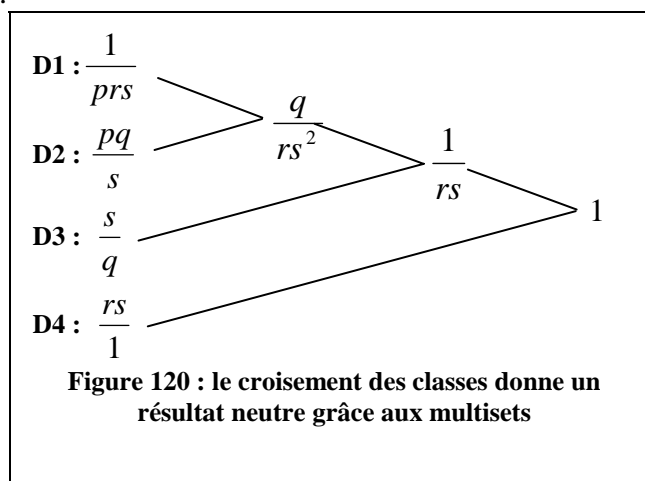
$$F1 : \frac{pq}{r} \quad F2 : \frac{pr}{q} \quad F3 : \frac{1}{rqp} \quad F4 : \frac{qr}{p}$$

Chaque classe pourrait neutraliser un trait polaire en se combinant avec n'importe quelle autre classe, mais chaque couple de classes a en commun un trait polaire de la même valeur, ce qui conduirait l'algorithme à repousser le croisement. L'algorithme constatera donc un interblocage, bien que des traits positifs et négatifs pourraient se neutraliser.

Il est possible de résoudre de tels cas problématique en implémentant le croisement avec des multiset, c'est-à-dire en comptant le nombre de besoins ainsi que le nombre de fournitures qui apparaissent pour un croisement. Ceci permet d'obtenir la neutralisation des traits polaires des classes F1 à F4 : on obtient le croisement suivant  $F1 \times F2 \times F3 \times F4$  en donnant comme résultat une classe neutre. Les modèles minimaux de cette classe seront calculés par la suite à partir de la description de ses arbres.

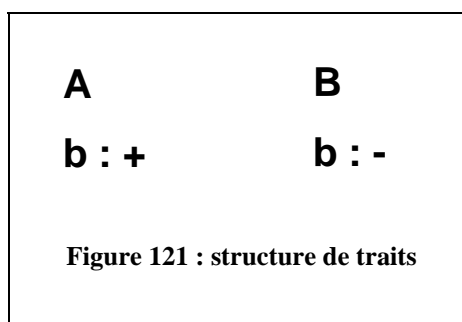
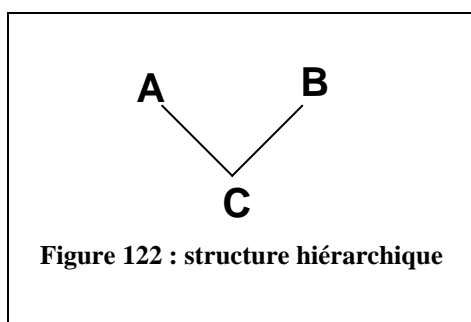


Cette solution permet donc d'aboutir à la bonne neutralisation des traits polaires des classes D1 à D4 (Cf. Figure 115), et ceci indépendamment de l'ordre dans lequel nous avons procédé à la combinaison des classes.



Bien que nous introduisions dans l'algorithme les multiset de besoins et de fournitures, nous gardons pour l'instant les besoins et les fournitures simples au niveau de la description. En fait, nous n'avons pas eu affaire à des cas où une classe doit exprimer deux fois le même besoin (ou qui parallèlement, une classe qui fournit deux fois le même trait). Cependant, nous n'excluons pas cette possibilité pour des descriptions qui sont adéquates à d'autres phénomènes linguistiques ou d'autres langues.

Notons encore que les traits polaires multiset représentent un outil très puissant qui rentre en compétition avec la structure hiérarchique. Il est toujours possible d'exprimer un croisement entre les classes A et B donnant ainsi la classe C à l'aide des traits polaires. Par exemple, la hiérarchie de la Figure 122 est équivalente à la structuration de l'information sur les traits de la Figure 121 : A a besoin de B et B le fournit, ce qui rend obligatoire le croisement des deux.



Ainsi, ce « double langage » nous fournit une plus grande expressivité qui permet de distinguer une structure analyse comme hiérarchique d'une structure analysée comme provenant d'un croisement de deux informations. La hiérarchie que nous obtenons est souvent plus propres et plus lisibles que si nous étions obligés d'exprimer tout dans la même hiérarchie.

## 4 DTAG

Nous arrivons maintenant à la partie propre à la construction d'une grammaire TAG de l'allemand. Nous commencerons par rappeler les objectifs de DTAG, puis nous présenterons les critères qui ont été retenus pour cette grammaire. Ceci nous amènera aux premiers résultats pour des phénomènes « problématiques » de l'allemand ? Nous proposerons quelques arbres élémentaires et nous discuterons des problèmes de verbe « second », des relatives et des compléments verbaux. La dernière sous-section sera consacrée à l'organisation de la DTAG dans une méta-grammaire.

### 4.1 Objectifs de DTAG

Dans le domaine de la génération des textes, nous avons affaire tout particulièrement à une tâche où on n'a pas besoin de tous les ordres des mots, car il existe des ordres des mots plutôt neutres qui sont acceptables dans tous les contextes. Ces ordres sont suffisants pour les tâches des systèmes actuels. TAG propose une interface entre les mots ordonnés et une structure de dérivation, qui, aussi faussée soit-elle sur le plan sémantique, est d'un grand intérêt pour les systèmes de génération. Pour la génération on pourrait même s'avancer à dire qu'avoir un choix limité d'ordres de mots facilite la tâche du générateur, car un module qui permet de générer tous les ordres possibles dans des contextes communicatifs très spéciaux, doit savoir choisir entre eux, ce qui représente une mission qui outrepassé les capacités des systèmes de génération en fonction aujourd'hui.

Le hic de l'approche TAG est que l'ordre neutre d'un contexte donné ne fait pas forcément parti des ordres qui sont imposés par le formalisme. Dans le casz de l'allemand, l'ordre neutre est défini par sa capacité de pouvoir exprimer la plupart des structures communicatives possibles (Lernerz 1977, Choi 1999)<sup>163</sup>, et c'est ordre est souvent présenté comme l'ordre des mots par défaut dans les systèmes qui ne prennent pas en compte la structure communicative. Cette structure coïncide heureusement souvent avec l'un des ordres que TAG peut générer (pour un contre-exemple, je renvoie au cas de pronominalisation de la phrase (209).b).

<sup>163</sup> Comme le montre Müller 1999, cet ordre standard n'est pas indépendant du lexème : par exemple, la plupart des verbes préfèrent le datif devant l'accusatif, mais quelques uns préfèrent l'inverse. Voir la discussion de ce phénomène en section I.6.



Le but de DTAG est donc de générer une grammaire dans le formalisme LTAG qui soit aussi bonne linguistiquement que possible, mais surtout, sous un point de vue pratique, aussi utile que possible dans la tâche de génération de texte.

La construction de la DTAG se base sur des exemples de textes à générer de nature plutôt simples. Ces exemples ont été rencontrés dans deux projets de génération de l'allemand auxquels j'ai participé, les projets MultiMétéo<sup>164</sup> et SmartKom<sup>165</sup>.

Avec la DTAG, nous cherchons donc à trouver un compromis entre les différents besoins pratiques et restrictions imposé par le formalisme qui ne sont pas tous à la fois satisfaisables : en effet, des configurations d'ordre des mots qui apparaissent trop souvent pour être laissées de côté doivent être traitées, même si les critères linguistiques ainsi que l'interprétabilité sémantique de l'arbre de dérivation en souffrent.

Les arbres des relatives extraposées représentent un cas de violation inévitable des critères linguistiques. Par contre, nous ne traitons pas les cas de brouillage où les éléments sous-catégorisés de deux verbes échangent leur place. La phrase (209).b ne sera donc pas générée. De plus, nous avons fait en sorte que la « topicalisation » dans le Vorfeld (voir chapitre I et II) reste possible pour tous les arguments et modificateurs verbaux, aussi profondément enchâssés que soient ces verbes. Dans cette grammaire, le placement des verbes non-finis n'est permis que dans la parenthèse droite et dans le Nachfeld (l'extraposition n'est possible que pour les infinitifs en *zu*). Pour les phrases à générer, nous considérons pouvoir éviter le « VP-fronting » dans le Vorfeld ainsi que l'intraposition.

Nous sommes conscients que les choix présentés ici sont des décisions complètement pragmatiques. Ils ne prétendent pas d'une scientificité autre que de montrer leur efficacité dans des tâches concrètes de génération<sup>166</sup>.

## 4.2 Les critères retenus pour DTAG

La difficulté majeure lors de la tâche de création de la grammaire DTAG découle principalement du manque de restrictions qui résulte des échecs des descriptions parfaites ; on a la *Qual der Wahl*, (litt. la peine du choix, 'l'embarras du choix') : on sait d'ores et déjà qu'on ne peut pas respecter tous les critères qu'on a stipulés sur l'utilisation du formalisme en linguistique. Mais en fonction des choix qu'on fait sur les arbres élémentaires, l'un sera vérifié tandis que l'autre non. Lequel préférer ? Entre deux structures de constituants résultantes (deux structures dérivées), comment choisir et sur quelles bases ?

Puisque le but est d'obtenir une grammaire parallèle à celle de l'anglais et du français, utilisable avec les mêmes outils, il est nécessaire de connaître les principes sous-jacents de ces grammaires. A la suite d'analyses de ces grammaires existantes et de discussions avec ses développeurs, j'ai retenu surtout un critère de base, le critère de prédominance de la structure de dérivation :

---

<sup>164</sup> Il s'agit d'un projet européen de génération multilingue de bulletins météorologiques, réalisé par la société Lexiquest, Coch 1998.

<sup>165</sup> SmartKom est un projet développé au centre allemand de recherche en intelligence artificielle DFKI de génération interactive d'indication du programme culturel d'une ville, <http://smartkom.dfki.de>.

<sup>166</sup> Bien entendu, il est partiellement clair pourquoi les ordres des mots qu'on obtient ainsi constituent l'essentiel des phrases apparaissant dans les besoins de la génération de texte : les autres ordres impliquent habituellement une structure communicative beaucoup plus spécifique et compliquée. Ce sont donc des constructions à éviter dans des tâches de génération aussi simples que les bulletins météorologiques (MultiMétéo) et l'indication du programme culturel d'une ville (SmartKom). De la même manière, il est prévisible qu'une analyse de texte avec DTAG sera d'autant plus prometteuse que le texte est simple, informatif, non littéraire, non parlé, car ces textes auront la même simplicité de la structure communicative.

*Critère de prédominance de la structure de dérivation* : Si le formalisme LTAG impose le choix entre une structure dérivée préférable et une structure de dérivation préférable, cette dernière est considérée comme la plus importante.

Autrement dit, la structure dérivée est d'un moindre intérêt que la structure de dérivation, parce que, comme on dispose pour elle d'un catalogue considérable, les différentes structures résultantes trouvent pratiquement toujours une justification linguistique : elle peut être de nature sémantique, elle peut indiquer la portée des éléments, elle peut être de nature syntaxique et aussi prosodique. Tout est permis.

En conséquence, dans les choix apparaissant pendant le développement, le but principal est de lier la chaîne des mots à une structure de dérivation qui est la plus sémantique possible.

Le 2<sup>e</sup> critère que j'ai pris en considération est le critère de modularité dans une métagrammaire, que j'utilise pour la génération des schèmes. Les *schèmes*, sont des arbres élémentaires qui ne sont pas encore lexicalisés et qui portent une place (*l'ancre*) où la lexicalisation a lieu dans un deuxième temps. Pour une explication plus précise, je renvoie à la section 4.

*Critère de la représentation modulaire* : parmi deux grammaires qui ont la même capacité générative dérivationnelle, celle qui permet plus facilement une représentation modulaire dans une métagrammaire est préférable.

Ce critère n'intervient que quand la meilleure interprétabilité sémantique a été obtenue. Ce critère pourrait donner parfois comme résultat des structures élémentaires plus compliquées que nécessaires si on ne considérait que l'utilisation de ces structures pendant la génération (ou l'analyse), car des parties de l'arbre provenant de différentes classes de la métagrammaire auront éventuellement des nœuds à part. L'impact de ce critère deviendra clair par la suite.

Enfin le troisième critère que j'ai pris en compte porte sur le nombre des *schèmes* différents nécessaires dans une grammaire :

*Critère de la minimalité de l'ensemble des schèmes* : Pour la même capacité générative dérivationnelle donnée, une grammaire avec moins de schèmes différents est préférable.

Un ordre d'importance est établi entre les critères, ce qui nous guide dans les choix apparaissant pendant le développement d'une grammaire TAG :

1. L'ordre des mots induit par les phénomènes syntaxiques considérés comme importants (voir ci-dessus), doit faire partie de la grammaire à tout prix, même si les autres critères en souffrent.
2. On essaie de maximiser la similarité entre la structure de dérivation et la structure sémantique<sup>167</sup>.
3. On essaie de choisir des arbres qui permettent facilement un partage d'information dans les différentes classes d'une métagrammaire
4. On essaie de faire des choix qui minimalisent le nombre d'arbres nécessaires.
5. S'il reste des choix à faire, on applique un dernier critère : on essaie d'obtenir une structure dérivée qui ressemble à une structure de constituants X-bar.

L'idéal serait, bien entendu, que tous les critères soient remplis, mais en réalité, on s'aperçoit que les deux premiers critères imposent déjà la quasi-totalité des décisions sur la forme des arbres et laissent peu de poids aux autres critères.

Après avoir exposé la raison d'être de la DTAG et les critères retenus pour sa génération, nous allons maintenant présenter les arbres concrets et discuter les choix sous-jacents.

---

<sup>167</sup> La forme de la structure sémantique n'est pas incontestée, et même pour une structure sémantique donnée, il se peut qu'on soit obligé de faire un choix entre deux liens prédicat-argument qui ne sont pas réalisables tous les deux. Nous rencontrons un tel cas avec les verbes à contrôle, voir section 2.2.1.

### 4.3 Des solutions compromettantes : les premiers arbres

Après avoir esquissé la raison d'être de la DTAG et les critères retenus, nous présentons dans cette section les arbres verbaux de base de la grammaire DTAG. Nous montrerons pourquoi il est nécessaire (mais aussi problématique) de faire des compromis entre les différents critères proposés.

Le choix des phénomènes grammaticaux qui sont encodés dans ces arbres élémentaires, ont été choisis en fonction des besoins pour le projet de génération SmartKom (Cf. la note de bas de page 165). L'élaboration des concepts de base est le résultat des discussions avec Tilman Becker du DFKI, Sarrebruck, qui est le responsable de ce projet.

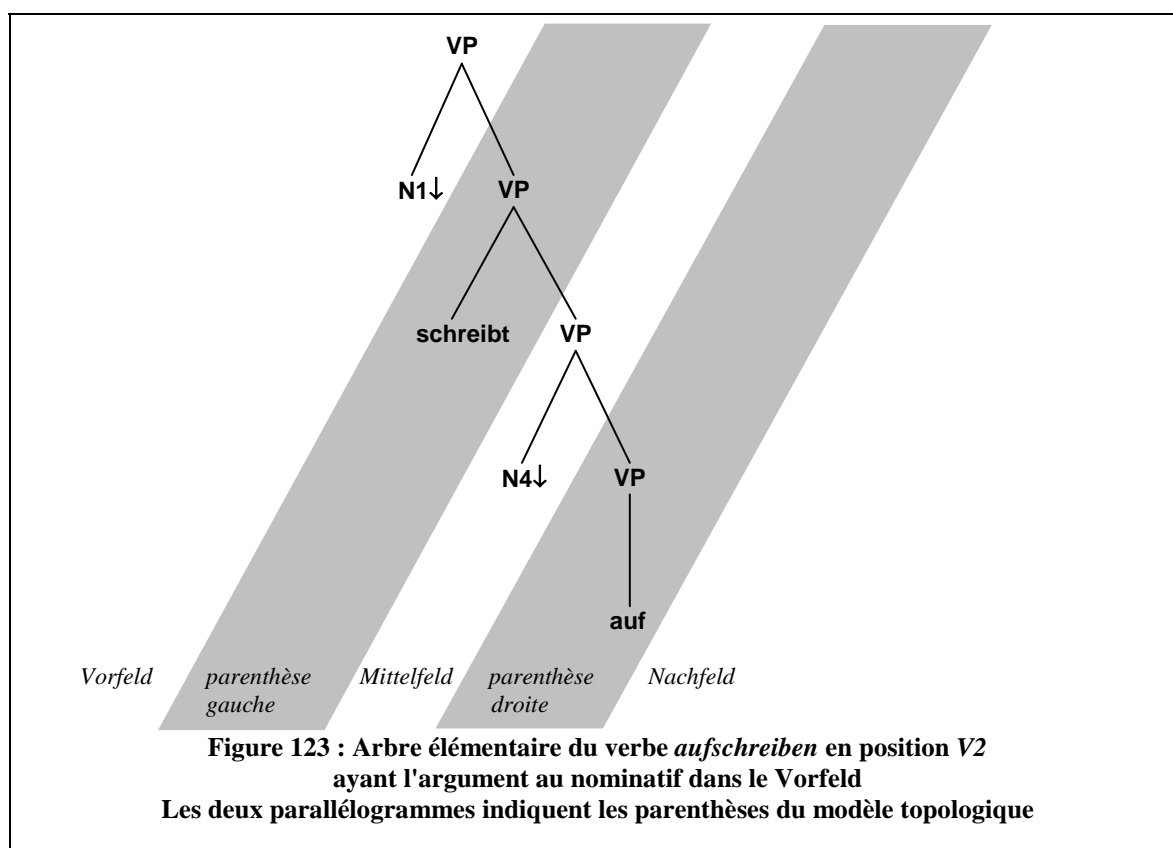
Dans les sous-sections suivantes nous allons présenter la forme de base qu'ont les arbres élémentaires proposés, discuter la réalisation du verbe second (dit V2) en TAG, proposer des ajustements pour résoudre les problèmes de la proposition relative (la problématique posée par les relatives a été présentée dans la section 2.4.4) et enfin illustrer avec des exemples des arguments verbaux. Notons qu'une classification complète des arbres verbaux de l'allemand ne peut être discutée qu'après l'introduction à la métagrammaire et à ses problèmes, que nous présentons dans la section suivante.

#### 4.3.1 Arbres élémentaires

La décision essentielle qui guide la forme de tous les arbres élémentaires verbaux, est la réalisation d'une structure de bifurcation droite (angl. '*right branching*') : il s'agit d'une structure où les nœuds sont au maximum binaires (ils ont deux enfants ou moins), et où le chemin de la racine au nœud de la parenthèse droite est toujours le chemin vers la droite. Le chemin entre la racine et la parenthèse droite est appelé la *colonne* de l'arbre<sup>168</sup>. Pour la plupart des cas, le nœud (ou l'un des nœuds) de la parenthèse droite représente l'ancre, sauf dans le cas où le verbe est en position V2. Dans ce dernier cas, le verbe pend d'un nœud de la colonne spécifiée pour gouverner le V2. L'image suivante permettra de clarifier ce propos. Prenons donc comme exemple le verbe *aufschreiben* 'noter'. il s'agit d'un verbe transitif avec particule, qui est composé de la racine verbale *schreiben* 'écrire' et de la particule *auf* 'sur' (d'origine prépositionnelle). Comme il a déjà été discuté dans la section II.3.2, dans le cas du V2, la racine du verbe sans particule est placée dans la parenthèse gauche, tandis que la particule prend sa place dans la parenthèse droite.

---

<sup>168</sup> Tous les nœuds pieds verbaux (sauf pour les relatives) apparaissent dans la parenthèse droite (Cf. section 4.3 suivante). Cette définition est donc une extension de la définition donnée dans la section 1.3.1 ci-dessus, qui se limitait au chemin qui existe entre racine et nœud pied.



On retrouve dans l'arbre élémentaire la partition du modèle topologique qui se projette sur l'arbre (ce qui n'a rien d'étonnant puisque l'arbre encode l'ordre linéaire, voir section II.2). Cette interprétation est représentée dans la Figure 123 par les parallélogrammes grisés renvoyant aux parenthèses du modèle topologique.

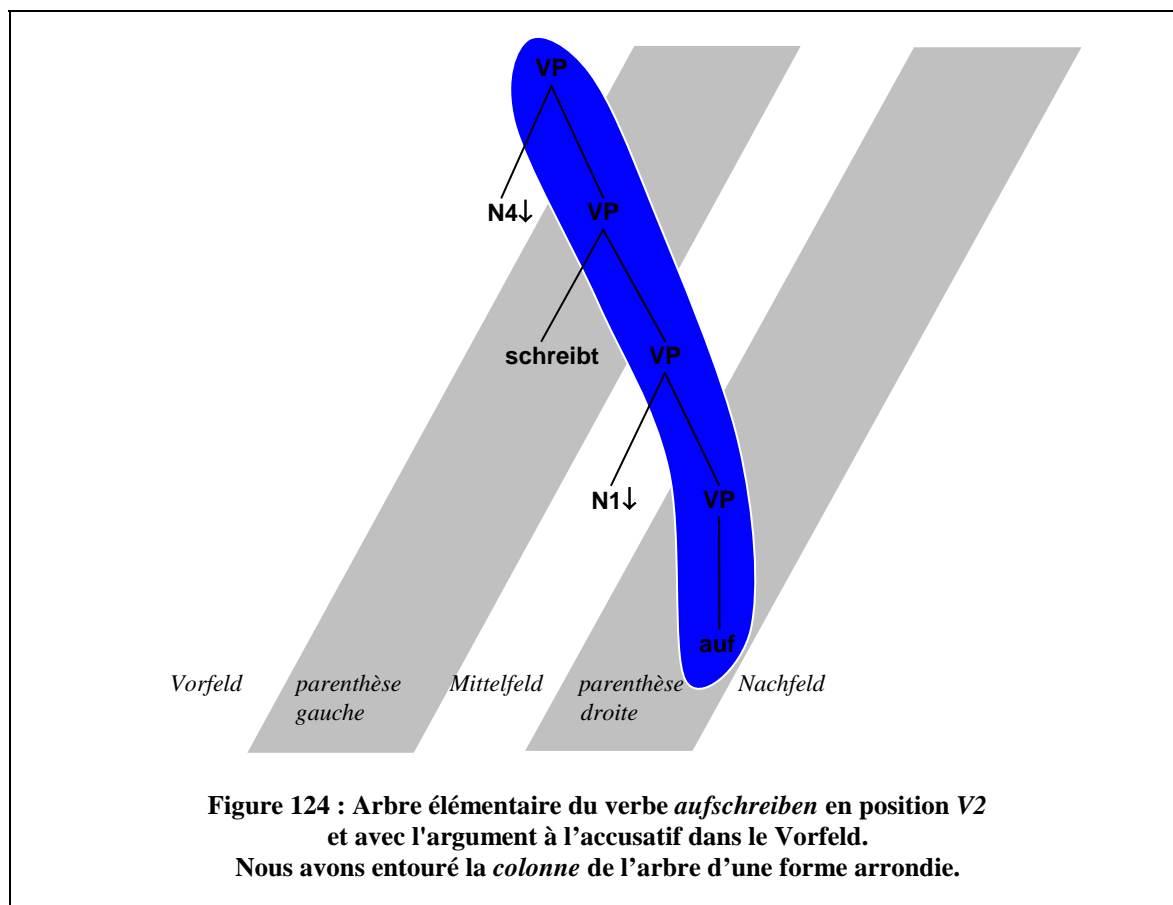
Les deux arguments nominaux du verbe apparaissent tous les deux comme nœud de substitution dans l'arbre élémentaire du verbe. Les cas grammaticaux des arguments nominaux sont indiqués par le numéro dans le nom du nœud de substitution. Suivant l'ordre habituel, nous notons

- 1 le nominatif
- 2 le génitif
- 3 le datif
- 4 l'accusatif<sup>169</sup>

La phrase (224) est un exemple qui peut être analysé à l'aide de l'arbre élémentaire dans la Figure 123 : le sujet (N1) se trouve dans le Vorfeld et l'objet direct (N4) se trouve dans le Mittelfeld.

<sup>169</sup> Il existe des verbes qui sous-catégorisent deux noms au même cas. En ce cas, un des noms a une fonction attributive ; il possède des propriétés syntaxiques différentes et son nœud de substitution est noté différemment.

- (224) Diese Frau schrieb ihre Reiseeindrücke auf.  
 cette femme écrivait ses voyage\_impression sur  
 'Cette femme a noté ses impressions de voyage'



Dans le cas des phrases où le Vorfeld est occupé par l'accusatif, nous avons besoin d'un autre schème où les nœuds de l'argument nominatif et de l'argument accusatif ont mutuellement échangé leur place, tel qu'il est présenté dans l'arbre de la Figure 124.

Cet arbre permet d'analyser la phrase (225).

- (225) Ihre Reiseeindrücke schrieb diese Frau auf.  
 ses voyage\_impression écrivait cette femme sur  
 'Cette femme a noté ses impressions de voyage'

Il est nécessaire maintenant d'expliquer les décisions qui ont été prises afin d'obtenir les arbres de cette forme. Notre but n'est pas du tout de reproduire le modèle topologique dans une structure syntagmatique classique, mais le résultat obtenu a été inévitable compte tenu des facteurs complètement intrinsèques au fonctionnement du formalisme TAG.

En effet, le choix d'une structure de bifurcation droite était inévitable, car, faute de faire mieux, cette dernière permet

- l'adjonction des modifieurs, qui peuvent intervenir à toutes les places dans le Mittelfeld, (Cf. section 4.3.2 ci-dessous).
- l'analyse des relatives extraposées, (déjà présenté dans la section 2.4.4), dont l'analyse TAG sera discutée ci-dessous.
- une représentation compacte et uniforme des arguments verbaux (Cf. section 4).

Par ailleurs, pour la même raison, le nœud VP unaire doit être dans la parenthèse droite, qui gouvernera tous les éléments apparaissant dans ce champ. Ce nœud s'avère aussi très utile pour la description de la grammaire dans la métagrammaire.

Enfin, donner le nom VP aux nœuds dans la colonne de l'arbre est complètement arbitraire, mais le fait que tous ces nœuds portent le même nom, facilite aussi bien le traitement de l'adjonction sur ces nœuds que la représentation dans la métagrammaire.

### 4.3.2 V2 en TAG

Bien que l'idée du verbe en deuxième position soit très simple à décrire, il s'agit d'une caractéristique de l'allemand qui est difficile à traduire dans le formalisme TAG. En effet, avec les outils formels proposés par TAG, l'expression du Vorfeld, et avec celle-ci l'expression de l'existence d'une place se trouvant devant le verbe fini principal, qui peut être occupée par un et un seul syntagme quelconque, pose problème. Aucune des opérations de combinaison d'arbres, ni la substitution ni l'adjonction, n'est adaptée à exprimer les restrictions de place qui peuvent agir sur un champ topologique. En effet, la substitution est introduite pour les arguments d'un prédicat et l'adjonction est prévue pour des processus itératifs comme la modification et les constructions à argument verbal. Par contre, un élément qui, malgré son statut de modifieur, ne peut apparaître qu'une seule fois, n'est pas prévu en TAG.

Rappelons brièvement les données sur les possibilités d'ordre des mots des modifieurs en allemand : dans le Mittelfeld, les adverbiaux peuvent intervenir entre tous les éléments, et on peut en trouver plusieurs d'affilée. Comme on le voit dans les exemples (226), tous les ordres des adverbiaux dans le Mittelfeld sont possibles (moyennant différentes interprétations communicatives). Par ailleurs, chaque élément adverbial peut aussi apparaître dans le Vorfeld comme on peut le voir dans les exemples (227). Par contre dans le Vorfeld, une seule de ces adjonctions est possible, comme l'illustrent les exemples (228).

- (226) a. Diese Frau schrieb am Morgen gerne ihre Reiseeindrücke am Frühstückstisch auf.  
cette femme écrivait le matin volontiers ses voyage\_impression à\_la table\_du\_petit\_déjeuner sur  
'Le matin, à la table du petit déjeuner, cette femme aimait de noter ses impressions de voyage'
- b. Diese Frau schrieb gerne ihre Reiseeindrücke am Frühstückstisch am Morgen auf.  
c. Diese Frau schrieb ihre Reiseeindrücke am Frühstückstisch am Morgen gerne auf.  
d. Diese Frau schrieb am Frühstückstisch am Morgen gerne ihre Reiseeindrücke auf.
- (227) a. Am Morgen schrieb diese Frau gerne ihre Reiseeindrücke am Frühstückstisch auf.  
b. Gerne schrieb diese Frau am Morgen ihre Reiseeindrücke am Frühstückstisch auf.  
c. Am Frühstückstisch schrieb diese Frau am Morgen gerne ihre Reiseeindrücke auf.
- (228) a. \*Am Morgen diese Frau schrieb gerne ihre Reiseeindrücke am Frühstückstisch auf.  
b. \*Am Morgen gerne schrieb diese Frau ihre Reiseeindrücke am Frühstückstisch auf.  
c. \*Gerne am Frühstückstisch am Morgen schrieb diese Frau ihre Reiseeindrücke auf.

Ces données de l'allemand nous amènent à poser deux prémisses pour le traitement en TAG :

- Le traitement habituel des adverbiaux et de considérer que le Vorfeld soit occupé par un syntagme adverbial grâce à l'adjonction de l'arbre élémentaire auxiliaire ancré par le syntagme adverbial. Cela revient à dire qu'on écarte les autres possibilités comme l'occupation du Vorfeld par un processus spécial, où les arbres élémentaires, spécialement conçus pour la tâche d'occuper le Vorfeld, sont substitués dans l'arbre verbal.
- Nous n'avons qu'un arbre auxiliaire pour les adverbiaux, indépendamment du champ topologique occupé par l'ancre de l'arbre auxiliaire : l'arbre auxiliaire ancré dans le Vorfeld est le même que celui ancré dans le Mittelfeld. D'un point de vue linguistique, il

est discutable de savoir si l'occupation du Vorfeld doit être traité comme un processus identique à la modification dans le Mittelfeld.<sup>170</sup> Par contre, dans notre optique de l'utilisation de TAG comme module entre la chaîne des mots et la structure sémantique, les modifications dans le Vorfeld et dans le Mittelfeld reçoivent les mêmes représentations dans l'arbre de dérivation. Nous ne faisons donc pas la distinction de placement dans les deux champs (nous considérons en effet que le choix de placement est véhiculé par la valeur communicative, ce qui n'est pas pris en compte par une structure sémantique basée sur l'arbre de dérivation).

Ces deux conditions, combinées à notre critère de minimalité de l'ensemble des schèmes, nous amènent à chercher un schème auxiliaire qui, en fonction de la structure des traits du nœud cible, s'adjoit une seule fois pour l'adjonction dans le Vorfeld, et un nombre illimité de fois pour l'adjonction dans le Mittelfeld.

En fait, on aimerait bien contrôler les adjonctions par des opérateurs des expressions régulières, tels que « ! » pour indiquer que l'adjonction à un nœud particulier doit être unique ou bien « \* » pour indiquer qu'on peut adjoindre zéro ou plusieurs arbres auxiliaires. Malheureusement, de telles expressions ne sont pas prévues pour l'adjonction en TAG ; l'adjonction est conçue pour des processus itératifs, donc qui peuvent être répétés à l'infini (comme l'opérateur « \* »), et seule l'interdiction complète d'une adjonction est prévue<sup>171</sup>, ce qui est nécessaire si l'on veut interdire, par exemple, l'adjonction à un nœud pied (voir section 2.2.2 *supra*). Cette interdiction est indiquée sur le nœud pied de l'arbre auxiliaire ① (Figure 125) par l'intermédiaire des lettres NA (abréviation de l'anglais *no adjunction*).

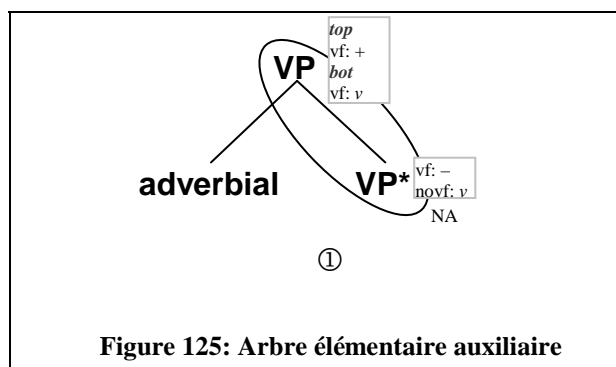
Au final, l'issue de secours se trouve, comme c'est souvent le cas en TAG quand les opérations de base s'avèrent insuffisantes, dans la structure des traits qui est attachée aux nœuds.<sup>172</sup> Cependant, même avec l'introduction des traits, la traduction générale de l'ensemble des opérateurs réguliers s'avère complexe. En effet, le fait qu'on ne puisse pas utiliser les opérateurs logiques dans la conception des traits rend complexe la tâche de la métagrammaire. Comme nous n'avons pas accès à l'opérateur « non » (« ¬ »), on est contraint à utiliser deux traits différents. Nous proposons comme solution l'arbre auxiliaire ① (Figure 90) ainsi que différentes configurations des traits conçus à partir d'opérateurs logiques réguliers, que nous présentons dans la Figure 126.

---

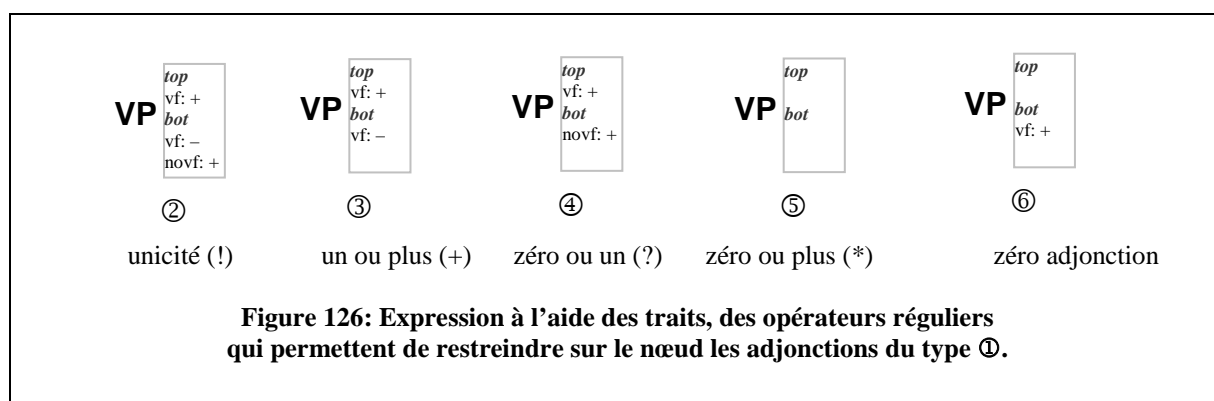
<sup>170</sup> Pour des avis contraires je réfère à toutes les approches du Vorfeld en GB, par exemple Sternefeld 2001, ou encore l'analyse LFG habituelle de l'allemand Butt *et alii* 1999.

<sup>171</sup> Des idées qui proposent d'exprimer directement des restrictions régulières d'adjonction apparaissent dans la TAGML, le format standardisé d'encodage des arbres apparaissant dans les analyses TAG, Lopez 2000. Cependant, à mes connaissances, aucun outil informatique TAG ne les prend en compte pour l'instant. Nous verrons ci-dessous qu'il est possible de simuler les expressions régulières à l'aide des traits et que de tels ajouts n'augmentent pas le formalisme.

<sup>172</sup> Une discussion intéressante qu'on ne peut malheureusement pas approfondir ici, est la question de savoir comment évaluer à quel point les traits doivent être considérés comme faisant partie intégrale du formalisme TAG ou comme un ajout, greffé au formalisme. Les analyses linguistiques TAG existantes n'ont jamais développé en détail la fonction des traits dans le formalisme. Ils jouent pourtant un rôle essentiel dans le traitement de nombreux phénomènes linguistiques. En même temps, ils posent des problèmes pour l'algorithme du parsing TAG. (Voir Lopez 1999 pour les détails de l'algorithme. Le passage d'information des traits d'un arbre via un 2<sup>e</sup> à un 3<sup>e</sup> est en particulier algorithmiquement délicat. Un tel passage a lieu par exemple pour l'accord du sujet dans les constructions du passé composé, où le sujet passe ses valeurs d'accord via l'arbre du participe passé à l'arbre de l'auxiliaire (Cf. section 4.3.4 *infra*)

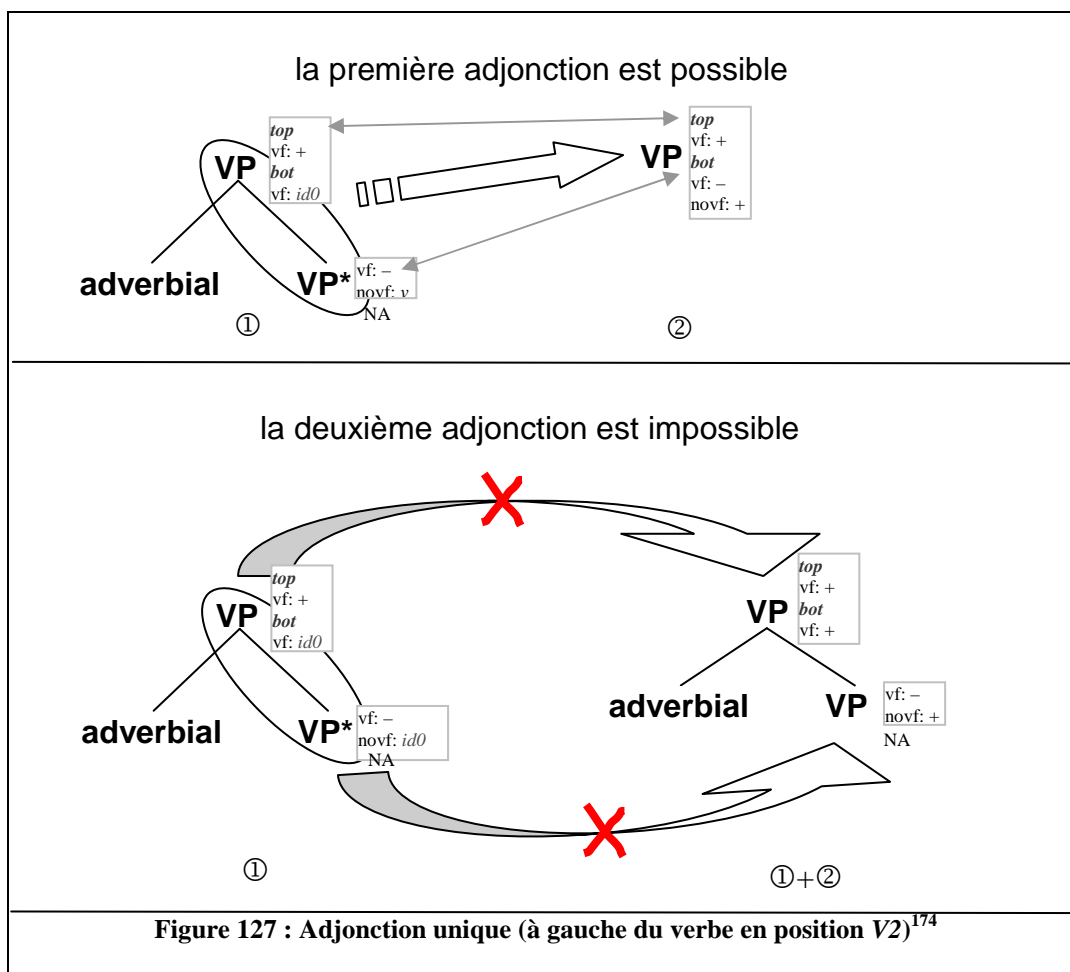


Afin d'expliciter le fonctionnement des traits utilisés dans la méta-grammaire, nous montrons dans la Figure 127 comment est exprimée l'unicité de l'adjonction, dont nous avons besoin pour l'adjonction dans le Vorfeld. Nous utilisons la séparation des traits amont et aval<sup>173</sup> d'un nœud cible pendant l'adjonction d'un arbre auxiliaire. Ce nœud cible *VP* sera la racine de tous les schèmes verbaux où le Vorfeld n'est pas rempli par un argument du verbe.



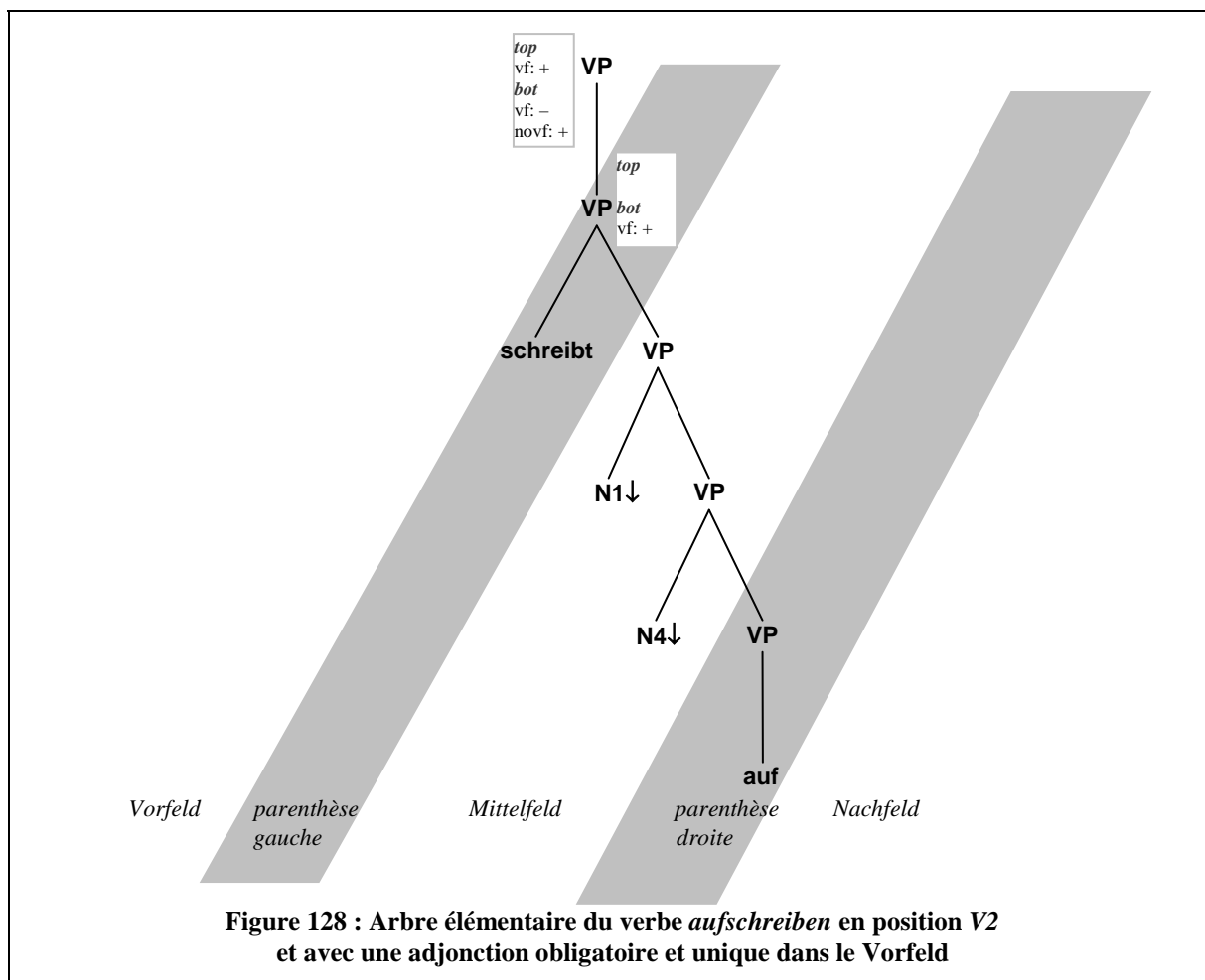
<sup>173</sup> Dans les représentations des arbres, tirées directement de la compilation avec la méta-grammaire, nous utilisons les termes anglais, *top* pour *amont*, et *bot* (comme *bottom*) pour *aval*.





Le processus d'adjonction unique fonctionne comme suit : le nœud cible porte un trait du nom *vf* (comme *Vorfeld*) avec la valeur positive en amont (*bot* : *vf* : +) et la valeur négative en aval (*top* : *vf* : -). Ces différentes valeurs rendent impossible l'unification des traits en amont avec les traits en aval sans qu'une adjonction ait lieu à ce nœud. De plus, en aval, nous avons le trait inverse du trait *vf*, le trait *novf* avec la valeur +. Nous avons besoin de ce trait pour la seule raison que nous n'avons pas d'opération de négation dans les restrictions portant sur les traits. Lorsqu'une adjonction a lieu sur le nœud ②, les traits en amont s'unifient avec la racine de l'arbre auxiliaire ①, et les traits en aval s'unifient avec les traits du nœud pied. C'est ainsi que la variable *v* passe sa valeur du trait *novf* au trait *vf* en aval au nœud racine. Le résultat de cette adjonction est représenté dans l'arbre ①+② de la Figure 127. Cet arbre a à sa racine les mêmes valeurs en amont et en aval pour le trait *vf* et la répétition d'une telle adjonction au nœud racine n'est pas possible, car l'adjonction exige que le trait *vf* en aval soit négatif. Une adjonction au nœud pied est de toute manière exclue. Dans la Figure 128, nous présentons un arbre élémentaire ancré par *aufschreiben* 'noter' qui prend en compte les exemples (227).

<sup>174</sup> La coïndexation des valeurs des variables est indiquée par des variables notées *id* suivies d'un index. Cette notation provient des schèmes générés à l'aide du compilateur de Gaiffe 2002.



Cet arbre porte comme racine le nœud ② introduit ci-dessus. Le fait qu'on garde ce nœud en tant que nœud unaire (avec un seul enfant) n'est pas obligatoire<sup>175</sup>. De plus, ce nœud au dessus du verbe jouera un rôle important dans le traitement des arguments verbaux. La valeur positive du trait *vf* en aval sur le nœud *VP* dominant le verbe interdit une adjonction à ce niveau. C'est ainsi qu'une adjonction multiple dans le Vorfeld est interdite. Par contre, sur les autres nœuds *VP* de la colonne, nous n'avons pas de restrictions d'adjonction, et les arbres auxiliaires modificateurs peuvent s'adjoindre l'un sur l'autre. Nous pouvons également voir la nécessité du nœud *VP* au dessus de la particule *auf* : il permet de garder le même type d'adjonction pour toutes les places dans le Mittelfeld.

Notons que cette structure de bifurcation droite permet de se contenter d'un seul arbre par modifieur de phrase : toute modification adverbiale se fait par la gauche sur les nœuds *VP*. Cela veut dire que contrairement à la Grammaire TAG du français, nous ne faisons pas la distinction entre la modification au niveau du verbe et la modification au niveau de toute la phrase. En fait, nous considérons qu'une structure de constituants ne peut tout simplement pas exprimer les questions de portée des modificateurs. Si on tentait de le faire, il faudrait développer une structure de bifurcation à droite et c'est la solution que nous avons choisie pour notre grammaire. Les modificateurs de phrase comme par exemple la négation par *nicht*, modifient généralement le ou les constituants qui suivent cette négation. Par contre, si le modifieur se trouve tout à la fin d'une

<sup>175</sup> En effet, on aurait pu unifier le nœud avec son enfant et on aurait obtenu les mêmes effets, mais, comme nous le verrons dans la section concernant les classes de la méta-grammaire (section 4), l'arbre que nous proposons est plus facilement séparable en descriptions partielles, donc en classes de la méta-grammaire (voir également *supra* la discussion sur le Critère de la représentation modulaire).

phrase laissant une parenthèse droite vide (exemple (229)), le modifieur porte sur toute la phrase. Pourtant, le modifieur se trouve clairement dans le Mittelfeld comme le montre le contraste entre (230)a et (230)b<sup>176</sup>. Une adjonction de l'arbre auxiliaire ne peut donc pas avoir lieu à un nœud supérieur à la parenthèse droite (d'autant plus que cela rendrait impossible l'analyse en TAG des relatives extraposées, comme nous le verrons dans la suite). Par ailleurs, plusieurs choix d'adjonction compliquent les restrictions de placement qui découlent de la topologie de la phrase. Nous nous contenterons donc de l'analyse de la modification unique par la gauche. Nous renvoyons à Gerdes 1998 pour une discussion détaillée sur l'interprétation sémantique d'une telle structure.

(229) Peter liebt diese Frau nicht.  
Peter aime cette femme pas  
'Peter n'aime pas cette femme'

(230) a. Peter schließt die Tür nicht ab  
Peter ferme\_à\_clé la porte non PARTICULE  
'Peter ne ferme pas la porte à clé'

b. \*Peter schließt die Tür ab nicht  
Peter ferme\_à\_clé la porte PARTICULE non

Malgré les avantages formels d'une structure de bifurcation droite dans les descriptions de l'allemand, il faut noter que dans le cas des verbes simples sans particule dans la position V2 (et dans ces cas uniquement), nous aboutirons à un arbre élémentaire d'apparence inhabituelle : la parenthèse droite est un nœud feuille non lexicalisé qui n'est en fait ni un nœud de substitution, ni un nœud pied. Nous développerons par la suite les avantages de cette structure : d'une part, la parenthèse droite vide marque une frontière (syntaxique et prosodique) importante dans la structure de l'allemand<sup>177</sup>. D'autre part, cette analyse est essentielle pour le regroupement en classes d'une métagrammaire comme nous le verrons dans la section 4. En fait, dans le cas d'un verbe V2 sans argument verbal ni particule (les seuls cas où la parenthèse droite peut être vide<sup>178</sup>), si introduire des nœuds feuille simples est gênant pour un outil utilisant la grammaire TAG, alors il est très facile d'ajouter un nœud vide (E) au nœud VP de la parenthèse droite<sup>179</sup>. Un tel nœud vide aurait un caractère intrinsèque au formalisme TAG et n'impliquerait pas une analyse basée sur un quelconque mouvement.

### 4.3.3 Les relatives en TAG

Le traitement des phrases relatives fait clairement apparaître les limites du formalisme TAG. Dans la section 2.4.4, nous avons soulevé deux problèmes distincts concernant le traitement des relatives en TAG : leur structure interne et leur positionnement dans la phrase.

Bien que notre analyse des relatives en TAG suit dans les grandes lignes les propositions de Kroch & Joshi 1986 pour XTAG et de Abeillé 1991, 2002 pour FTAG, nous avons été obligé de faire des ajustements afin de pouvoir traiter des données de l'allemand qui sont très différentes de celles du français.

<sup>176</sup> C'est d'ailleurs ce phénomène qui amène certains générativistes à considérer que le verbe en position V2 était déplacé à partir de la position finale (Grewendorf *et alii* 1987).

<sup>177</sup> Pour des preuves de l'existence de la parenthèse droite vide, voir Müller 1999:174.

<sup>178</sup> Nous ne traitons pas dans l'analyse TAG l'extraposition avec parenthèse droite vide, ce qui peut être une analyse possible de l'exemple suivant :

[Ich]<sub>VF</sub>[versuche]<sub>I</sub>[ ] [das Buch zu lesen]<sub>NF</sub>.

Pour une discussion plus détaillée, voir section I.6 .

<sup>179</sup> Il faudrait créer une classe spécifique dans la métagrammaire pour cette sorte de verbe, ce qui rendrait la description un peu moins générale.

Ainsi, dans le cas de la relativisation d'arguments verbaux, nous n'avons pas pu garder l'analyse proposée pour le français, qui présente le pronom relatif comme une co-tête. Nous avons opté pour une analyse où ce dernier est substitué dans l'arbre du verbe de la même manière qu'un argument plein. La raison de ce choix réside dans le fait que le pronom relatif fait l'accord en genre et en nombre avec l'antécédent<sup>180</sup>. Le cas grammatical du pronom relatif est déterminé à l'intérieur de la proposition relative par son gouverneur syntaxique, qui peut être un verbe ou une préposition. En français, il existe des constructions où le du pronom est régi par le verbe de la relative (les pronoms *qui*, *dont*, *à qui*, que respectivement pour le nominatif, le génitif, le datif et l'accusatif), et seuls certains pronoms (lequel, duquel etc.) font l'accord en genre et en nombre avec l'antécédent (mais la relativisation par le biais de ces pronoms ne sont pas traités dans la FTAG de Abeillé 1991:158). Par contre, en allemand, le pronom fait systématiquement l'accord en nombre et en genre avec l'antécédent. Si dans l'arbre élémentaire de la proposition relative, on encodait la forme du pronom en dur (c'est-à-dire en co-tête), on aboutirait à un très grand nombre d'arbres élémentaires pour ces propositions, car on aurait un arbre par forme du pronom. Ainsi, le verbe *schlafen* 'dormir' des phrases (231) à (234), aurait besoin d'un arbre élémentaire différent pour l'analyse de chacune de ces phrases.

- (231) der Mann, der[nominatif, sg, masc] schläft.  
'l'homme qui dort'
- (232) die Frau, die[nominatif, sg, fem] schläft.  
'la femme qui dort'
- (233) das Kind, das[nominatif, sg, neut] schläft.  
'l'enfant qui dort'
- (234) die Menschen, die[nominatif, pl] schlafen.  
'les êtres humains qui dorment'

L'accord du pronom relatif avec l'antécédent est réalisé à l'aide des variables liant les traits du genre (*gen*) et de nombre (*nb*) présents dans le nœud pied de l'arbre relatif avec les traits du nœud de substitution du pronom. Or le cas grammatical est fixe dans un arbre donné, car il fait partie de la sous-catégorisation du verbe. Ces deux accords séparés illustrent bien le double rôle syntaxique (transférème et anaphorème, voir section II.4) que remplit le pronom relatif. Remarquons aussi que l'accord en personne (*pers*) entre le nœud de substitution du pronom relatif et le verbe, force le verbe à être à la 3<sup>e</sup> personne pour les pronoms relatifs habituels qui portent ce trait.<sup>181</sup> Nous considérons que le traitement de la 3<sup>e</sup> personne dans les relatives, qui peut aussi, rappelons-le, servir pour la modification des pronoms à la 1<sup>ère</sup> et à la 2<sup>e</sup> personne, est tout à fait suffisant pour les besoins des projets actuels de génération de texte.

---

<sup>180</sup> Mais pas en cas grammatical, sauf quand un autre pronom personnel fait cet accord, voir section I.1 pour une discussion plus détaillée.

<sup>181</sup> Lorsque les pronoms sont séparés en une partie complémenteur et une partie anaphore (section II.4.2.), l'analyse est également possible avec les arbres élémentaires proposés, mais elle n'est pas tout à fait satisfaisante, car il n'y a qu'un seul nœud de substitution qui est proposé pour les deux parties du pronom relatif : l'arbre ancré par la partie complémenteur possède obligatoirement un nœud de substitution pour le pronom personnel, ce qui gênerait l'interprétabilité sémantique de l'arbre de dérivation qui en résulte. Cependant, ce défaut nous semble sans grande importance dans une application concrète car la complexité communicative de telles structures le met de toute manière hors jeu de tout système existant de génération.

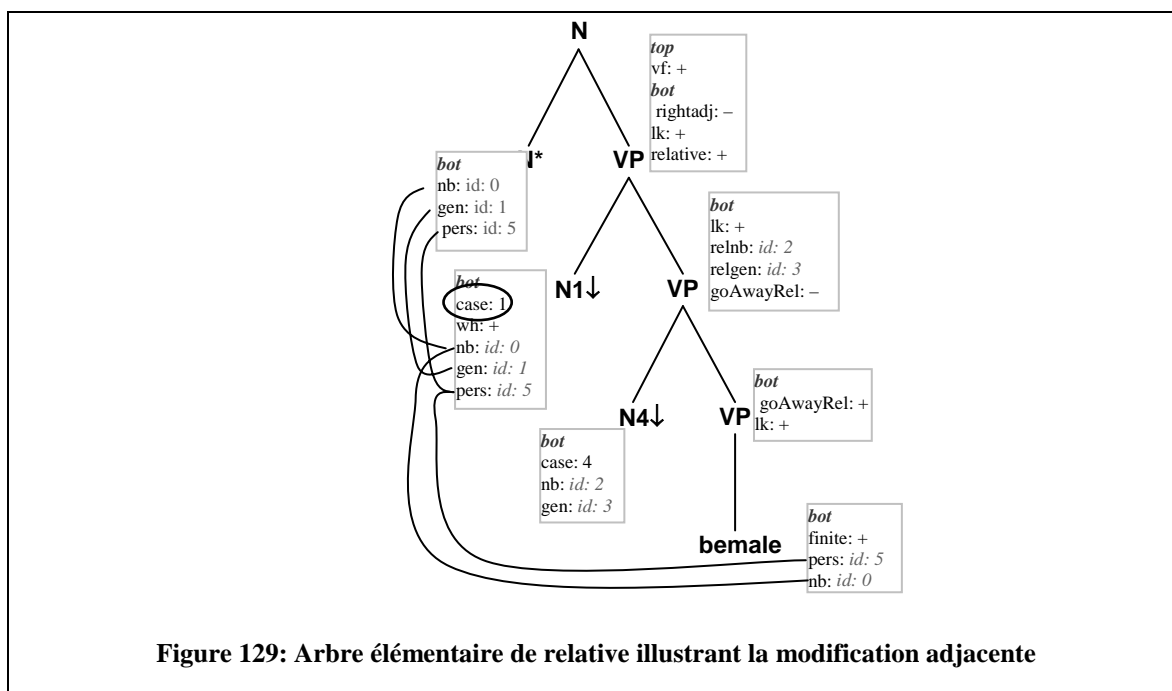


Figure 129: Arbre élémentaire de relative illustrant la modification adjacente

L'allemand et le français n'ont par ailleurs pas le même comportement en ce qui concerne la place de la relative dans la phrase, et les arbres élémentaires doivent être différents dans les deux langues. Comme nous l'avons vu dans la section II.4, les cas d'extraposition de relative sont beaucoup plus fréquents en allemand qu'en français : aucun système de génération traitant la proposition relative en allemand ne peut éviter d'inclure une analyse de l'extraposition.<sup>182</sup> Comme la réalisation d'une modification de longue distance d'un élément substitué est impossible en LTAG (section 2.4.4), nous devons avoir recours à un « hack »<sup>183</sup> : il s'agit

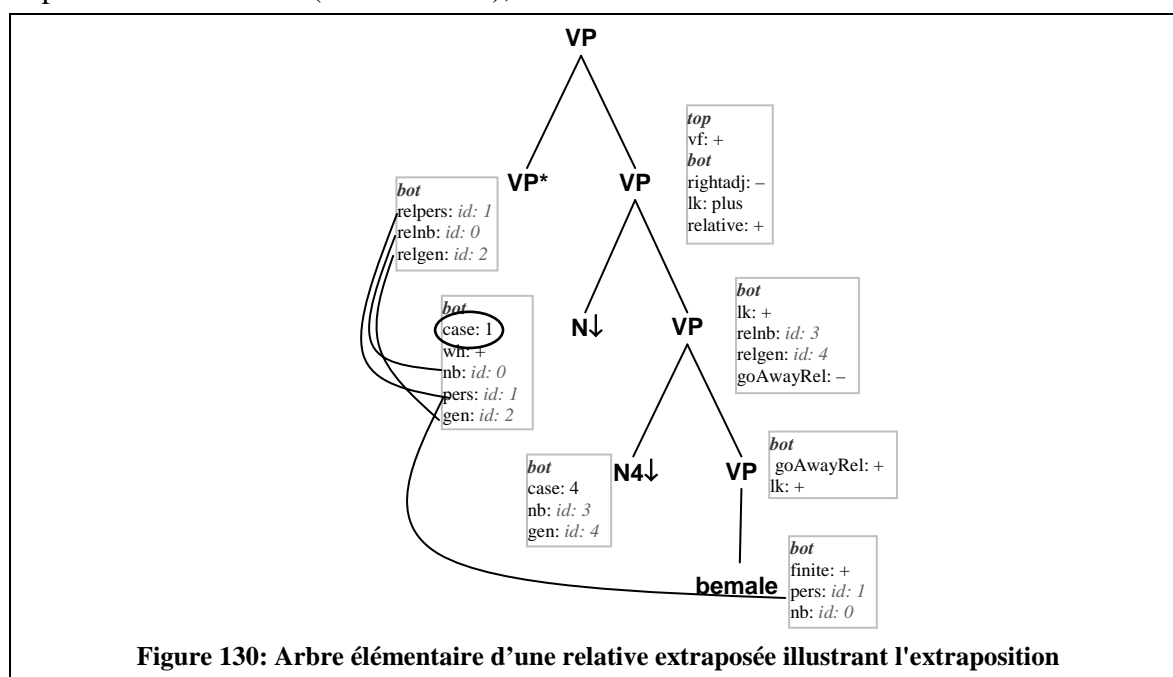


Figure 130: Arbre élémentaire d'une relative extraposée illustrant l'extraposition

<sup>182</sup> Au début du projet SmartKom, on a donné au groupe de génération 20 phrases d'exemples illustrant le type de phrases que le système devra générer un jour. Parmi ces 20 phrases, on trouvait déjà trois exemples contenant des relatives extraposées.

<sup>183</sup> Ce terme, des fois traduit par le mot *bidouillage*, appartient « à l'argot des informaticiens » et recouvre le concept d'un « assemblage temporaire (matériel ou logiciel) d'éléments hétéroclites, grossier et rudimentaire, conçu sans respecter les règles prescrites. Cet assemblage peut être perçu comme ingénieux, s'il est le fait d'un informaticien amateur, par exemple, ou, au contraire, comme désolant, s'il est le fruit du

d'utiliser les traits lexicaux pour le passage des informations lexicales dont la relative extraposée a besoin. Concrètement, nous n'avons pas d'autres choix que de proposer un nœud par élément nominal apparaissant, qui soit accessible à partir du Nachfeld et qui porte l'information d'accord. La structure de la colonne binaire est pour cette raison essentielle, car elle permet à chaque nœud nominal de posséder son propre nœud dans la colonne, sur lequel il peut passer l'information de genre et de nombre, dont son modifieur relatif dans le Nachfeld a besoin. Nous avons donc un premier arbre relatif pour une modification adjacente de la relative, portant des nœuds racine et pied de catégorie *N*, et un deuxième arbre relatif pour l'extrapolation de la relative qui s'adjoit sur un nœud de catégorie *VP*.

Considérons donc un cas d'extrapolation illustré sous (235). Il s'agit d'un exemple que nous avons déjà présenté dans la section 2.4.6.

- (235) Peter schließt die Tür ab, die ich bemale.  
 Pierre ferme la porte que je peins PARTICULE  
 'Pierre ferme à clef la porte que j'ai peinte hier'

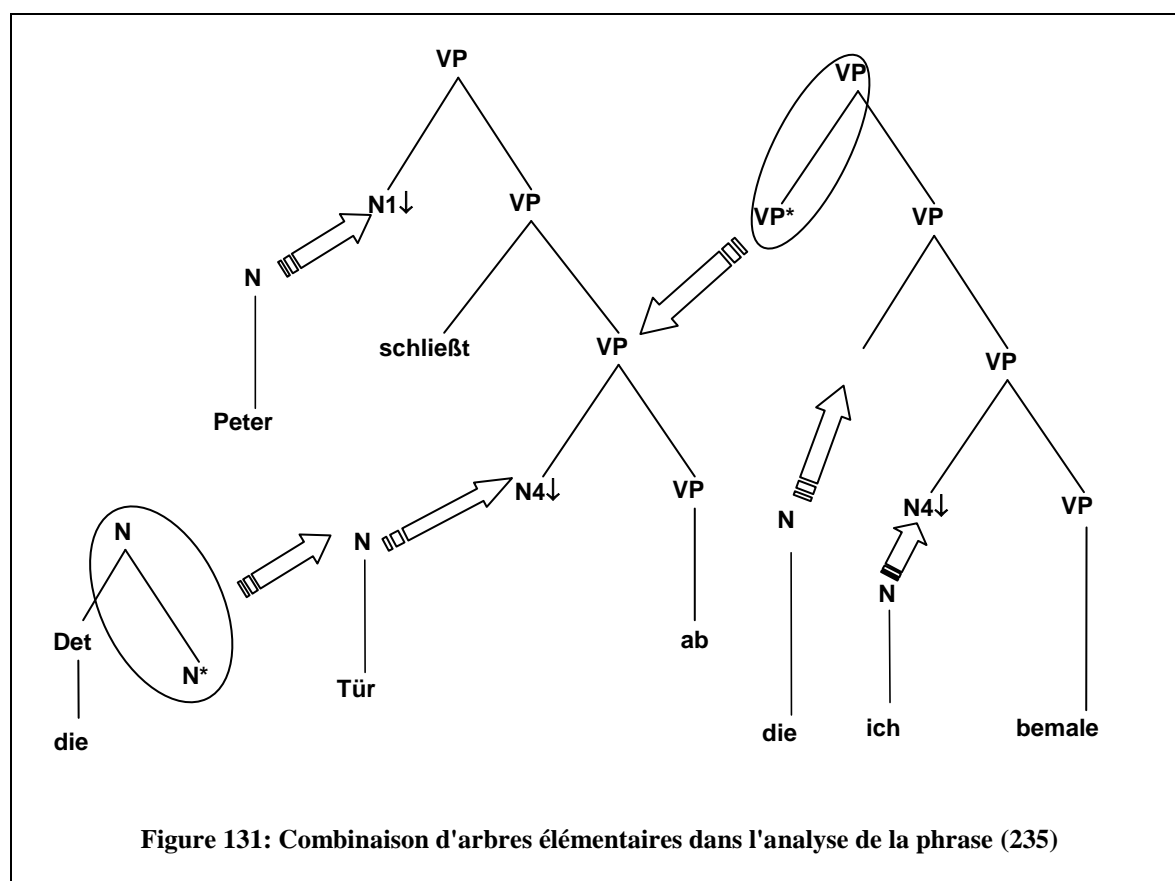


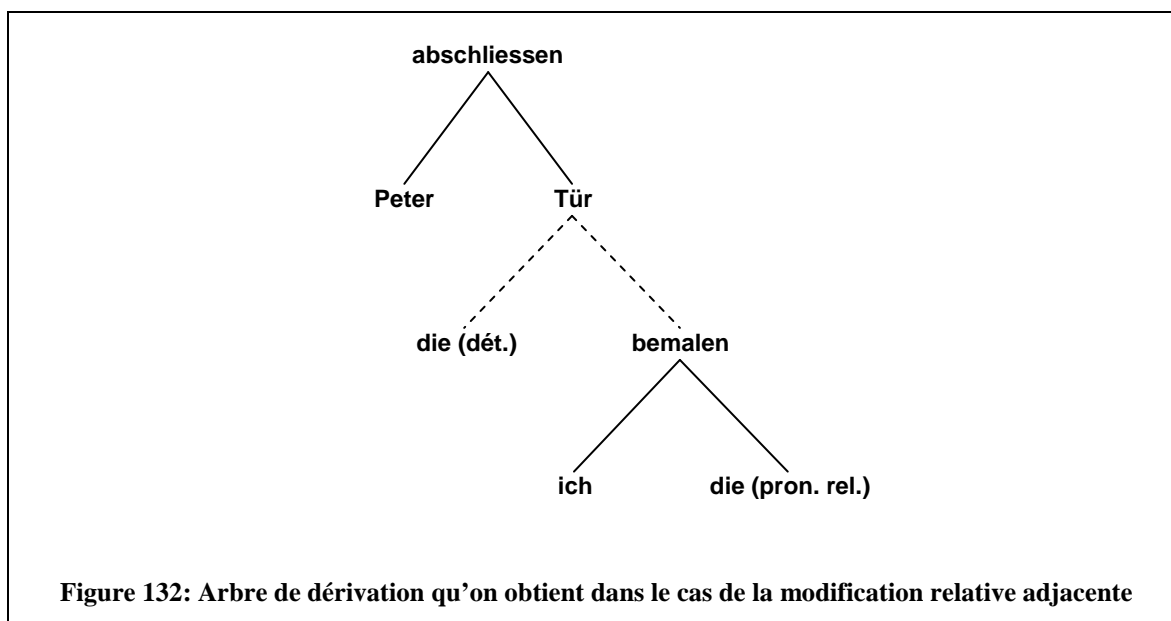
Figure 131: Combinaison d'arbres élémentaires dans l'analyse de la phrase (235)

Le verbe avec particule *abschließen* 'fermer à clé' ancre le même arbre élémentaire que le verbe *aufschreiben* de la Figure 123. Chaque nœud *VP* qui se trouve sur la colonne porte des traits *relnum* et *relgen*, que l'arbre de la relative passe au nœud de substitution pour le pronom relatif, ce qui permet au pronom de faire l'accord avec son gouverneur « distant ». Il est donc possible de réaliser une telle relation à l'aide des traits. Par contre, le prix à payer est que l'arbre de dérivation ne reflète pas le lien de gouvernance entre le nom et son modifieur : dans l'exemple donné ci-dessus, le SN *die Tür* 'la porte' est substituée dans l'arbre de *abschließen* 'fermer à clé', tandis que le modifieur de *die Tür*, l'arbre de la relative ancré par *bemalen* 'peindre',

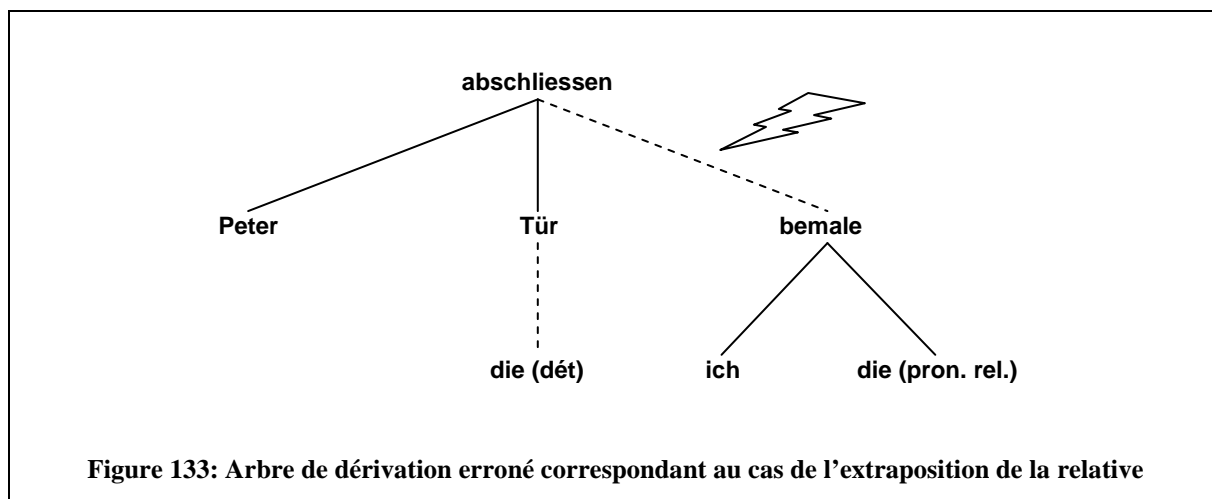
travail d'un mauvais informaticien professionnel. » Office de la langue française, 2000. Je laisse au lecteur le soin d'apprécier lui-même de quel type de « hack » il s'agit ici...

s'adjoit à l'arbre de *abschließen* et non à l'arbre de son gouverneur *die Tur*. Nous illustrons par la Figure 131, la combinaison des différents arbres élémentaires permettant d'obtenir la phrase (235).

Les conséquences de cette structure qui n'est motivée ni syntaxiquement ni sémantiquement sont assez graves, aussi bien pour la génération que pour l'analyse. En effet, un système qui veut utiliser cette grammaire pour la génération doit faire un précalcul sur le poids de la relative (éventuellement en comparaison avec le poids de la parenthèse droite). Dans le cas où l'extraposition est inévitable, l'arbre de dérivation (ayant un caractère sémantique) qui sert d'entrée à la grammaire doit être modifié et faire en sorte que la relative monte jusqu'au premier verbe qu'il rencontre.



Ainsi, pour analyser la phrase donnée sous (235), nous devons modifier l'arbre de dérivation que l'on obtient dans le cas de la modification relative adjacente (Figure 132) afin de pouvoir générer une phrase avec extraposition : le nœud de *bemalen*, modifiant *die Tur* doit monter au nœud de *fermer à clé*, comme il apparaît dans l'arbre de la Figure 133. Par ailleurs, cet arbre doit en plus (et c'est ici qu'on voit bien la problématique d'un arbre de dérivation erroné) amener avec lui l'information sur l'accord avec son gouverneur *porte*. Cela est indispensable parce que l'arbre du verbe sur lequel l'arbre de la relative doit s'adjoindre porte plusieurs nœuds dans sa colonne, présentant ainsi à la relative plusieurs nœuds où cette dernière pourrait s'adjoindre. Pour que la relative arrive à choisir le bon nœud où s'adjoindre, il faut mettre l'adresse du 'bon' nœud de la colonne dans l'arbre de dérivation. Un tel pré-calcul n'est donc pas trivial, car il mêle les changements dans l'arbre de dérivation avec l'information sur la structure des arbres élémentaires qui correspondent à chaque nœud de l'arbre de dérivation.



D'un point de vue de l'analyse, une structure de sortie qui ne contient que l'arbre de dérivation de la Figure 133 n'est pas interprétable, car la modification du verbe *abschließen* par *bemale* n'a pas de sens, et un module de post-calcul qui essaie de récupérer le sens dans le sens inverse du module décrit ci-dessus ne saurait pas qui (*die Tür* ou *Peter*) est peint. Il faudrait donc garder en mémoire quelque part que le nœud sur lequel *bemale* s'est adjoint est bel et bien le nœud de la colonne appartenant à *die Tür*. Une possibilité est de donner un numéro à chaque argument et de mettre un trait sur le nœud de la colonne qui porte ce numéro, ce que nous avons fait pour nos arbres. Le traitement de ces numéros pour la récupération de la vraie relation sémantique constitue en fait un formalisme à part. Développer un formalisme permettant de récupérer les informations perdues pendant l'analyse TAG n'a pas beaucoup de sens. Nous considérons qu'il est préférable de « bidouiller » l'outil, s'il s'agit juste d'un besoin urgent d'une application qui doit tourner avec la grammaire TAG. Nous remettons à plus tard la réflexion sur des alternatives, qui permettent un traitement généralisé des phénomènes de longue distance (Cf. troisième section du quatrième chapitre).

#### 4.3.4 Les compléments verbaux en TAG

Dans cette sous-section, nous présentons les schèmes que nous avons retenu pour le traitement des verbes à argument verbal dits *verbes matriciels*. Comme indiqué dans la section 4 intitulée "les objectifs de DTAG", nous traiterons uniquement deux possibilités de placement des arguments verbaux, le placement dans la parenthèse droite (qui est la position par défaut) et l'extraposition dans le Nachfeld. Avec ces deux possibilités de placement, nous arrivons à traiter tous les types de verbes principaux, car, comme nous l'avons vu dans la section II.3, il existe des verbes qui se construisent obligatoirement de manière « cohérente » (en formant un complexe verbal), et d'autres qui se construisent obligatoirement de manière « incohérente », (via l'extraposition dans le Nachfeld).

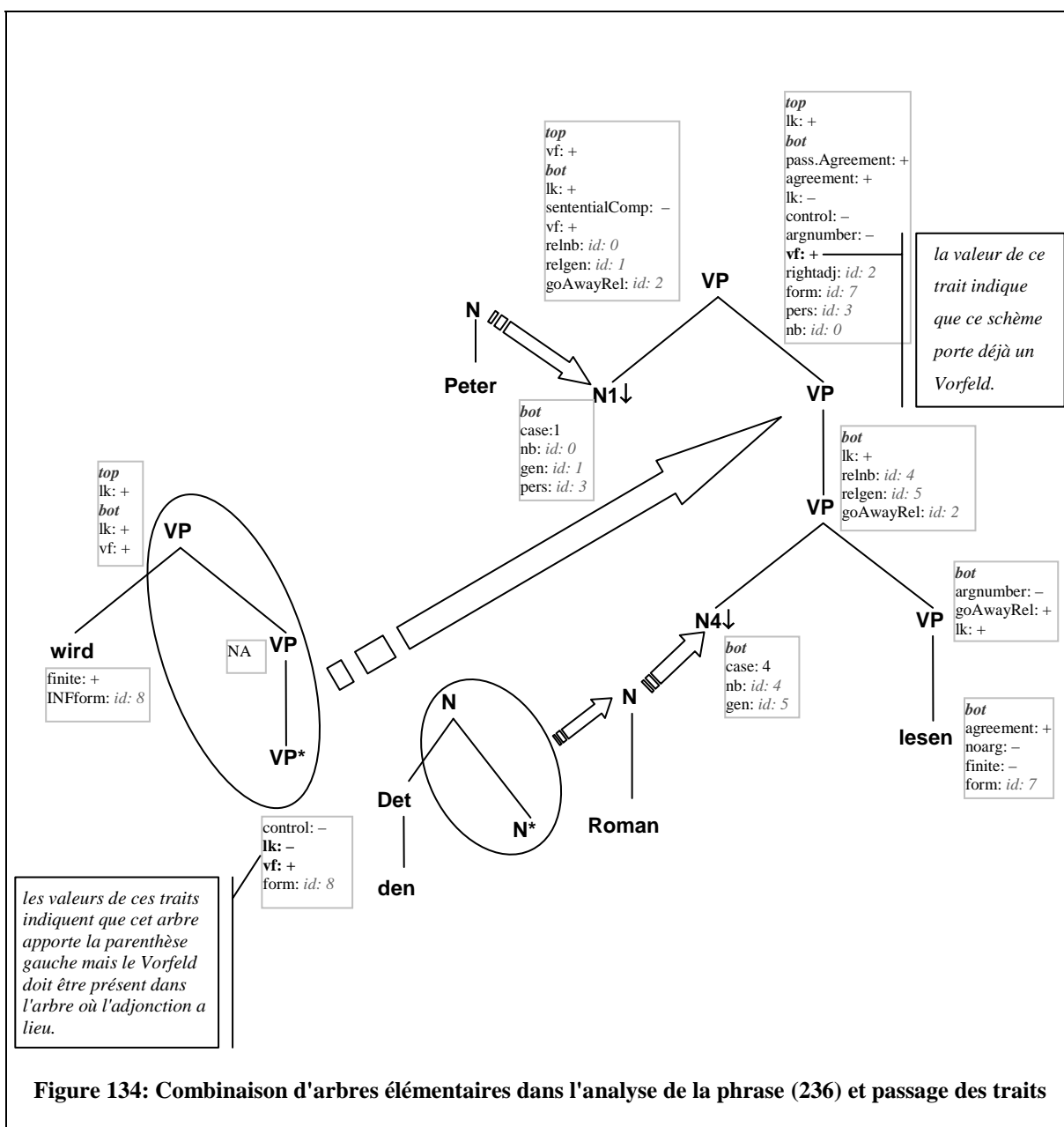
Dans une perspective d'utilisation de la grammaire de l'allemand dans un système de génération, parallèle aux TAG développées pour d'autres langues, nous avons préservé autant que possible l'analyse faite pour le français par Abeillé 1991, 2002. Ainsi, nous avons retenu le fait qu'un verbe matriciel s'adjoint à son dépendant verbal<sup>184</sup>. L'autre possibilité, c'est-à-dire l'utilisation de la substitution, est exclue en allemand pour les mêmes raisons que pour le français et l'anglais. En effet, il ne serait pas possible de traiter des cas de dépendance de longue distance à partir du verbe enchâssé ; par exemple, on ne pourrait pas analyser une phrase où l'argument nominal du verbe enchâssé est placé dans le Vorfeld.

<sup>184</sup> La seule exception vient des verbes à contrôle avec argument ou modifieur relativisé, voir section 2.4.4.



Une telle analyse divise les verbes matriciels en deux groupes. Un premier groupe est formé par les auxiliaires et les verbes à montée qui ne sous-catégorisent pas de sujet et qui font donc l'accord via l'arbre élémentaire du verbe enchâssé. Le deuxième groupe comporte les arbres des formes finies des verbes à contrôle qui portent un nœud de substitution pour leur sujet. En conséquence de cette dichotomie, les arbres des verbes enchâssés devront être doublés, d'une part en arbres qui portent le sujet (afin de permettre l'adjonction des auxiliaires ou des verbes à montée) et d'autre part, en arbres qui ne portent pas de sujet (afin de permettre l'adjonction des verbes à contrôle).

Une deuxième condition que nous nous sommes fixés était de permettre à tous les arguments et même aux modifieurs des verbes enchâssés, d'apparaître dans le Vorfeld. Or, la gestion du placement des dépendants dans le Vorfeld représente une complication considérable pour l'analyse. Notons par ailleurs que nous avons délibérément choisi de ne pas traiter le brouillage d'arguments dans le Mittelfeld même si certains phénomènes peuvent être analysés par TAG (c'est le cas par exemple, de l'échange de la place des arguments entre deux verbes voisins).



Le cas le plus simple d'une construction à argument verbal est celle d'un temps composé, comme par exemple le futur. Dans ce cas, l'infinitif doit porter tous les arguments verbaux, incluant le

sujet. Nous considérons qu'un auxiliaire ne peut pas être modifié par un adverbial ; remplir le Vorfeld est une tâche qui doit donc être accomplie par l'arbre de l'infinitif.

- (236) Peter wird den Roman lesen  
 Pierre va le roman lire  
 'Pierre lira le livre'

Considérons d'abord l'analyse de la phrase à ordre des mots dit standard :

Pour obtenir l'arbre de *lesen* 'lire', il faut donc faire en sorte que l'arbre auxiliaire de *wird* 'va' se place dans la position V2, c'est-à-dire entre l'argument au nominatif *Peter* et l'argument à l'accusatif *den Roman* 'le roman'. Comme TAG ne possède pas d'outil spécifique pour gérer le placement global à l'intérieur d'une phrase, nous simulons ces places globales sur la colonne de l'arbre des verbes enchâssés. Nous gardons donc le 'nœud colonne' *VP* de la parenthèse gauche dans l'arbre dont l'un des dépendants occupe le Vorfeld.

Dans le cas de l'exemple (236), le verbe *lesen* 'lire', dont l'argument nominal *Peter* occupe le Vorfeld, ancre un arbre avec un nœud unaire *VP* à la position de la parenthèse gauche. Le nœud *VP* de la parenthèse gauche a, pour le trait *lk* (abréviation de l'allemand : *linke Klammer*, 'parenthèse gauche') des valeurs en amont et en aval qui sont différentes. De plus, ce même nœud porte en aval un trait *vf* (Vorfeld) positif, qui indique que l'arbre s'occupe déjà du Vorfeld et donc que seul un arbre sans Vorfeld peut s'y adjoindre. Nous présentons l'analyse TAG de cette phrase dans la Figure 134.

L'arbre de la phrase (237) a la même structure qu'une phrase avec des nœuds de substitution qui ont été échangés. Dans le cas de la phrase (238), les deux arguments nominaux de *lesen* se trouvent dans le Mittelfeld et l'adverbe *Morgen* 'demain' modifie *lesen*. C'est ainsi que l'arbre de *lesen* possède un autre nœud avec adjonction obligatoire, le nœud racine. Ce dernier porte les mêmes traits que le nœud racine de l'arbre V2 avec modifieur dans le VF, présenté *supra*. Il n'y a donc qu'une seule adjonction possible à ce nœud racine.

- (237) Den Roman wird Peter lesen.  
 le roman va Pierre lire  
 'le roman, Pierre le lira' ou 'C'est le roman que Pierre lira'

- (238) Morgen wird Peter den Roman lesen.  
 Demain va Pierre le roman lire  
 'Demain, Pierre lira le roman'

Passons maintenant à la construction des schèmes des verbes à contrôle. Ces derniers se construisent de la même manière que les verbes auxiliaires ou à montée<sup>185</sup>, moyennant le fait que, cette fois-ci, le sujet est argument du verbe matriciel et que le verbe matriciel peut lui même être modifié, soit devant le verbe dans la position de Vorfeld, soit après le verbe et donc dans le Mittelfeld. Nous ne traitons évidemment pas le cas du brouillage de modifieurs dans le Mittelfeld et nous posons comme condition que le modifieur doit être adjacent soit au verbe qu'il modifie, soit à l'un des arguments du verbe.

- (239) Peter versucht den Roman zu lesen.  
 Pierre essaie le roman à lire  
 'Pierre essaie de lire le roman'

---

<sup>185</sup> D'un point de vue topologique, la deuxième analyse où l'infinitif est extraposé dans le Nachfeld est préférable, mais dans la grammaire TAG, nous ne faisons pas la distinction entre les deux cas, et les deux analyses coïncident.

- (240) Heute versucht Peter den Roman zu lesen.  
 Aujourd'hui essaie Pierre le roman à lire  
 'Aujourd'hui, Pierre essaie de lire le roman'

Deux analyses sont possibles pour la phrase (240). Dans la première analyse, le verbe matriciel *versucht* est responsable du Vorfeld faisant en sorte que ce soit lui qui est modifié par l'adverbe *Heute* 'aujourd'hui' ; dans le deuxième cas, c'est l'arbre de *zu lesen* 'de lire' qui possède le nœud du Vorfeld et c'est donc lui que l'adverbe *Heute* modifie. Cependant, dans cet exemple, cette distinction ne semble pas détenir une grande importance. Par contre, dans d'autres exemples comme celui que nous présentons ci-dessous (exemple (241)), la différence est plus claire.

- (241) Morgen versprach Peter den Roman zu lesen.  
 Demain promettait Pierre le roman à lire  
 'Demain, Pierre a promis de lire le roman'

Dans l'exemple suivant (242), qui illustre un cas de complexe verbal composé des deux verbes *zu lesen* 'de lire' et *versuchen* 'essayer', nous avons affaire à un enchâssement du verbe dans la parenthèse droite. Une question essentielle qui se pose à ce niveau est la question de savoir où a lieu l'adjonction de l'arbre du verbe matriciel qui vient de droite. Topologiquement, il est clair qu'une analyse qui crée une unité de constituants dans le Vorfeld est préférable ; autrement dit, on opterait pour une adjonction directement sur le nœud colonne de la parenthèse droite.

- (242) Morgen wird Peter den Roman zu lesen versuchen.  
 Demain va Pierre le roman à lire essayer  
 'Demain, Pierre essayera de lire le roman'

En fait, l'inconvénient d'une telle analyse est que l'arbre de dérivation ne suit pas. La construction d'un syntagme de parenthèse droite, marqué dans la figure par le point d'interrogation, provoque plusieurs effets qui ne sont pas souhaités. D'abord, les places les moins marquées du sujet se trouvent dans une place adjacente (à gauche ou à droite) à la parenthèse droite. Dans le cas de notre exemple, cette analyse permettrait d'obtenir la phrase (243) où les deux arguments verbaux ont mutuellement échangé leur place, mais pas la phrase (242) qui présente un ordre des mots plus neutre.

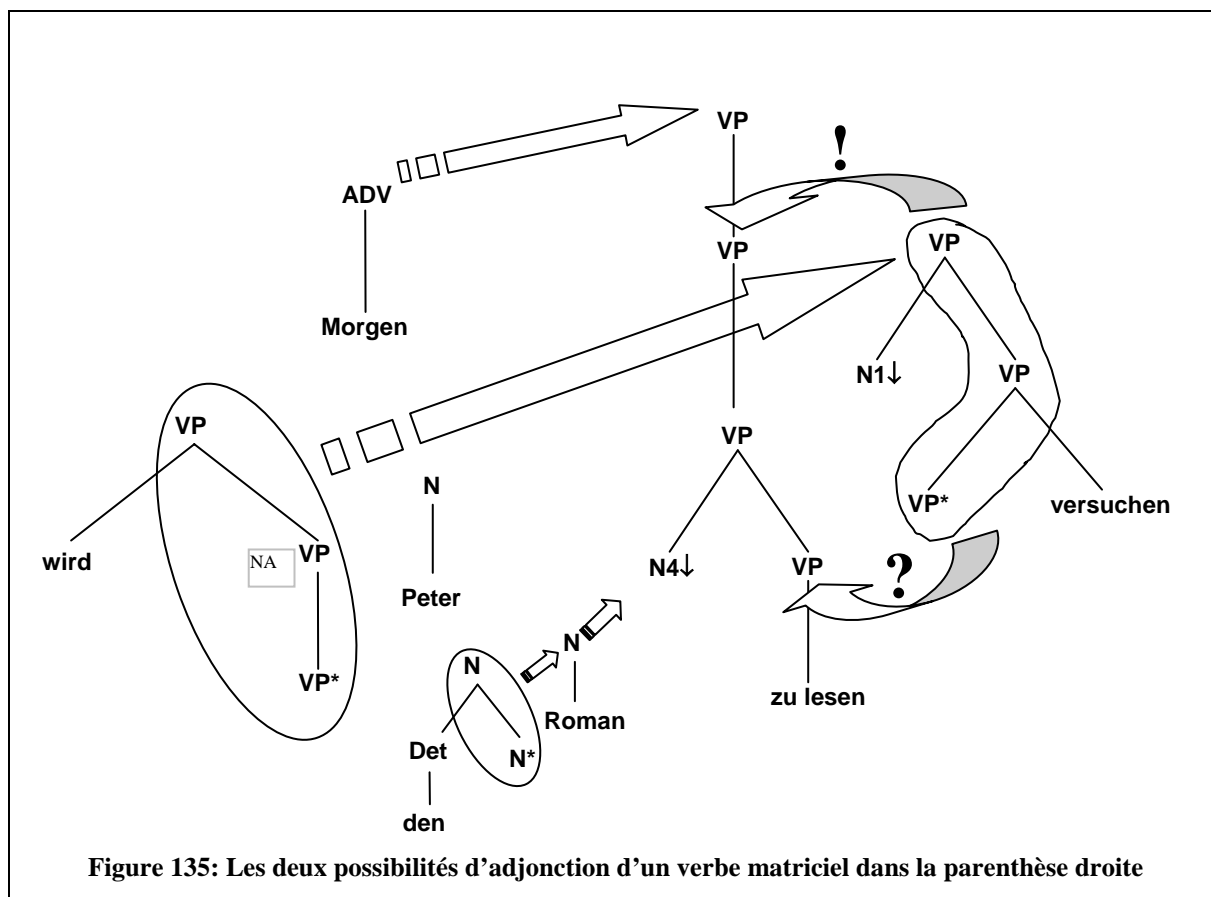
- (243) Morgen wird den Roman Peter zu lesen versuchen  
 Demain va le roman Pierre à lire essayer  
 'Demain, Pierre essayera de lire le roman'

Considérons maintenant l'adjonction dans un tel cas.

L'adjonction en bas (sur le nœud *VP* de la parenthèse droite) ne permet pas à l'auxiliaire d'avoir accès à son argument verbal, le verbe *versuchen* 'essayer'. Il est possible de réparer les dégâts en faisant passer l'information d'accord, par l'arbre de *zu lesen* 'de lire', tel que *wird* sous catégorise une infinitive nue et non une infinitive avec une préposition *zu*. Cependant, un passage des traits d'accord par le biais d'un autre arbre est toujours suivi du problème que nous avons rencontré lors de l'examen des relatives extraposées : l'arbre de dérivation ne représente plus la structure prédicat-argument.

Par contre, l'adjonction en haut, c'est-à-dire l'adjonction au nœud colonne de la parenthèse gauche, permet d'obtenir un bon arbre de dérivation. Ainsi, dans l'exemple donné ci-dessus, l'arbre de *wird* peut s'adjoindre sur son argument verbal *versuchen* 'essayer'.

Le critère de prédominance de la structure de dérivation implique que la priorité doit être donnée à cette possibilité, malgré le fait que la structure dérivée ne corresponde pas à une structure de constituants qui est souhaitable.



Ce constat nous a finalement amené à considérer que toute adjonction d'un arbre d'un verbe matriciel a toujours lieu au nœud colonne de la parenthèse gauche (ou à la racine si l'arbre sur lequel on effectue l'adjonction n'a pas de nœud colonne de la parenthèse gauche).

Cependant, la situation se complique car le fait que nous ayons à intervenir au niveau de la structure dérivée (qui s'approche en fait de la structure topologique), a comme résultat que le bidouillage que nous avons effectué pour sauver les relatives extraposées ne marche plus pour les éléments nominaux qui se trouvent dans le Mittelfeld. En effet, l'extrapolation des relatives a lieu dans le Nachfeld, c'est-à-dire après que tous les verbes dans la parenthèse droite. Si l'arbre matriciel s'adjoint au nœud colonne de la parenthèse gauche, il barre l'accès par la droite à des nœuds qui se trouvent plus pas que la parenthèse gauche, c'est-à-dire aux nœuds colonne du Mittelfeld. Il n'est donc pas possible d'avoir des analyses pour des phrases comme (244).

- (244) Morgen wird Peter den Roman zu lesen versuchen den ich sehr mag  
 Demain va Pierre le roman à lire promis que je beaucoup aime  
 'Demain, Pierre a promis de lire le roman que j'aime beaucoup'

Il me semble qu'il est pratiquement impossible d'étendre notre bidouillage à ces exemples, car il faudrait passer une liste des éléments nominaux modifiables à l'arbre matriciel, ce qui dépasse même les limites de bidouillage grâce aux traits. Nous voyons clairement que tout traitement d'un phénomène syntaxique qui ne suit pas les généralités observées mais qui est imposé par le formalisme coince à partir du moment que les phénomènes se répètent ou se compliquent. Par contre, dans le cadre d'une tâche concrète de génération, il est fort probable que le besoin de générer une phrase du type de (244) n'arrive jamais ; s'il fallait pour une raison ou pour une autre générer une telle phrase, le module de précalcul, qui est de toute manière présente pour les relatives, pourrait aussi la traiter à part.

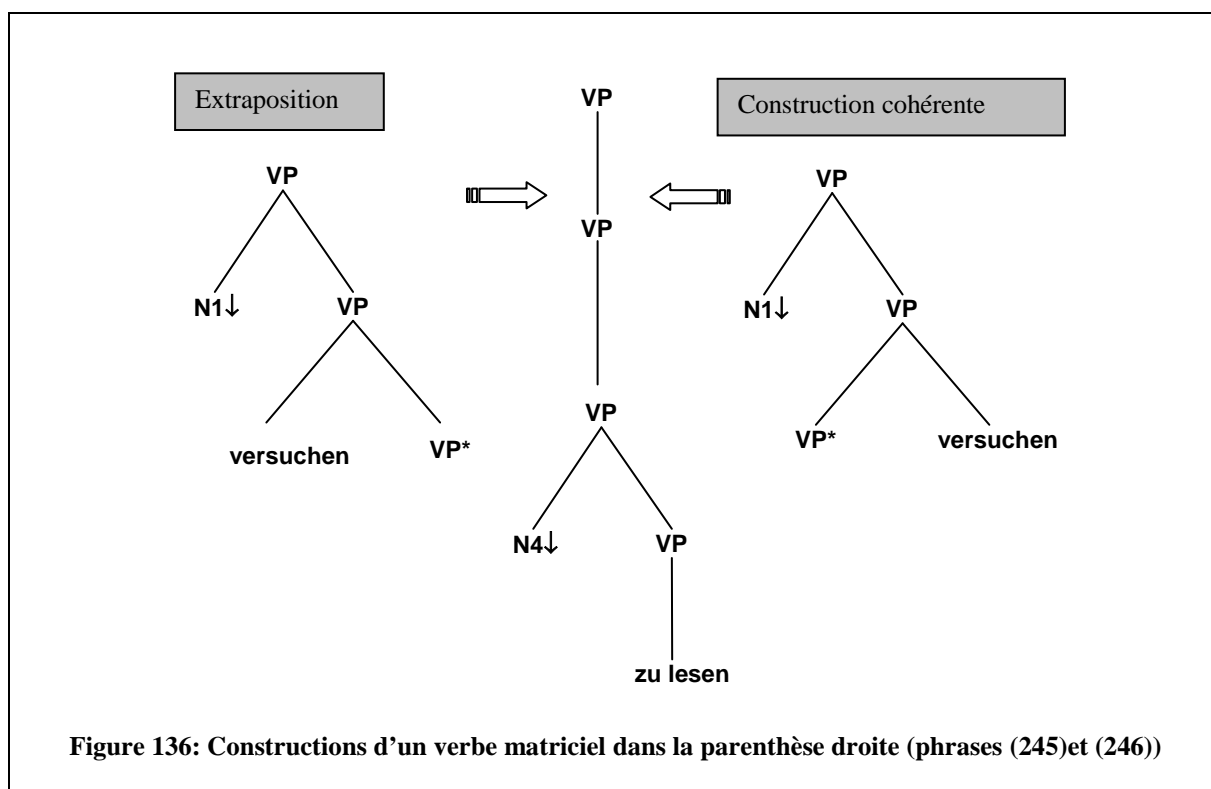
Ainsi, grâce à l'adjonction au nœud colonne de la parenthèse gauche, nous arrivons à construire aussi bien la structure cohérente (V2V1) que l'extrapolation (V1V2). La structure de l'arbre du

verbe matriciel qui permet l'extrapolation de son argument verbal, se distingue de la structure des constructions cohérentes seulement par le fait que la place de l'ancre verbale et celle du nœud pied soient échangées (Cf. Figure 136).

- (245) Morgen, wird Peter versuchen den Roman zu lesen  
 Demain va Pierre essayer le roman à lire  
 Demain, Pierre essayera de lire le roman'

La phrase (245) est une construction extrapolée, parallèle à la phrase (242) ; les deux phrases peuvent être analysées avec les arbres élémentaires proposés. Afin d'éviter l'adjonction d'une relative extrapolée dans le cas d'une structure cohérente (et éviter que le complexe verbal soit interrompu par une relative extrapolée tel qu'on peut le voir dans l'exemple (246)(246), nous avons recours à un trait qui empêche l'adjonction d'une relative extrapolée. Ainsi, une adjonction d'un verbe matriciel qui se fait par la droite porte le trait *rightAdjunction* avec une valeur positive sur son nœud pied, et l'arbre du verbe enchâssé fait passer sa valeur au trait *goAwayRel* qui, quand elle est positive, bloque l'adjonction d'une relative extrapolée.

- (246) \* Morgen wird Peter den Roman zu lesen, den ich sehr mag, versuchen  
 Demain va Pierre le roman à lire que je beaucoup aime promettre

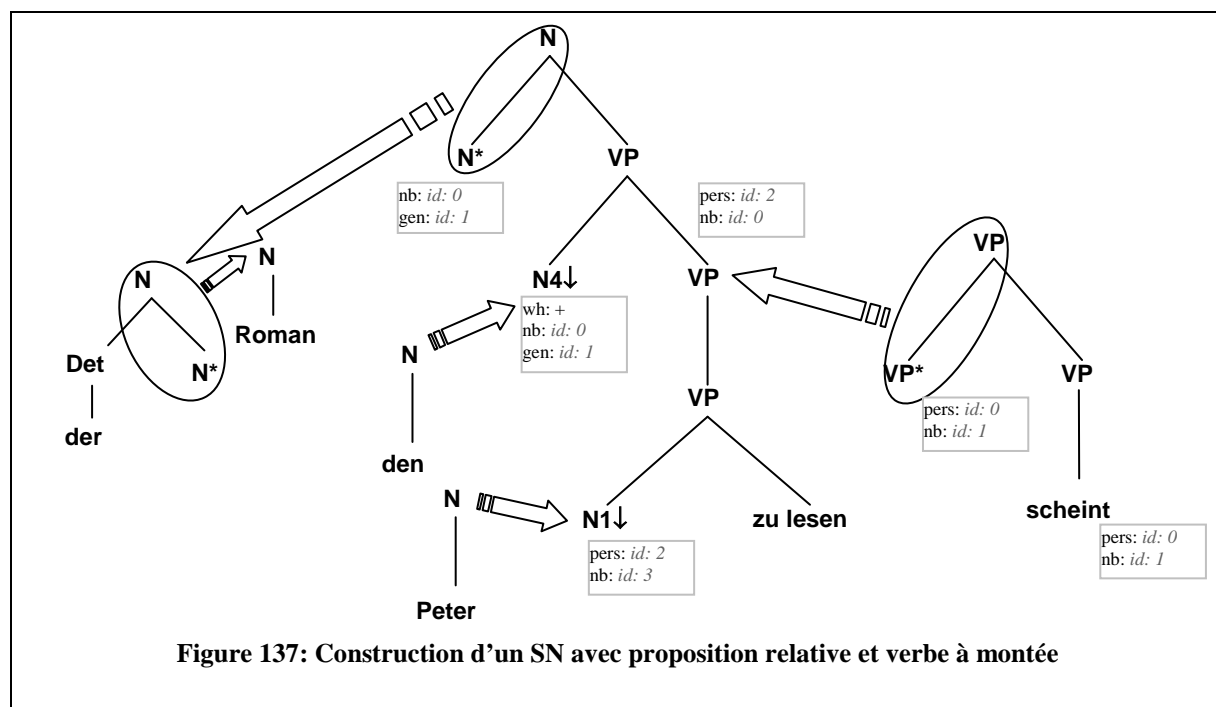
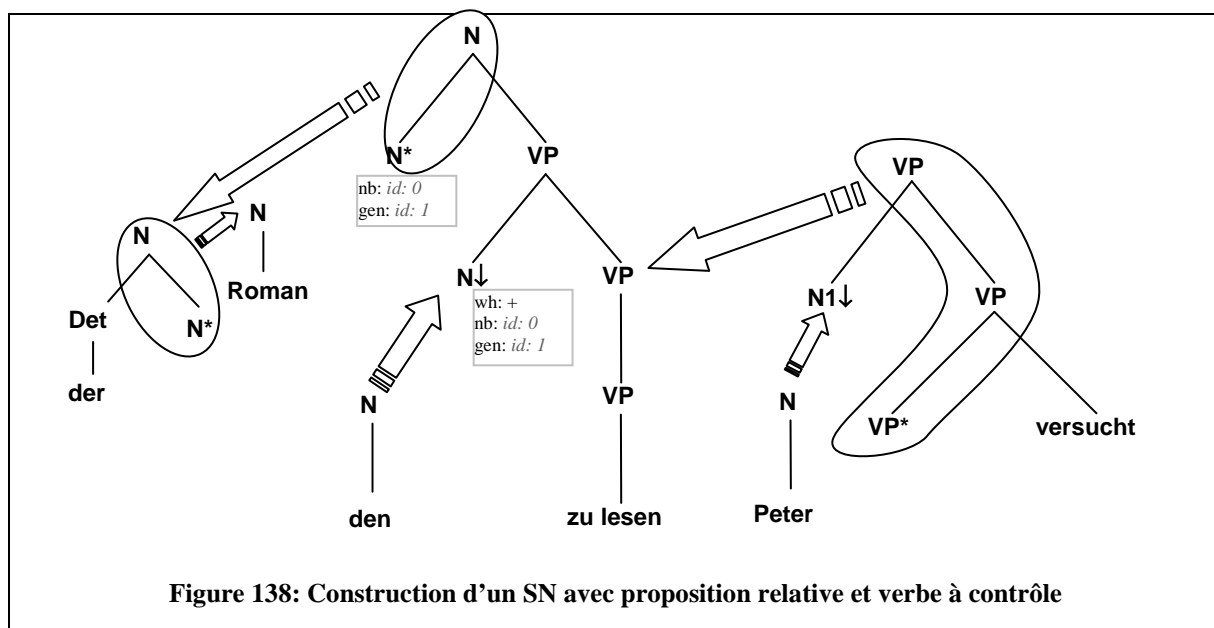


Nous voyons donc que mêmes des structures plutôt simples, parmi les structures que nous avons retenues pour une grammaire TAG de l'allemand (nous avons mis de côté des phénomènes comme le *fronting* et l'intraposition), les résultats ne sont pas tout à fait satisfaisants d'un point de vue théorique. Cependant, nous considérons que ces résultats sont suffisants pour la plupart des tâches de génération.

Nous allons mettre une fin à ce tour d'horizon des arbres élémentaires de DTAG, en revenant encore une fois sur la question des relatives, cette fois-ci, dans des constructions à argument verbal, que nous illustrons dans les exemples ci-dessous (247) et (248) :

- (247) der Roman, den Peter zu lesen versucht  
 le roman que Pierre à lire essaie  
 'le roman que Pierre essaie de lire'
- (248) der Roman, den Peter zu lesen scheint  
 le roman que Pierre à lire semble  
 'le roman que Pierre semble lire'

L'arbre de la Figure 138 permet d'analyser le SN qui apparaît sous (247) car, comme le sujet est pris en charge par le verbe à contrôle, il ne porte pas de nœud de substitution pour le sujet. Par contre, dans le cas d'une construction auxiliaire ou à montée, nous avons besoin d'introduire un autre arbre. Cet arbre, que nous montrons dans la Figure 137, possède un nœud de substitution pour le sujet et permet donc l'analyse du SN qui apparaît sous (248). L'accord est réalisé par le passage de traits, comme il est de coutume dans ce type de construction.



En fait, comme nous l'avons déjà montré dans la section 2.2.1, qui traite du problème du verbe à contrôle, l'analyse de la relativisation d'un argument nominal d'un verbe matriciel est une toute autre analyse que le traitement des relativisations et des structures matricielles habituelles. La raison est simple : tout comme la relativisation, l'argument verbal est réalisé à l'aide d'un nœud pied ; or un arbre ne peut avoir qu'un seul nœud pied. Il faut donc faire son choix : Faut-il réaliser l'adjonction de la relative, ou celle de l'argument verbal en tant que substitution ? La réponse est claire : la première adjonction est une adjonction modifiante (dans le sens de Schieber & Shabes 1994, voir section 2.2.2) qu'il faut à tout prix garder en tant que telle ; par contre, la deuxième adjonction est prédicative, ce qui veut dire qu'on utilise l'opération de substitution pour réaliser l'argument uniquement pour des raisons techniques. Le deuxième cas d'adjonction nous a permis de traiter des phénomènes comme les « dépendances de longue distance ». Cependant, comme la proposition relative constitue un îlot, ses éléments ne sortent pas de la proposition et il nous est possible de réaliser l'argument verbal à l'aide d'un nœud de substitution. Les exemples sous (249) et (250) illustrent ces deux cas. Nous présentons les arbres élémentaires qui permettent d'analyser ces deux phrases dans la Figure 139.

(249) der Mann, dem Peter zu schlafen scheint  
 le homme que Pierre à dormir semble  
 'l'homme à qui il semble que Pierre dort'

(250) der Mann, der zu schlafen versucht  
 le homme qui à dormir essaie  
 'l'homme qui essaie de dormir'

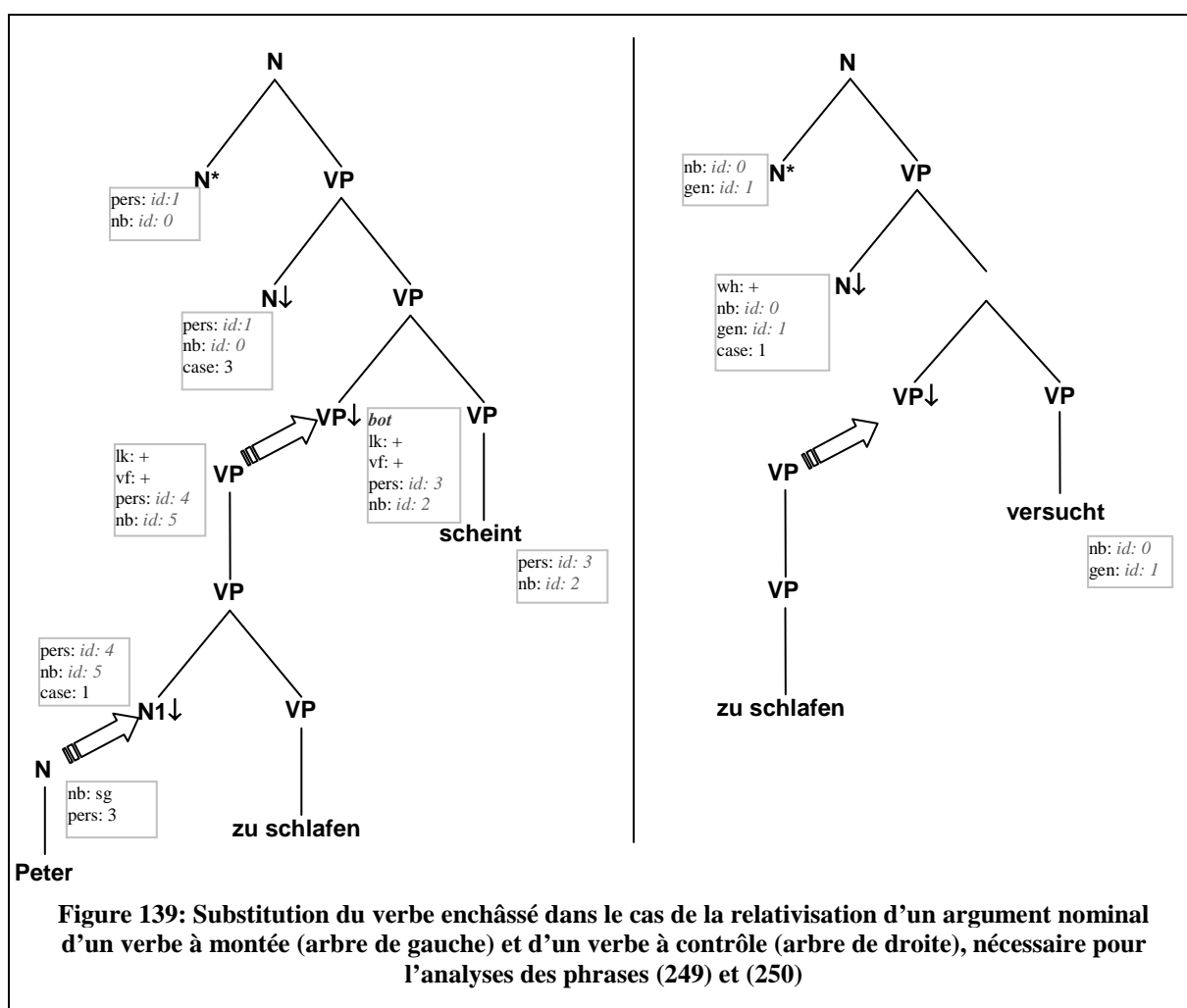


Figure 139: Substitution du verbe enchâssé dans le cas de la relativisation d'un argument nominal d'un verbe à montée (arbre de gauche) et d'un verbe à contrôle (arbre de droite), nécessaire pour l'analyses des phrases (249) et (250)

Notons toutefois que l'arbre de dérivation d'une telle analyse conserve les bons liens de prédicat-argument, même si l'ordre de représentation est inversé par rapport au cas d'adjonction des verbes matriciels : l'arbre qui fait une adjonction est marqué comme dépendant et l'arbre qui remplit son nœud de substitution est le gouverneur. Une discussion de la relation entre l'arbre de dérivation et le graphe sémantique se trouve dans Candito 1999:313-321 et Candito et Kahane 1998a.

#### **4.4 L'organisation de la DTAG dans une métagrammaire**

Dans cette sous-section, nous présentons les classes qui ont été utilisés dans la métagrammaire de l'allemand ainsi que leur organisation dans différentes hiérarchies. L'objectif de la structuration des schèmes retenus est de garder une modularité maximale permettant de compléter facilement la grammaire avec des phénomènes qui ont été laissés de côté. C'est la raison pour laquelle nous remettons à plus tard le traitement des structures à nombre premier, ou encore certains cas de sous-catégorisation comme l'attributif ou l'accusatif. Mais dès que la nécessité de traitement se présente dans le système de génération qui utilise cette grammaire, il sera très facile d'assembler la nouvelle classe à partir des classes existantes.

Par ailleurs, afin d'obtenir une représentation graphique des descriptions d'arbres, nous représentons la parenté directe ( $\rightarrow$ ) entre deux constantes par une ligne continue et la dominance ( $\Rightarrow$ ) par une ligne pointillée. La précedence linéaire est indiquée par le symbole '<', ce qui implique qu'en l'absence de ce symbole, l'ordre entre deux fils reste non spécifié. Dans un tel cas, l'ordre dans la représentation graphique est purement arbitraire, tout comme c'est le cas de la représentation des arbres de dépendance.

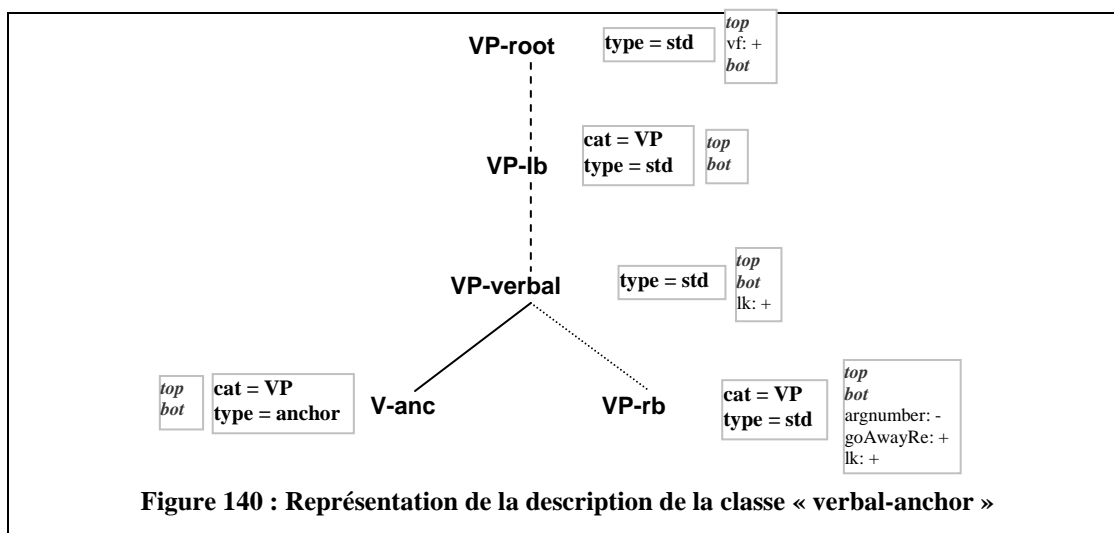
##### **4.4.1 La construction d'un schème verbale**

Pour comprendre la séparation de l'information contenue dans les arbres élémentaires, il convient de commencer présenter d'abord la hiérarchie des réalisations verbales qui constituent la charpente de la construction des schèmes verbaux. Toute classe de cette hiérarchie dépend de la classe « verbal-anchor », qui détermine déjà les liens de base entre les constantes qui apparaissent dans cette hiérarchie.

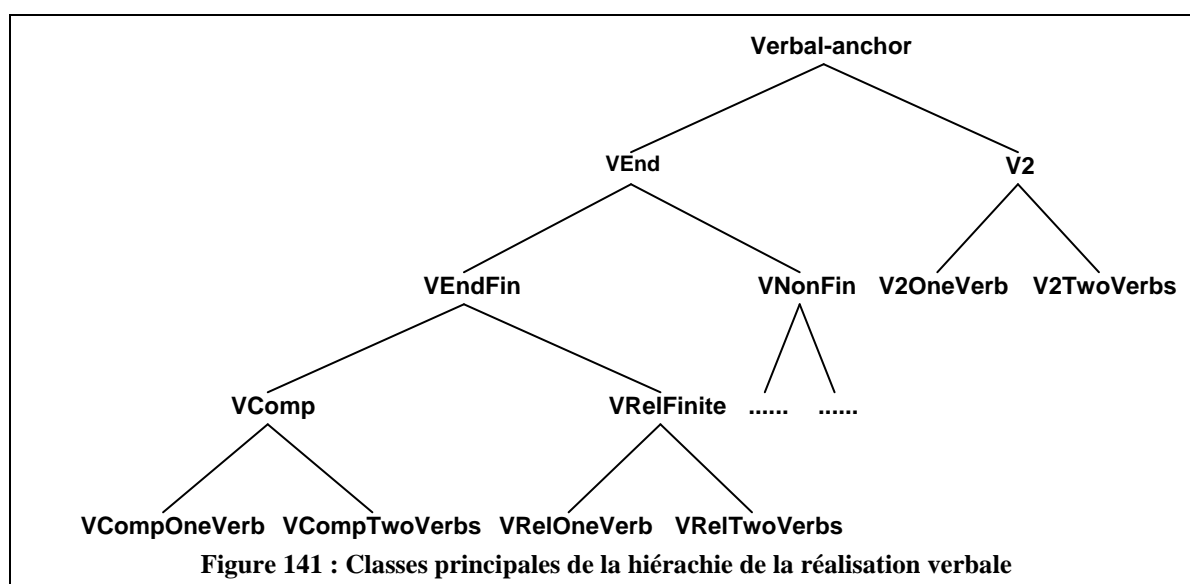
La constante racine s'appelle « VP-root » et domine toutes les autres constantes d'un schème verbal (par exemple, la constante « VP-lb », *lb* étant l'abréviation de l'anglais *left bracket*, parenthèse gauche). Nous rappelons que la dominance peut se réaliser par un chemin d'une longueur qui est égale ou supérieure à 0 ; autrement dit, l'égalité n'est pas exclue. Nous pouvons soulever une conséquence assez importante de cette dominance : pour que les deux constantes où l'une domine l'autre ne se réalise pas en un nœud dans le modèle minimal, il faut qu'une autre classe établisse la parenté ou bien que des valeurs différentes des traits interdisent l'unification.

Dans la Figure 140, nous représentons la description que comporte la classe « verbal-anchor », dont héritent tous les schèmes verbaux de DTAG.

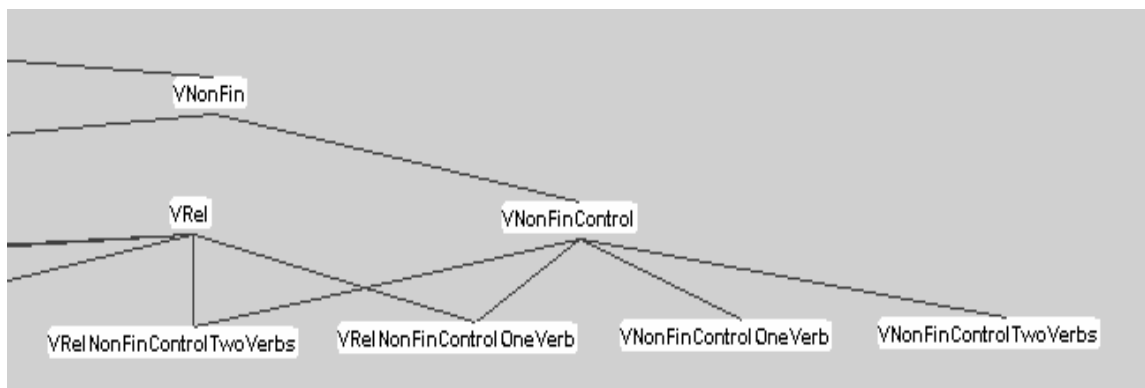




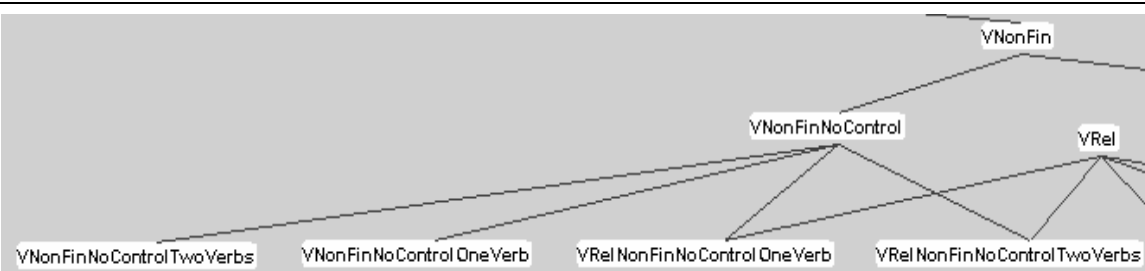
La constante « VP-verbal » qui est parent de l’ancre du schème décrit, se trouve dans la colonne future, entre la constante de la parenthèse gauche et celle de la parenthèse droite. Selon qu’il s’agisse d’une construction à verbe second ou à verbe final, cette constante va s’unifier avec l’une des deux constantes des parenthèses. Il s’agit là d’un point important car la hiérarchie complète de la réalisation verbale est divisée en fonction de ce choix : la constante « verbal anchor » a deux fils, d’une part la constante « V-end » où le verbe se trouve dans la parenthèse droite et d’autre part la constante « V2 » où le verbe se trouve dans la parenthèse gauche. La réalisation verbale pour les structures des verbes non finis est divisée dans une partie



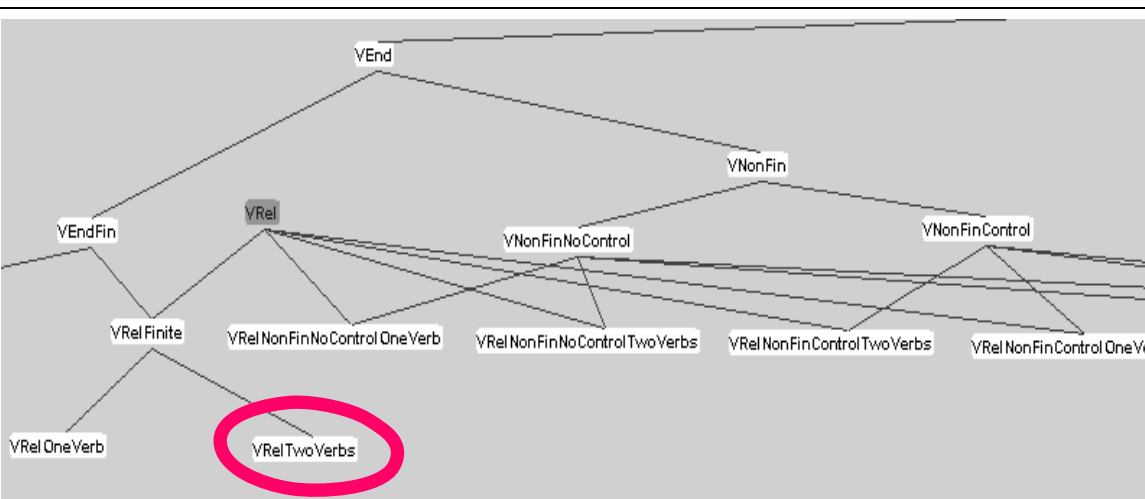
« VNonFinNoControl » où le verbe non fini réalise tous les arguments nécessaires dans le cas des constructions auxiliaires ou à montée, et dans une partie « VNonFinControl » où le sujet est supprimé dans les constructions à contrôle. Les figures suivantes montrent le détail de la hiérarchie de la réalisation verbale qui était restée non spécifiée dans la Figure 141.



**Figure 142 : Détail de la capture d'écran de l'éditeur de métagrammaire de Gaiffe 2002<sup>186</sup>. L'image montre la hiérarchie de réalisation verbale pour les verbes non finis utilisés dans des constructions à contrôle.**



**Figure 143 : Hiérarchie de réalisation verbale pour les verbes non finis utilisés dans des constructions auxiliaire et à montée.**



**Figure 144 : Hiérarchie de réalisation verbale autour des constructions pour les propositions relatives. La classe *VRelTwoVerbs* nous servira comme exemple dans la suite.**

Chaque classe terminale est redoublée en fonction de l'existence éventuelle d'une co-tête verbale. Dans la métagrammaire actuelle, de telles structures sont obtenues dans le traitement de différents passifs, où l'auxiliaire est réalisé en co-tête. En fait, les cas de co-tête doivent être

<sup>186</sup> Nous avons réimplémenté l'algorithme pour l'arrangement des graphes directionnels et nous avons ainsi changé l'apparence de l'interface.

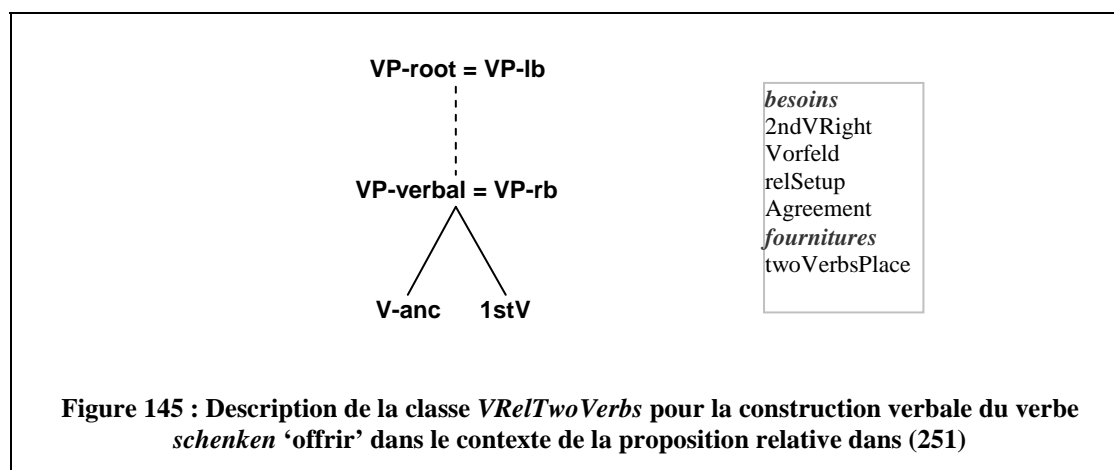
traités à part car l'existence d'une telle co-tête influence fortement la structure complète de l'arbre. Dans le cas du V2 par exemple, l'auxiliaire prend la parenthèse gauche et la tête verbale se place dans la parenthèse droite.

#### 4.4.2 Un exemple de neutralisation de traits polaires

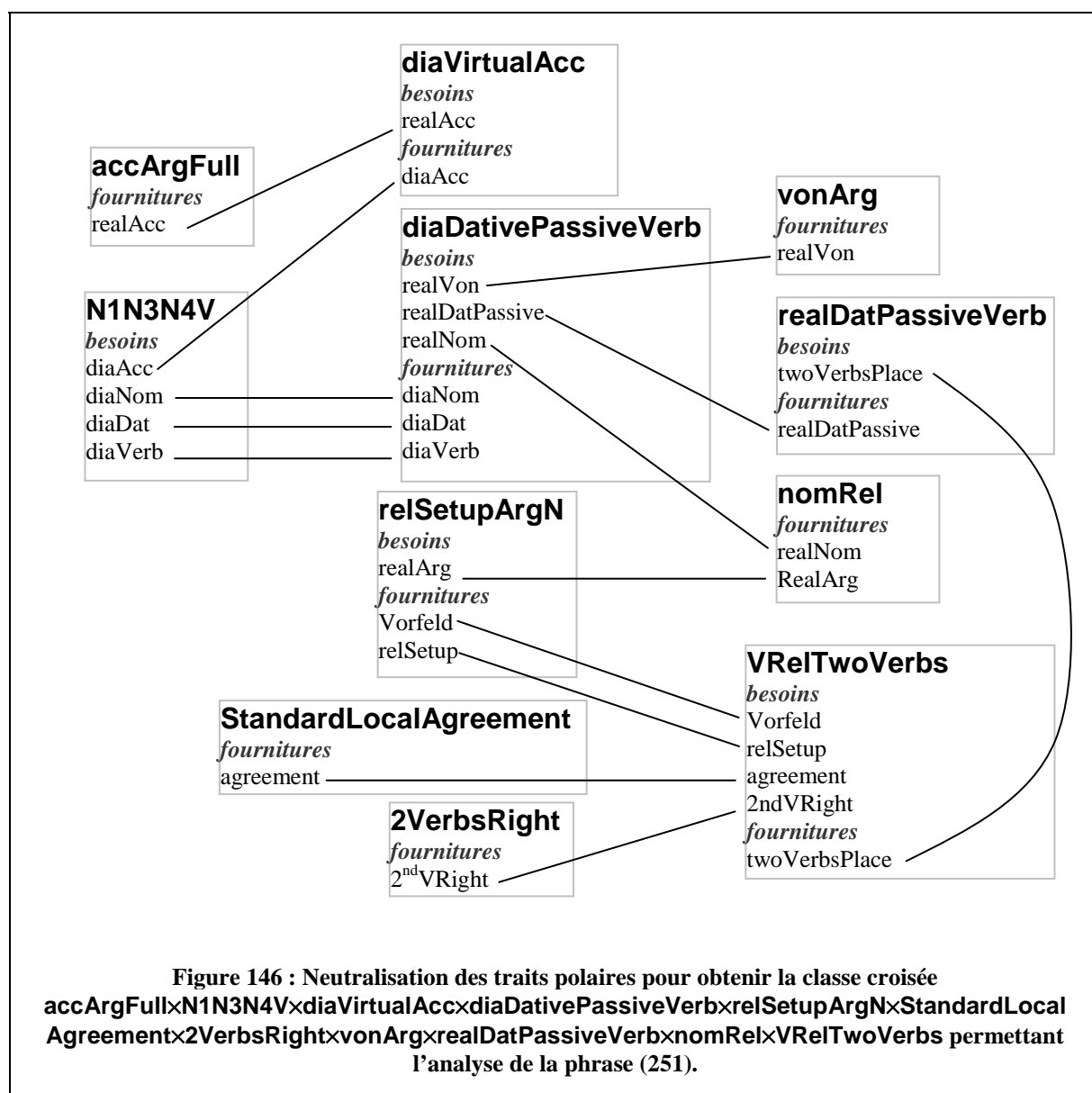
Examinons plus précisément le cas de la relative telle qu'elle apparaît dans l'exemple ci-dessous :

- (251) Der Mann, der von Maria einen Roman geschenkt bekam  
 Le home qui de Marie un roman offert recevait  
 'L'homme qui se voyait offert un roman par Marie'

La description de la structure de la réalisation verbale dont a besoin le verbe *schicken* 'offrir' est représentée dans la Figure 145. Notons que l'ordre entre les constantes « V-Anc » et « 1stV » n'est pas encore spécifié. Nous avons factorisé cette information qui est commune à toutes les structures ayant deux verbes dans la parenthèse droite dans une classe hors hiérarchie du nom « 2ndVRight ». Par ailleurs, si nous avions fait hériter chaque classe de cette classe pour la construction verbale à deux verbes, cela aurait créé un effet non souhaitable car la hiérarchie deviendrait illisible. Afin d'éviter cet effet, ces classes, tout comme la classe *VRelTwoVerbs*, ont besoin d'un trait polaire qui est fourni par la classe « 2ndVRight ». De cette manière, la description de la structure verbale se complète en se croisant avec la classe « 2ndVRight ».

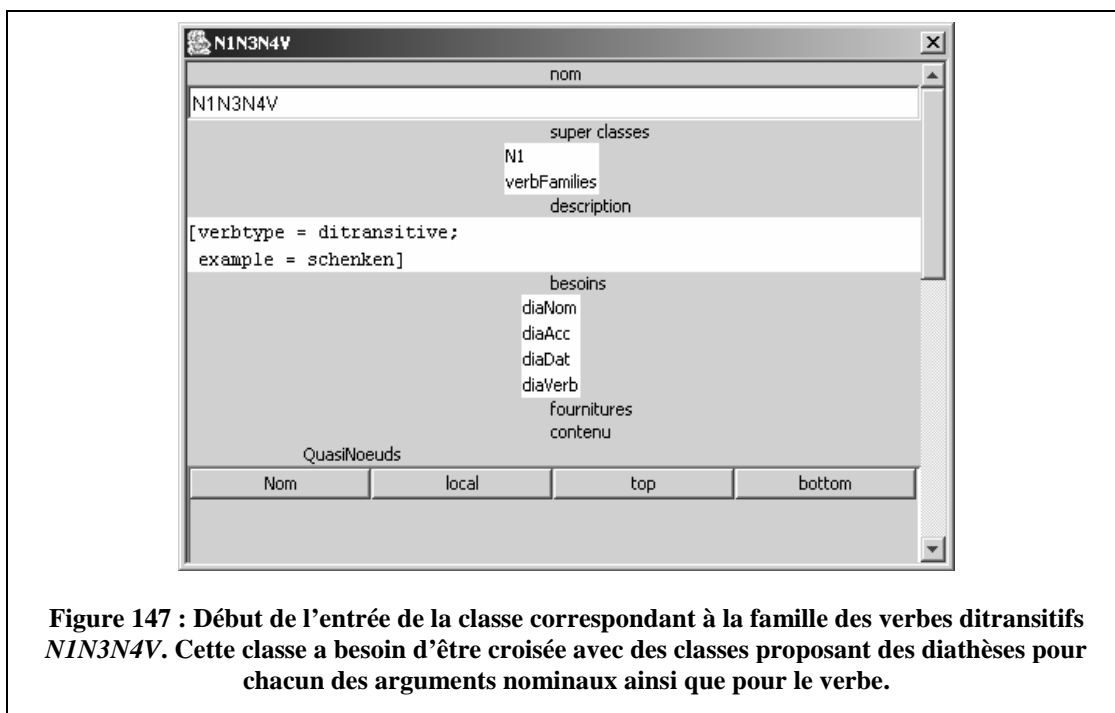


Notre but est de présenter la métagrammaire à l'aide de classes qui sont nécessaires dans l'obtention de l'analyse TAG du syntagme nominal (251). La Figure 146 montre comment se neutralisent les différents traits polaires des classes qui décrivent finalement l'arbre dont nous avons besoin.



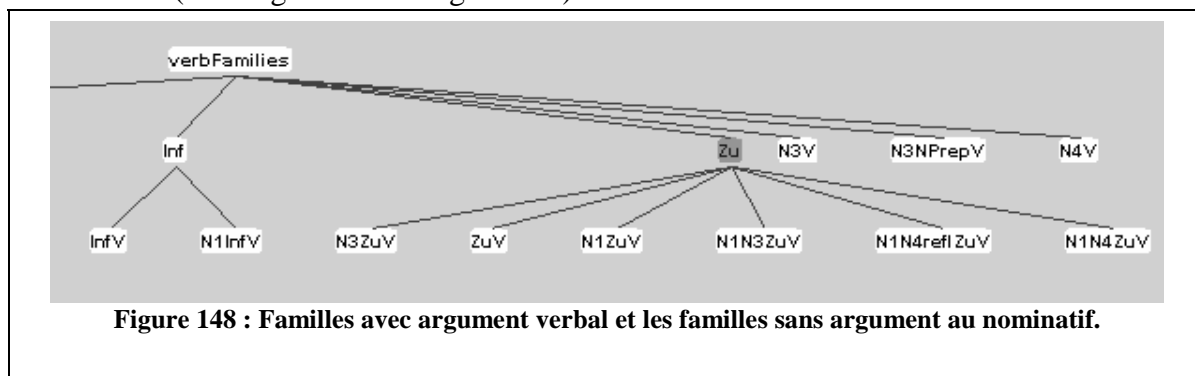
### 4.4.3 Familles

Passons maintenant aux familles où la sous-catégorisation initiale est déterminée. Nous avons une structure connexe à part qui correspond à la première dimension de Candito 1999. En ce qui concerne les noms des classes de la sous-catégorisation initiale, nous nous sommes reportés à la notation classique des familles ; ainsi pour distinguer les différents arguments, nous ajoutons le nom qui indique le cas de cet argument (plus le nom « Attr » pour indiquer lequel des deux arguments ayant le même cas est attributif).

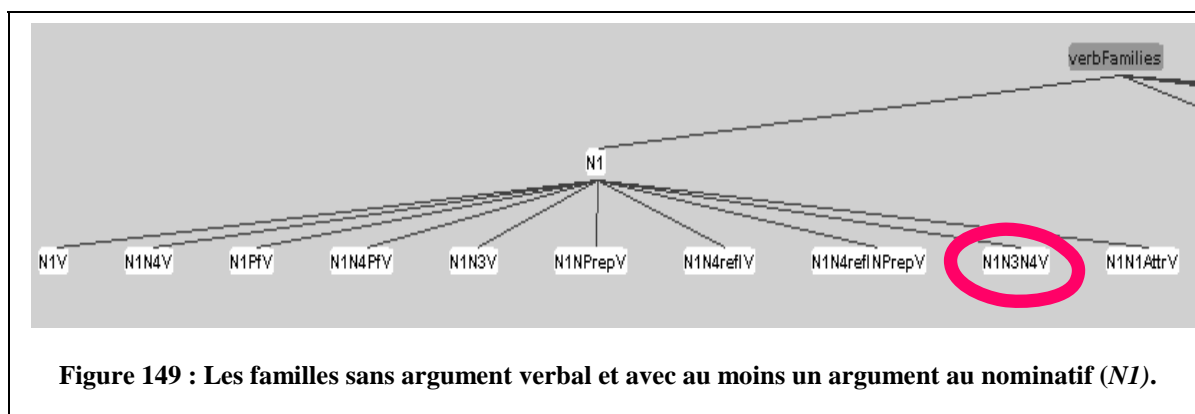


**Figure 147 : Début de l'entrée de la classe correspondant à la famille des verbes ditransitifs N1N3N4V. Cette classe a besoin d'être croisée avec des classes proposant des diathèses pour chacun des arguments nominaux ainsi que pour le verbe.**

Dans la mesure du possible, nous avons essayé de rester fidèles à l'approche de Candito 1999 : les classes peuvent donc être analysées comme des déclencheurs d'une série de croisements, parce que les classes de sous-catégorisation initiale ne fournissent rien et qu'ils ont besoin d'une série de traits polaires. Par exemple, la classe des verbes ditransitifs (*N1N3N4V*) a besoin d'une diathèse pour sa tête ainsi que chacun de ses arguments, c'est-à-dire le verbe (Cf. Figure 147). Dans le souci d'une lecture facile, nous avons regroupé les classes de la première dimension par type d'argument sous-catégorisé. De cette manière, nous avons regroupé les classes pour les verbes à argument verbal et ensuite toutes les classes pour les verbes avec au moins un argument au nominatif (voir Figure 148 et Figure 149).



**Figure 148 : Familles avec argument verbal et les familles sans argument au nominatif.**



**Figure 149 : Les familles sans argument verbal et avec au moins un argument au nominatif (N1).**

Il est difficile de partager l'information entre deux constantes qui ont des arguments différents, et c'est peut-être ce qui dévoile le plus clairement les lacunes de ce formalisme concernant le partage de descriptions d'arbres. En effet, une constante ne peut pas distribuer son information sur deux autres constantes qui sont réalisés comme deux nœuds différents ; autrement dit, on ne peut pas mentionner de manière abstraite une constante et appliquer ses propriétés à d'autres constantes. Par exemple, on ne peut pas exprimer qu'un argument au datif et à l'accusatif se comporte exactement de la même manière ; on est obligé de répéter chaque fois la même information pour chacune des constantes. C'est ainsi qu'on voit une hiérarchisation plus poussée des classes représentant des familles n'amène pas à un plus grand partage d'information<sup>187</sup>.

Dans le tableau suivant, nous illustrons par un exemple chaque famille traitée.

<i>Famille</i>	<i>type</i>	<i>Exemple</i>
N1V	intransitif	<i>erzittern</i> 'trembler'
N1N4V	transitif	<i>besprenkeln</i> 'moucheter'
N1PfV	intransitif avec préfix	<i>einschlafen</i> 's'endormir'
N1N4PfV	transitif avec préfix	<i>aufschreiben</i> 'noter'
N1N3V	avec nominatif et datif	<i>helfen</i> 'aider'
N1NPrepV	intransitif avec argument prép.	<i>denkenAn</i> 'penser à'
N1N4reflV	réfléchi	<i>sichWaschen</i> 'se laver'
N1N4reflNPrepV	réfléchi avec argument prép.	<i>sichErinnern</i> 'se rappeler de'
N1N3N4V	ditransitif	<i>schenken</i> 'offrir'
N1N1AttrV	attributif	<i>werden</i> 'devenir', <i>bleiben</i> 'rester'
InfV	auxiliaire avec inf. nu ou participe	<i>haben</i> 'avoir', <i>werden</i> 'aux futur', <i>bleiben</i> 'rester'
N1InfV	contrôle avec inf. nu	<i>wollen</i> 'vouloir', <i>koennen</i> , 'pouvoir', <i>müssen, sollen</i> 'devoir'
N3ZuV	montée avec datif	<i>scheinen</i> 'sembler à'
N1ZuV	contrôle avec zu inf	<i>Versuchen</i> 'essayer', <i>meinen</i> 'penser'
N1N3ZuV	contrôle avec zu inf et datif	<i>erlauben</i> 'permettre', <i>verbieten</i> 'interdire', <i>versprechen</i> 'promettre'
N1N4ZuV	contrôle d'accusatif <sup>188</sup> (ou de nominatif) avec zu inf	<i>bitten</i> 'prier', <i>verpflichten</i> 'obliger'
N1N4reflZuV	réfléchi et contrôle avec zu inf	<i>sich entschliessen</i> 'se décider', <i>sich verbitten</i> 'ne pas admettre'
N3V	avec datif	<i>grauen</i> 'avoir horreur'
N3NprepV	avec datif et argument prép.	<i>liegen an</i> 'tenir à'
N4V	avec accusatif	<i>frieren</i> 'avoir froid'

<sup>187</sup> Contrairement à Candito 1999, nous avons mis à part les classes de réalisation verbale, ce qui représente une information partagée par les différentes familles.

<sup>188</sup> Le mot *bitten* 'prier' peut contrôler l'accusatif comme dans (i) (l'enfant dort), mais aussi le nominatif comme dans (ii) (l'homme a permission de dormir)

- (i) Der Mann bittet das Kind zu schlafen.  
L'homme prie l'enfant de dormir
- (ii) Der Mann bittet das Kind schlafen zu dürfen.  
L'homme prie l'enfant de lui permettre de dormir.

Pour l'analyse TAG on peut regrouper les deux cas dans la même famille, car le lien qu'entretient le verbe enchâssé dans une construction à contrôle n'est pas exprimé de toute manière. Ces constructions constituent ainsi un cas problématique pour la reconstruction du lien manquant dans la structure sémantique à partir de l'arbre de dérivation.

#### 4.4.4 Diathèses

Chaque classe qui représente une famille exige une diathèse par argument ainsi qu'une diathèse pour la tête verbale. Cette étape correspond à la dimension 2 de Candito 1999.

En ce qui concerne le choix de la forme active du verbe, nous avons utilisé le terme de *diathèse virtuelle*, introduit par Tesnière 1957:243. La plupart des classes de diathèse sont responsables de la création de la sous-catégorisation finale pour un seul argument. La raison qui explique ce choix est qu'on veut partager le maximum d'information. C'est ce qui explique pourquoi par exemple, le passif de l'accusatif d'un verbe reste toujours le même qu'il s'applique à un verbe intransitif, transitif, ou ditransitif; de la même manière, la diathèse d'un argument prépositionnel est toujours virtuelle car elle n'est pas affectée par le passif ni de l'accusatif, ni du datif. D'autre part, il faut faire en sorte que la redistribution d'un argument soit accompagnée de la redistribution des autres arguments qui sont concernés par le changement de diathèse en question. Ainsi, il ne faut pas que la sous-catégorisation initiale du nominatif subisse une diathèse virtuelle et reste nominatif pendant que l'accusatif est transformé en nominatif à cause du changement de diathèse du passif. Pour éviter de telles erreurs, nous disposons de deux traits binaires, l'un pour la diathèse du passif de l'accusatif et l'autre pour la diathèse du passif du datif. La Figure 150 montre les différentes classes de diathèse.

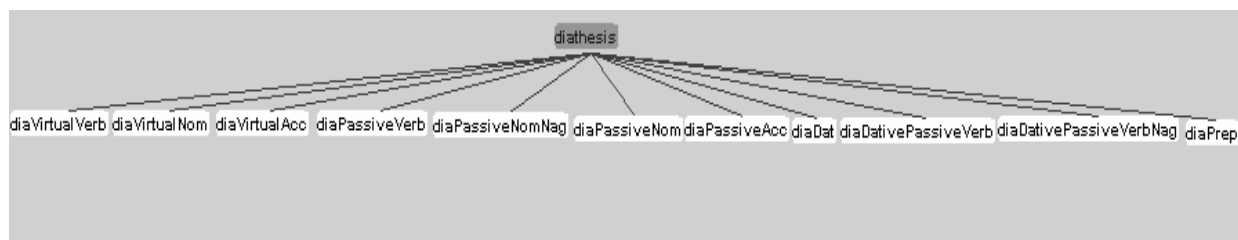


Figure 150 : La diathèse de la métagrammaire

La Figure 151 illustre les classes de la diathèse du passif du verbe (qui ne fait rien que de fournir une diathèse du verbe et d'exiger une réalisation verbale du passif (*realVerbPassive*)).

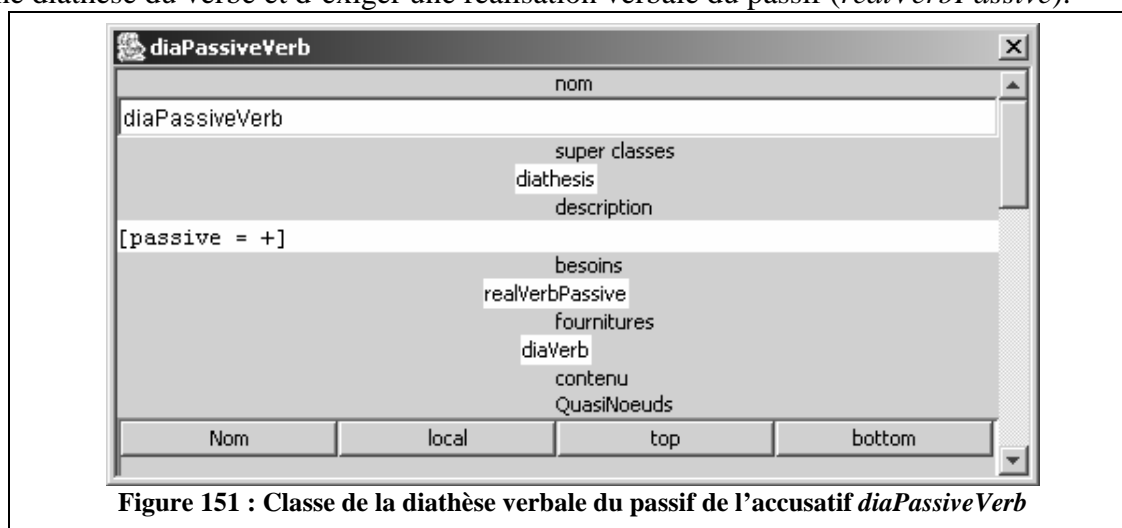


Figure 151 : Classe de la diathèse verbale du passif de l'accusatif *diaPassiveVerb*

Enfin la Figure 152 montre la diathèse du passif pour le nominatif avec réalisation de l'agent. La classe fournit une diathèse du nominatif et elle a besoin d'une réalisation d'un syntagme prépositionnel construit avec la préposition *von* 'par'. De plus, elle indique que l'argument du syntagme prépositionnel portera la catégorie *NI*, ce qui nous permet de reconnaître la sous-catégorisation initiale dans l'arbre élémentaire.

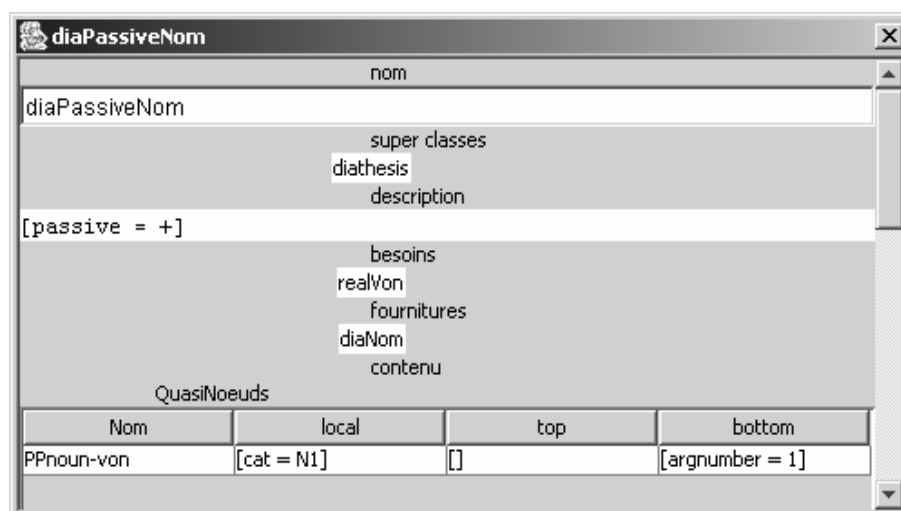


Figure 152 : Classe de la diathèse du nominatif au passif *diaPassiveNom*.

Par contre, la diathèse du passif du datif contient la diathèse du verbe, du nominatif et du datif car elle ne peut s'appliquer que dans le nominatif et le datif sont présents dans la sous-catégorisation initiale. La Figure 153 montre l'information qui est contenue dans la classe de diathèse *diaDativePassiveVerb*.

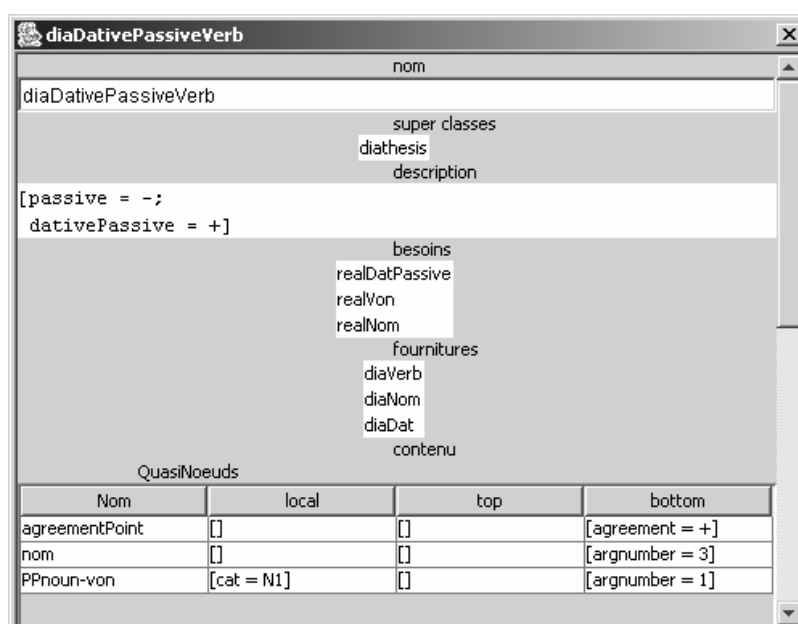


Figure 153 : Classe de la diathèse du passif du datif *diaDativePassiveVerb*.

La classe *diaDativePassiveVerb* est compatible avec la diathèse virtuelle de l'accusatif (Cf. Figure 154) parce que ce cas n'est pas concerné par la diathèse du passif du datif. Autrement dit, un argument à l'accusatif conserve son cas dans une construction du passif du datif.



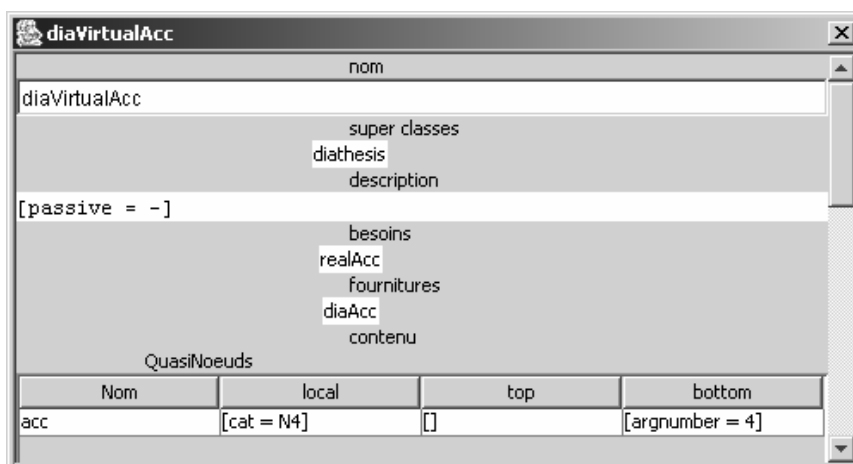


Figure 154 : Classe de la diathèse virtuelle (à l'actif) de l'accusatif qui sert aussi pour l'accusatif dans le contexte du passif du datif.

#### 4.4.5 Réalisations

Revenons à l'exemple concernant la description des classes croisées permettant d'obtenir l'arbre *schicken* 'offrir' que nous reportons sous (252).

- (252) Der Mann, der von Maria einen Roman geschenkt bekam  
 Le home qui de Marie un roman offert recevait  
 'L'homme qui se voyait offert un roman par Marie'

Dans cette phrase, la classe *NIN3N4V* a besoin de quatre diathèses : celui du verbe (*diaVerb*), du nom (*diaNom*), du datif (*diaDat*) et de l'accusatif (*diaAcc*). En fait, la classe *diaDativeVerb* fournit à lui seul les trois premières diathèses. Et pour réaliser l'accusatif, nous pouvons tout simplement recourir à la classe de la diathèse virtuelle de l'accusatif, car l'accusatif n'est pas impliqué dans la diathèse du passif du datif. Comme nous l'avons vu dans la Figure 154, la diathèse active de l'accusatif où la valeur négative pour le trait *passif* est compatible avec les traits de la diathèse du passif du datif.

Le croisement suivant *NIN3N4V* × *diaDativeVerb* × *diaVirtualAcc* a donc les besoins de réalisation suivants : *realDatPassive*, *realVon*, *realNom* et *realAcc*.

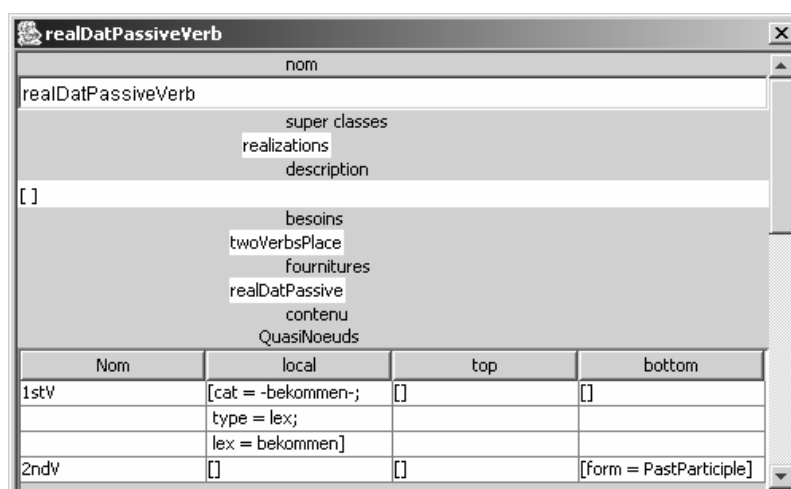
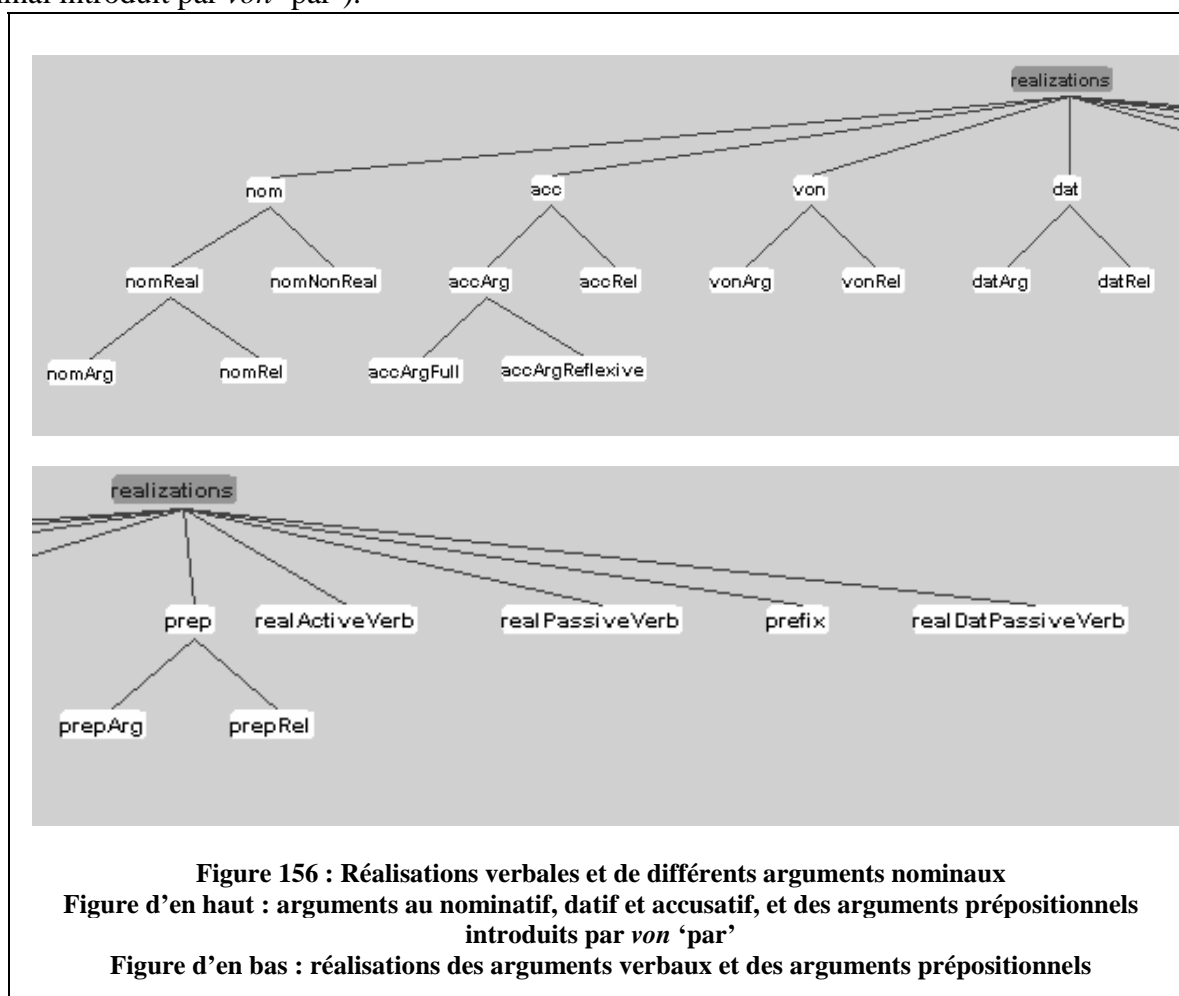
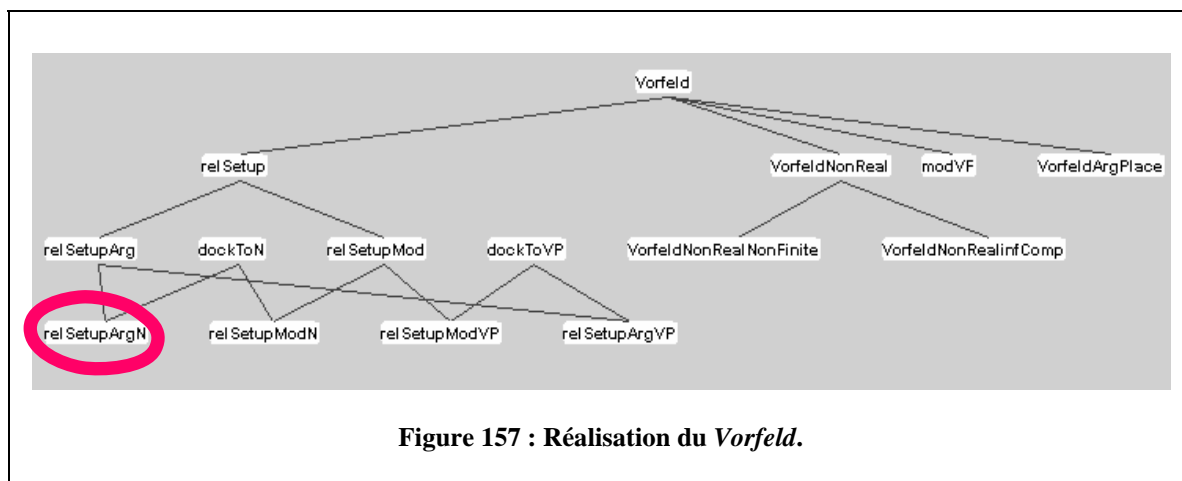


Figure 155 : Classe de la réalisation de la structure verbale du passif du datif.

La classe *realDatPassiveVerb* (Cf. Figure 155) fournit le trait *realDativePassive* et introduit dans l'arbre, l'auxiliaire *bekommen* 'recevoir'. Pour chaque réalisation nominale, nous avons le choix entre une réalisation d'argument nominal et celui d'un argument relativisé (cf. figureRéalisation). Notons que la double relativisation est exclue par le trait polaire *relArg* que fournit chaque classe d'argument relativisé et qui est consommé par la classe de l'arrangement de la relative. En ce qui concerne le cas nominatif, il peut ne pas être réalisé (notamment dans les constructions à contrôle). Dans le cas de notre exemple, nous nous intéressons à la relativisation du nominatif et à la réalisation des arguments à l'accusatif ou introduits par *von*. Nous utilisons donc les classes *nomRel* (relativisation du nominatif), *accArgFull* (réalisation d'un argument au nominatif « plein » par rapport à un accusatif réfléchi) et *vonArg* (réalisation d'un argument nominal introduit par *von* 'par').

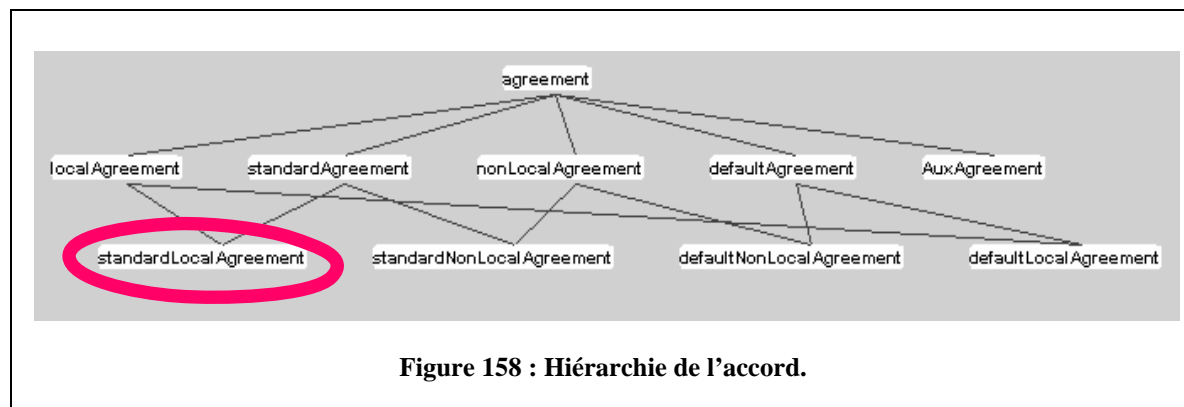


Passons maintenant à la classe *realDatPassivVerb*. Cette classe a besoin d'une structure verbale à deux verbes, un besoin qui est rempli par exemple par la classe *VRelTwoVerbs* que nous avons présenté au début de la section. En allemand, toute réalisation verbale dans la métagrammaire a besoin d'une construction de *Vorfeld*, qui peut être réalisé ou non ; dans le cas où il est réalisé, le *Vorfeld* peut être occupé par un modifieur, ou un argument qui peut être ou non relativisé. La hiérarchie complète de la réalisation du *Vorfeld* est montrée dans la Figure 157.

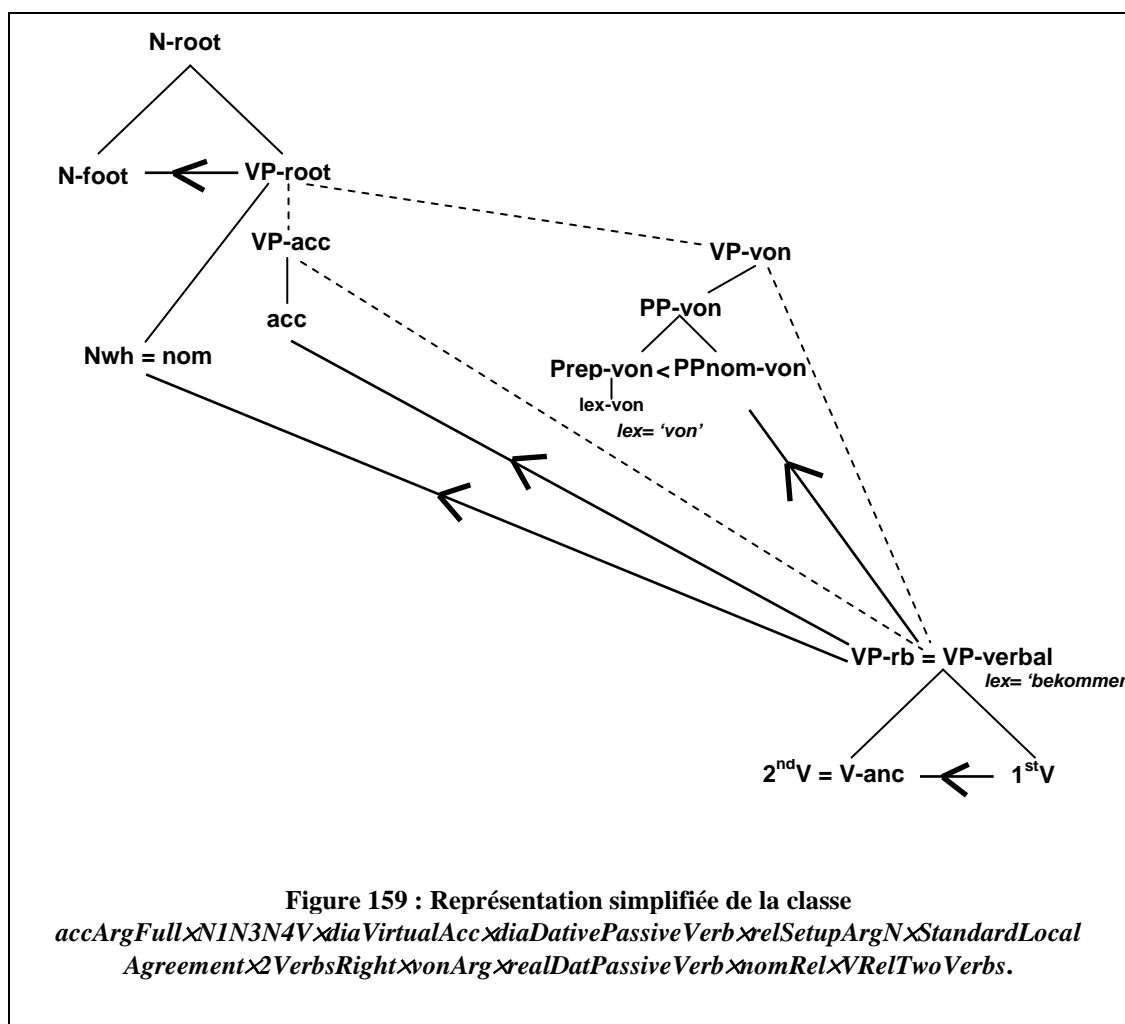


Les besoins de la classe *VRelTwoVerbs* d'un *Vorfeld* et d'un *relSetup* sont remplis par la même classe *relSetupArgN* : il s'agit d'une classe qui, dans le cadre de la modification d'un nom, apporte une nouvelle racine N et un nœud pied de la même catégorie<sup>189</sup>. Afin d'assurer la coïncidence entre la classe *relSetupArgN* et la relativisation d'un argument, cette classe demande le trait *relArg* qui, dans le cas de notre exemple, est fourni par le trait *nomRel* (la relativisation du nominatif).

Passons maintenant aux classes d'accord. Dans le formalisme de la métagrammaire où l'implication ne peut pas être exprimé, il est en fait impossible de réaliser directement le fait linguistique suivant, qui est pourtant très simple : une verbe fait l'accord avec son argument au nominatif s'il est présent, sinon il fait l'accord à la troisième personne du singulier par défaut. Pour simuler une telle implication, nous avons besoin d'un jeu de traits permettant le bon accord, même si, par exemple, le nominatif de la sous-catégorisation disparaît sous une diathèse du passif. De plus, dans le cas de l'accord « de longue distance » induit par l'analyse des constructions avec auxiliaire ou à montée, il faut faire en sorte que la bonne information d'accord passe à l'arbre auxiliaire (cf. Figure 158).



<sup>189</sup> Dans la métagrammaire, il existe une classe parallèle pour la relativisation extraposée, qui s'adjoit à la colonne VP. Voir 2.4 et Figure 157.



Les différentes constantes de la colonne future porte des traits qui interdisent l'unification de ces constantes dans un seul nœud. La seule contrainte non spécifiée est donc l'ordre de dominance dans la colonne des constantes *VP-acc* et *VP-von*, ce qui correspond aux deux possibilités d'ordres des constituants dans le Mittelfeld.

Nous obtenons ainsi les deux schèmes des Figure 160 et Figure 161 : le premier sert pour l'analyse de la phrase car il permet de remplacer l'ancre *\_V\_* par le verbe *geschenkt* 'offert'.

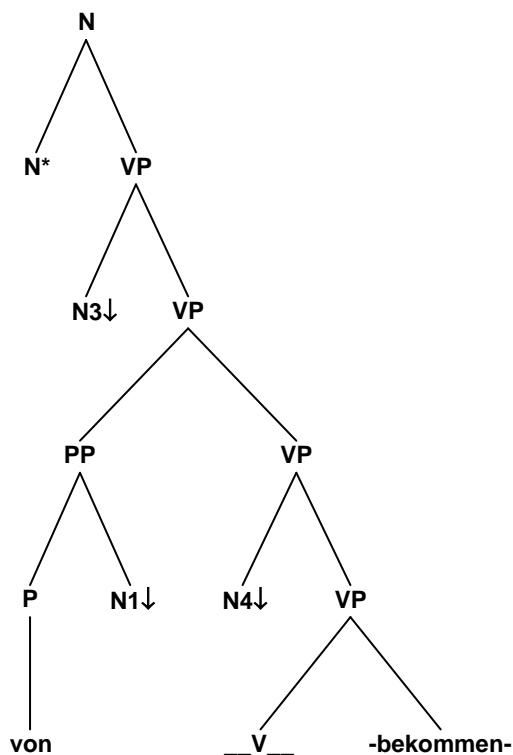


Figure 160: Premier modèle minimal de la classe croisée

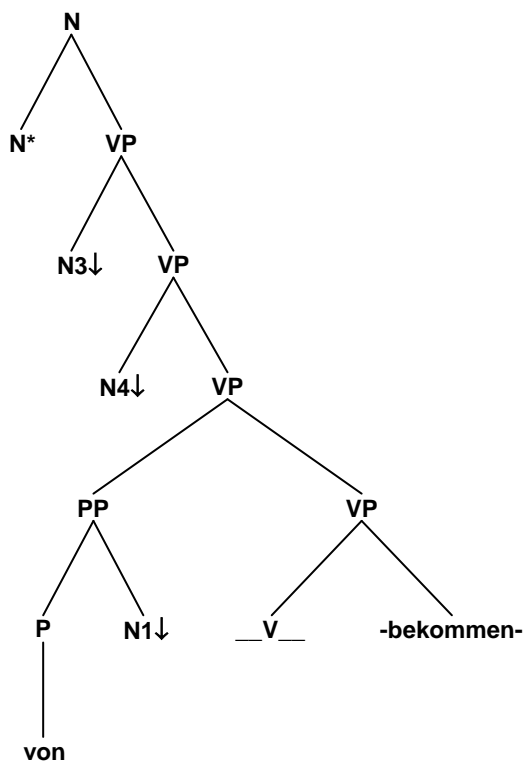
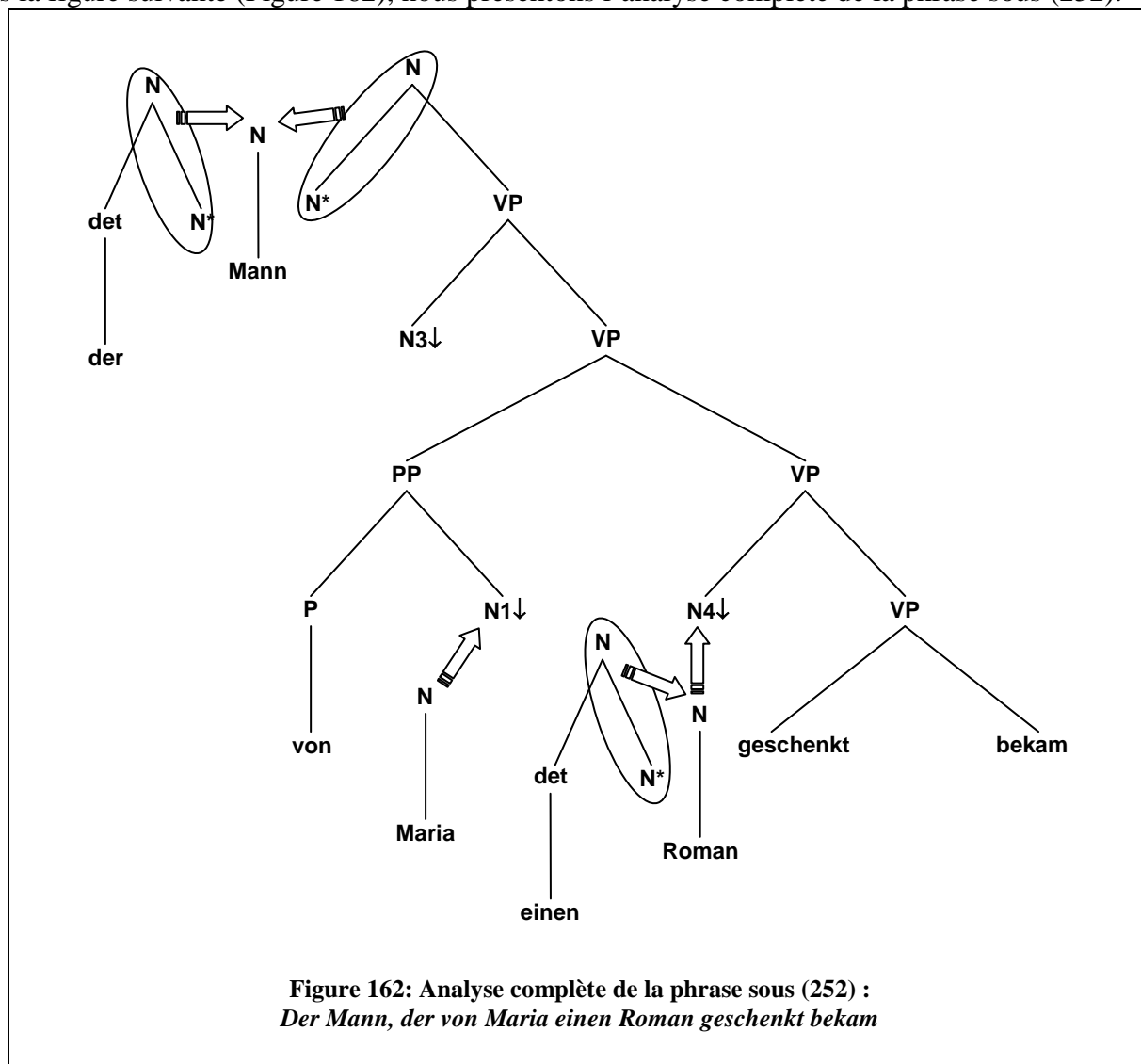


Figure 161: Deuxième modèle minimal de la classe croisée

Dans la figure suivante (Figure 162), nous présentons l'analyse complète de la phrase sous (252).



## 5 Remarques conclusives sur DTAG

Nous avons présenté cet exemple afin d'illustrer la métagrammaire de l'allemand que nous avons construite. Il est certes clair que cette grammaire n'est pas à un stade qui lui permettrait de revendiquer qu'elle traite tous les phénomènes au complet. Néanmoins, elle présente plusieurs avantages. Son plus grand avantage se trouve dans sa modularité conséquente qui lui permet de s'agrandir très facilement à des structures si cela est exigé par les structures des phrases spécifiques que nous devons générer. Par conséquent, comme la puissance de cette grammaire repose sur sa grande adaptabilité aux tâches requises durant la génération, la question de savoir combien d'arbres sont générés par cette grammaire est d'une importance mineure.

Dans son état actuel, cette grammaire de l'allemand génère déjà plusieurs milliers d'arbres, mais il sera très facile de l'augmenter de quelques centaines si l'on traite d'autres phénomènes linguistiques, comme par exemple les arguments nominaux au génitif que nous avons choisi de laisser de côté pour l'instant à cause de sa rareté d'apparition dans la langue réelle.

Il est notable de souligner que cette grammaire TAG de l'allemand peut traiter les principaux phénomènes syntaxiques de l'allemand, en structurant les schèmes à l'aide d'une métagrammaire. L'utilité d'une telle grammaire semble donc évidente dans le cadre d'un

développement de systèmes de génération de textes. Ceci est vrai au moins pour les systèmes qui marchent déjà grâce à TAG et qui tournent pour d'autres langues mais aussi les systèmes qui n'ont pas besoin de structures dérivées de sortie, spécifiques pour une étape de traitement postérieure, comme par exemple la synthèse de la parole. Cette dernière restriction est causée par le peu de liberté que nous avons dans la détermination de la structure dérivée ; je renvoie à la discussion présentée dans la section 6.

Toutefois, malgré les aspects prometteurs de la construction de la DTAG, nous nous trouvons confrontés à des lacunes du formalisme : d'un point de vue théorique, il se trouve que ce dernier est trop restreint pour qu'il puisse donner des analyses qui sont satisfaisantes à la fois d'un point de vue sémantique et syntaxique. Notre tentative sur une grande échelle a permis en quelque sorte d'affirmer la nécessité de revoir ce formalisme, voire à rechercher un autre formalisme permettant une plus grande articulation entre syntaxe et sémantique.

Un autre point faible de cette grammaire porte sur les critères permettant de construire un formalisme : la seule existence d'un formalisme ne peut pas servir de justification à une structure, et il est impératif de disposer de critères purement linguistiques dans la détermination des structures syntaxiques.

A la suite de ce chapitre, nous allons donc nous diriger vers un formalisme alternatif. Nous allons proposer une grammaire d'arbres lexicalisés basée sur le modèle topologique que nous avons présenté et développé dans le deuxième chapitre. Cette grammaire, que l'on intègre comme un module à part dans le modèle Sens-Texte, tente de surmonter les lacunes que nous avons rencontré lors de la création de la DTAG, tout en préservant les avantages et la beauté d'une grammaire d'arbres lexicalisés.





## Chapitre IV

# **Formalisations et extensions de la topologie**

# Formalisations et extensions de la topologie

*Anstatt mit objektiven Eigenschaften ... finden wir uns mit den von uns selbst gebauten Apparaten konfrontiert, und anstatt der Natur oder dem Universum zu begegnen, begegnen wir gewissermaßen immer nur uns selbst<sup>190</sup>.*

## 1 Formalisation d'une analyse topologique

Cette section reprend les notions présentées de manière descriptive dans le chapitre II. Dans un premier temps, nous formalisons la forme et l'expressivité des règles que nous avons considérées pour la linéarisation topologique. Dans la section 1.2, nous présentons une augmentation simple du formalisme pour la prise en compte d'un trait de type communicatif pour la linéarisation. L'implémentation informatique dans un système de génération de phrases que nous présentons dans la section 1.3 est basée sur ce formalisme augmenté.

### 1.1 Formalisation de base

Nous appelons « grammaire de dépendance topologique », une grammaire conçue dans le formalisme que nous avons introduit. Cette grammaire se définit par un vocabulaire  $V$ , un ensemble de catégories (lexicales)  $C$ , un ensemble de relations syntaxiques  $R$ , un série de boîtes  $B$ , une série de champs  $F$ , un champ initial  $i$ , un ordre de perméabilité des boîtes qui représente un ordre partiel de  $B$  (utilisé notamment pour l'émancipation) et quatre ensemble de règles que nous présentons ci-dessous<sup>191</sup> :

#### 1. Règles de description des boîtes :

La règle  $b \rightarrow f_1 f_2 \dots f_n$  indique que la boîte  $b$  est constituée d'une liste de champs  $f_1, f_2, \dots, f_n$ .



#### 2. Règles de description des champs :

La paire  $(f, \epsilon)$  de  $F \times \{!, ?, +, *\}$  indique que le champ  $f$  doit contenir exactement un élément (!), au plus un élément (?), un moins un élément (+) ou autant d'éléments voulus (\*).

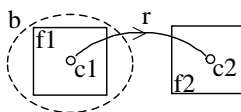
<sup>190</sup> « Au lieu de propriétés objectives, ... nous sommes confrontés à des appareils construits par nous-mêmes, et au lieu de rencontrer la nature ou l'univers, nous ne rencontrons en quelque sorte toujours que nous-mêmes. » Hannah Arendt, *Vita Activa*, Piper, 1967, page 256. (ma propre traduction)

<sup>191</sup> Nous ne présentons pas les règles lexicales qui indiquent les caractéristiques des entrées lexicales, et notamment leur catégorie.

<sup>192</sup> Dans nos figures, nous représentons les boîtes par des ovales et les champs par des rectangles ou des sections d'un ovale.

### 3. Règles de correspondance entre la structure de dépendance et la structure topologique :

La règle  $(r,c1,f1,c2,f2,b)$  indique qu'un mot  $w_2$  de catégorie  $c_2$ , qui exhibe une dépendance de type  $r$  sur un mot  $w_1$  de catégorie  $c_1$  placé dans un champ  $f_1$ , peut se placer dans le champ  $f_2$  d'une boîte contenant  $w_1$ , si cette boîte est séparée de  $w_1$  par une frontière de type  $\leq b$  (autrement dit, le paramètre  $b$  contrôle l'émancipation)



### 4. Règles de création de boîtes :

La règle  $(c,f,b,f')$  indique qu'un mot de catégorie  $c$  placé dans un champ  $f$  peut créer une boîte  $b$  et aller dans le champ  $f'$  de cette boîte. Les règles de création de boîtes sont appliquées d'une manière récursive jusqu'à ce qu'on rencontre une règle lexicale de type  $(c,f,b,-)$ , où  $b$  est une boîte lexicale contenant un champ lexical unique dans lequel le mot doit être placé.

### 5. Dérivation de la structure des constituants topologiques à partir d'un arbre de dépendance :

Le mot qui se trouve noeud racine d'un arbre de dépendance syntaxique est placé dans le champ initial  $i$ . S'appliquent alors les règles de création de boîtes et de description de boîtes en alternance, jusqu'à ce que le mot soit placé dans un champ lexical (-). Une règle de correspondance est alors activée pour l'un des dépendants du noeud racine, ce qui permettra à ce dépendant de se placer dans un champ accessible. Tout comme pour le noeud racine, les règles de création de boîtes s'appliquent jusqu'à ce que le mot soit assigné dans un champ lexical. Ce processus se répète jusqu'à placer tous les éléments de l'arbre de dépendance. Chaque fois qu'une règle de création de boîtes est déclenchée, une boîte est créée et une règle de description des champs concernant cette boîte doit être activée. Enfin, on doit vérifier que les contraintes des règles de description des champs sont bien respectées (par exemple, un champ qui demande au moins un élément ne peut rester vide).

Maintenant que le formalisme de base est établi, nous pouvons établir une grammaire de l'allemand (voir chapitre II pour une présentation descriptive de cette grammaire).

Notre grammaire de l'allemand se définit par les composantes suivantes :

$V$  = mots allemands  
 $C = \{V, AV, EV, V_{fin}, V_{inf}, VB_{se}, V_{pp}, \dots, C, N, X\}^{193}$   
 $R = \{r\}$  (on considère qu'une relation syntaxique unique  $r$  subsume toutes les autres relations)  
 $B = \{md, ed, cd, vc, vb, v, xp\}^{194}$   
 $F = \{i, vf, [, mf, ], nf, cf, h, o, u, -\}^{195}$   
 $i$  est le champ initial

et les règles suivantes :

<sup>193</sup>  $V$  = verbe,  $AV$  = verbe auxiliaire,  $EV$  = verbe avec Ersatzinfinitiv,  $V_{fin}$  = verbe fini,  $V_{inf}$  = verbe à l'infinitif avec *zu*,  $VB_{se}$  = base infinitive,  $V_{pp}$  = participe passé,  $C$  = complémenteur,  $X$  = élément non verbal,  $Y$  = un quelconque élément.

<sup>194</sup>  $md$  = domaine principal,  $ed$  = domaine enchâssé,  $cd$  = domaine de complémenteur,  $vc$  = complexe verbal,  $vb$  = boîte verbale,  $v$  = verbe,  $xp$  = syntagme non verbal.

<sup>195</sup>  $i$  = champ initial,  $vf$  = Vorfeld, '[' = parenthèse gauche,  $mf$  = Mittelfeld, ']' = parenthèse droite,  $nf$  = Nachfeld,  $cf$  = champ du complémenteur,  $h$  = tête,  $o$  = Oberfeld,  $u$  = Unterfeld,  $-$  = champ lexical,  $xf$  =  $vf/mf/nf/cf$  = champ majeur

*Ordre de perméabilité :*

vb < vc < xp = ed < cd < md

*Règles de description des boîtes :*

md	→	vf [ mf ] nf
ed	→	mf ] nf
cd	→	cf mf ] nf
vc	→	o h u
vb	→	o h
v	→	-
x	→	non décrit

*Règles de description des champs :*

(i,!), (-,!), (vf,!), (cf,!), (mf,\*), (nf,\*), ([,!), ([,?), (h,!), (o,?), (u,?)

*Règles de correspondances :*

- Placer le premier verbe dans la parenthèse droite<sup>196</sup> :  
(r, Y, [, V, ], -)
- Placer le verbe à la gauche du verbe précédant de la parenthèse droite :  
(r, V, h, V-fin, o, -)
- Placer le verbe à la droite du verbe précédant de la parenthèse droite :  
(r, AV-inf, h, EV, u, -)
- Placer un élément non verbal dans un champ majeur<sup>197</sup> :  
(r, V, ?, X, f, ed)

*Règles de création de boîtes :*

- Création du domaine principal dans le champ initial :  
(Vfin, i, md, [)
- Création d'un domaine enchâssé dans un des champs majeurs :  
(V-fin, f, ed, ])
- Création d'un complexe verbal dans la parenthèse droite ou le Unterfeld :  
(V,]/u,vc, h)
- Création d'une boîte verbale dans le Oberfeld :  
(V, o, vb, h)
- Création d'un constituant verbal :  
(V, [/h, v, -)
- Création d'un constituant non verbal :  
(X, f, xp, ?)
- Création d'un domaine pour une proposition relative<sup>198</sup> :  
("C", f, cd, "cf")

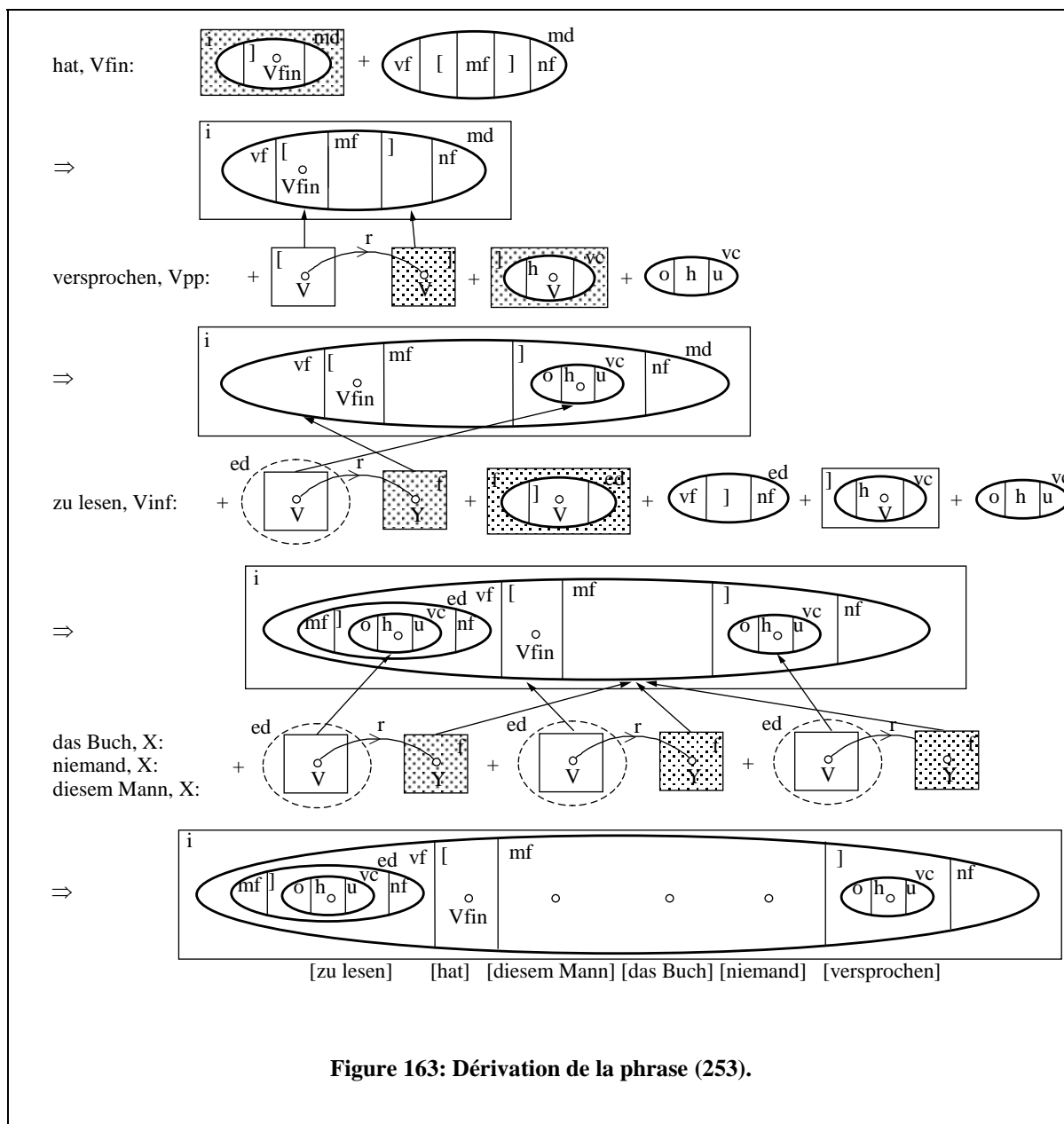
<sup>196</sup> Le dernier paramètre (-) indique que la parenthèse droite d'un domaine donné n'est pas accessible quand il y a émancipation d'un élément à partir d'un domaine enchâssé.

<sup>197</sup> Ce dernier paramètre indique qu'il est possible d'émanciper des éléments à partir de n'importe quelle boîte qui est inférieure au domaine enchâssé d'un point de vue de l'ordre de perméabilité. On peut donc faire l'émancipation à partir d'un ed, xp, vb ou vc mais pas à partir d'un cd. De plus, cette règle ne retient pas le champ du gouverneur. On doit raffiner cette règle si on veut rendre compte des phénomènes d'ordre à l'intérieur d'un syntagme nominal.

<sup>198</sup> Les guillemets indiquent que la partie complémenteur du pronom relatif n'est pas un mot réel et donc n'occupe pas le champ du complémenteur, qui par conséquent, doit prendre un autre élément.

Dans la Figure 163, nous donnons la dérivation de l'exemple sous (253) à partir de ces règles topologiques :

- (253) Zu lesen hat diesem Mann das Buch niemand versprochen.  
 A lire a cet homme (datif) le livre (accusatif) personne promis  
 'Lire, personne n'a promis à cette homme de le faire avec le livre.'



## 1.2 Augmentation des règles des traits communicatifs

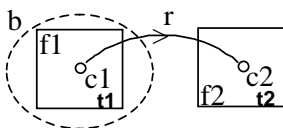
Il est essentiel pour un système de linéarisation qu'il puisse générer tous les ordres grammaticaux pour un arbre de dépendance syntaxique donné, car c'est ainsi qu'on peut analyser et améliorer les règles syntaxiques qui sont à la base du formalisme. Nous avons vu dans la section I.6 que cet ordre libre par rapport à la syntaxe est à la fois hautement restreint par la structure communicative. Il est donc aussi intéressant de modéliser cette influence à l'intérieur du système de linéarisation afin de n'obtenir que les ordres qui correspondent à la structure communicative voulue. Dans la section I.6.3, nous avons introduit ce que nous entendons par structure communicative comme une tripartition en topique, focus et tail. Nous limitons alors

l'expression de la structure communicative à un seul trait communicatif porté par chaque nœud de l'arbre de dépendance syntaxique<sup>199</sup>.

Dans le formalisme de linéarisation augmenté par l'information communicative, les règles de description des boîtes et les règles de description des champs restent inchangées. Dans les règles de correspondance entre la structure de dépendance et la structure topologique, nous ajoutons la possibilité de restreindre la règle en fonction de la valeur du trait communicatif porté par le gouverneur ou par le dépendant.

### 3. Règles de correspondance augmentées :

La règle  $(r, c_1, t_1, f_1, c_2, t_2, f_2, b)$  indique qu'un mot  $w_2$  de catégorie  $c_2$  et de trait communicatif  $t_2$ , qui exhibe une dépendance de type  $r$  sur un mot  $w_1$  de catégorie  $c_1$  et de trait communicatif  $t_1$  placé dans un champ  $f_1$ , peut aller dans le champ  $f_2$  d'une boîte contenant  $w_1$ , si cette boîte est séparée de  $w_1$  par une frontière de type  $\leq b$  (autrement dit, le paramètre  $b$  contrôle l'émancipation)



De la même manière, il est possible que la création d'une boîte puisse être conditionnée par la valeur du trait communicatif que porte l'élément qui veut ouvrir la boîte :

### 4. Règles de création de boîtes augmentées:

La règle  $(c, t, f, b, f')$  indique qu'un mot de catégorie  $c$  de trait communicative  $t$  et placé dans un champ  $f$  peut créer une boîte  $b$  et aller dans le champ  $f'$  de cette boîte. Les règles de création de boîtes sont appliquées d'une manière récursive jusqu'à ce qu'on rencontre une règle lexicale de type  $(c, t, b, -)$ , où  $b$  est une boîte lexicale contenant un champ lexical unique dans lequel le mot doit être placé.

Ces ajustements simples ont permis de faire la description de l'ordre des mots dans le Mittelfeld de l'allemand, un champ où l'ordre dépend principalement de facteurs communicatifs et non syntaxiques (cf. section 4). De la même manière, nous pouvons décrire des langues qui, dans leur fonctionnement global, ont peu de restrictions purement syntaxiques d'ordre des mots, mais présentent en même temps un lien direct entre ordre et structure communicative, comme nous le montre l'exemple le grec moderne (cf. section 7).

## 1.3 L'implémentation DepLin

L'informatique offre à la linguistique non seulement une raison d'être économique mais également sociale ; généralement, l'implémentation d'une idée dans un logiciel permet une compréhension plus profonde du phénomène et des analyses plus précises que la seule étude théorique. Peu de temps après le développement des idées de bases sur la topologie de l'allemand, nous avons implémenté les règles que nous avons retenues, dans un petit logiciel baptisé DepLin (comme Dependency Linearizer). Cela était d'autant plus nécessaire parce que malgré une simplicité apparente des croisements possibles des règles, ces derniers dépassaient

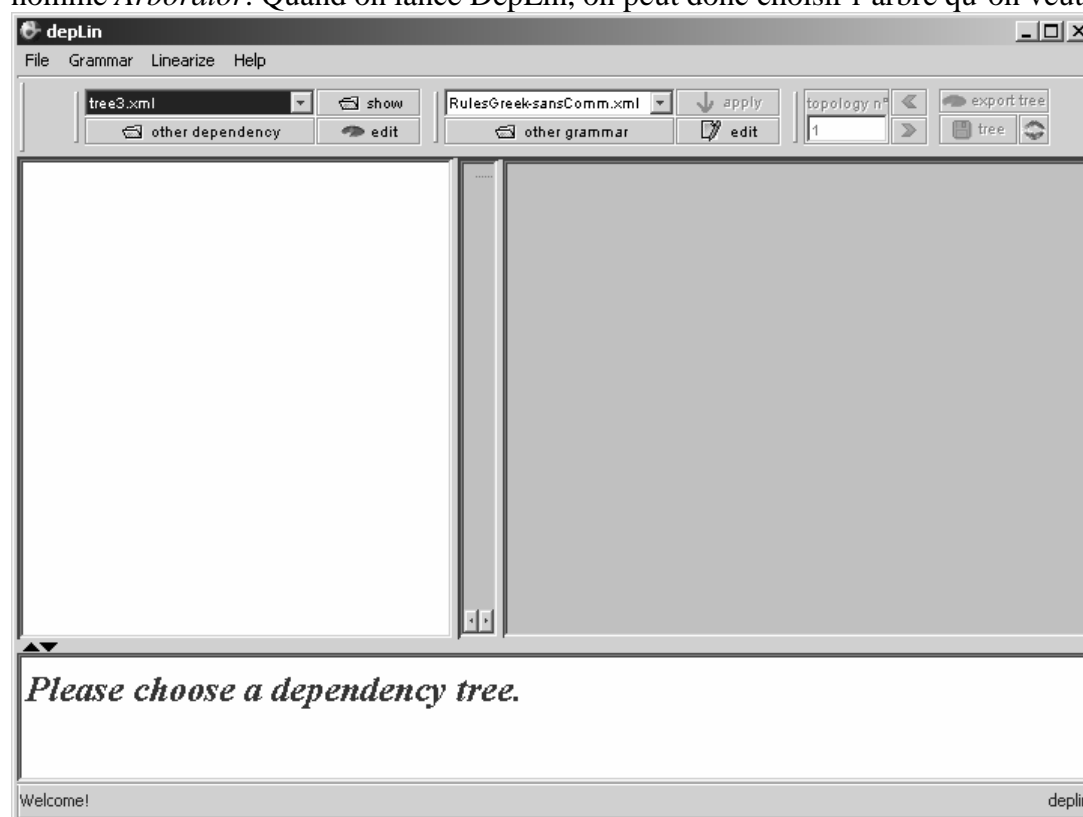
<sup>199</sup> Remarquons d'abord que si une analyse linguistique plus fine nous obligerait à utiliser plus qu'un seul trait communicatif, il est facile d'envisager la prise en compte de toute une structure de traits communicatifs. Par ailleurs, il est possible de traduire et prendre en compte toute structure complexe en croisant toutes les différentes valeurs dans un seul trait. Cela risque, bien entendu, de brouiller les généralités qu'on veut exprimer sur l'influence de ces traits sur l'ordre des mots.

souvent les capacités de calcul d'un humain<sup>200</sup>. DepLin est un outil qui calcule automatiquement tous les ordres possibles pour un arbre de dépendance donné. Ce logiciel a été très utile pour affiner les règles de l'allemand et dans l'établissement des règles du grec moderne (cf. section 7), et de manière générale pour comprendre la signification de la structure topologique. DepLin est programmé en Java. Il a une interface graphique qui permet de visualiser en entrée l'arbre de dépendance syntaxique, et en sortie, les arbres topologiques et tous les ordres obtenus dans un affichage texte.

Les structures arborescentes d'entrée et de sortie sont encodées en XML. Voici notre exemple de référence dans cet encodage :

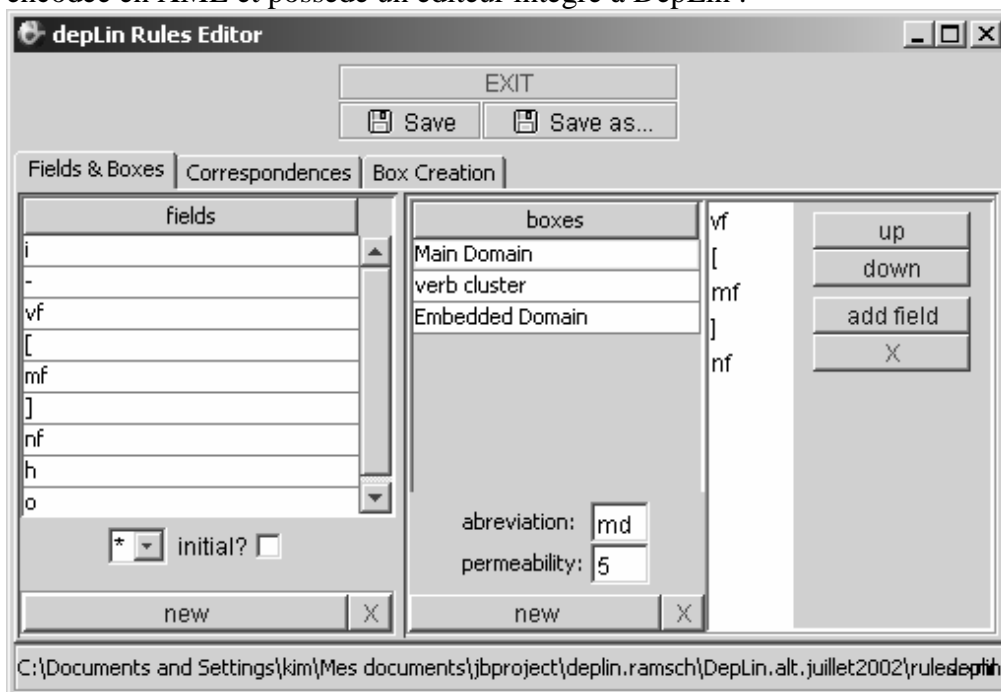
```
<?xml version='1.0' encoding='utf-8'?>
<!-- A SAMPLE dependency tree -->
<lingtree
  title="Sample dependency tree"
  type="dependency">
  <word name="hat" cat="VFin">
    <word name="Peter" relation="subj" cat="noun"></word>
    <word name="versprochen" relation="pp" cat="Vppinf">
      <word name="diesem Mann" relation="indirobj" cat="noun"></word>
      <word name="zu lesen" relation="zuinf" cat="Vzuinf">
        <word name="den Roman" relation="dirobj" cat="noun"></word>
      </word>
    </word>
  </word>
</word>
</lingtree>
```

Ces structures peuvent aussi être créées à l'aide d'un éditeur d'arbres que nous avons programmé, nommé *Arborator*. Quand on lance DepLin, on peut donc choisir l'arbre qu'on veut linéariser :

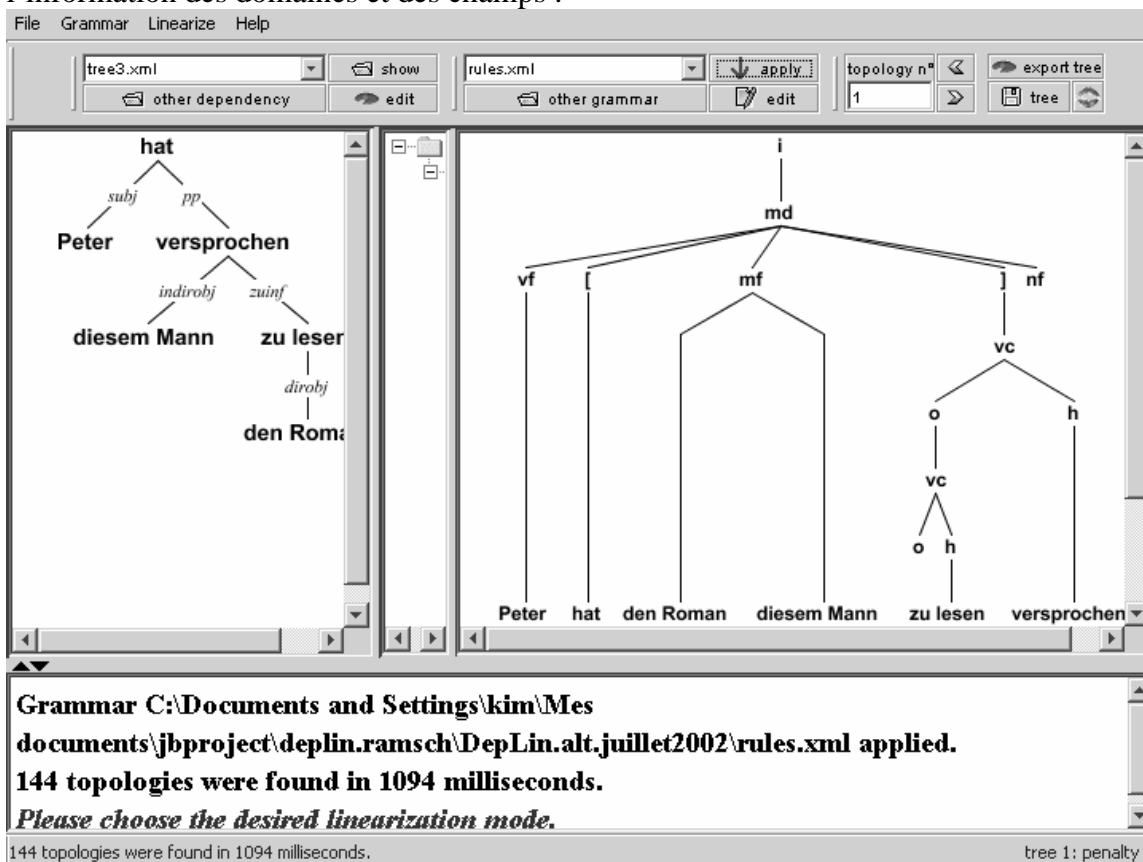


<sup>200</sup> Il peut paraître paradoxal qu'un objet qui modélise une activité humaine (la faculté de parler) dépasse les capacités de calcul d'un être humain. Mais un être humain ne cherche pas à calculer tous les ordres possibles dans une situation ordinaire de communication.

Une fois que l'arbre est chargé, on peut choisir la grammaire de linéarisation. Elle aussi est encodée en XML et possède un éditeur intégré à DepLin :



Après l'application des règles, on peut afficher les résultats sous forme d'arbres ou sous forme de liste de phrases, ordonnée selon différents critères, et augmentée, si on le souhaite, de l'information des domaines et des champs :





The screenshot displays a software application for generating phrase structures. The interface is divided into several sections:

- Top Bar:** Contains menu items 'File', 'Grammar', 'Linearize', and 'Help'. It also features buttons for 'show', 'apply', 'export tree', 'quality order', 'edit', and 'tree', along with a 'topology n\*' field set to 2 and a 'Linearize' button.
- Left Panel:** Shows a tree structure for the sentence 'Peter hat diesem Mann versprochen zu lesen den Roman'. The root node is 'hat', which branches into 'subj' (Peter) and 'pp' (versprochen). 'versprochen' further branches into 'indirobj' (diesem Mann) and 'zuinf' (zu lesen). 'zu lesen' branches into 'diobj' (den Roman).
- Right Panel:** Shows a larger tree structure with root node 'i', which branches into 'md'. 'md' branches into 'vf', 'mf', and 'nf'. 'mf' branches into 'vc', which further branches into 'o' and 'h'. 'o' branches into 'vc', which further branches into 'o' and 'h'.
- Bottom Panel:** Lists 37 generated phrase structures, each represented as a sequence of nodes and their children in brackets, such as:
  1. [<sub>md</sub> Peter<sub>vf</sub> hat<sub>i</sub> den Roman<sub>mf</sub> diesem Mann<sub>mf</sub> [<sub>vc</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ]<sub>o</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  2. [<sub>md</sub> Peter<sub>vf</sub> hat<sub>i</sub> diesem Mann<sub>mf</sub> den Roman<sub>mf</sub> [<sub>vc</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ]<sub>o</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  3. [<sub>md</sub> Peter<sub>vf</sub> hat<sub>i</sub> diesem Mann<sub>mf</sub> [<sub>vc</sub> versprochen<sub>h</sub> ] ]<sub>i</sub> [<sub>ed</sub> den Roman<sub>mf</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ] ]<sub>nf</sub> ]<sub>i</sub>
  14. [<sub>md</sub> diesem Mann<sub>vf</sub> hat<sub>i</sub> den Roman<sub>mf</sub> Peter<sub>mf</sub> [<sub>vc</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ]<sub>o</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  15. [<sub>md</sub> diesem Mann<sub>vf</sub> hat<sub>i</sub> Peter<sub>mf</sub> den Roman<sub>mf</sub> [<sub>vc</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ]<sub>o</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  16. [<sub>md</sub> diesem Mann<sub>vf</sub> hat<sub>i</sub> Peter<sub>mf</sub> [<sub>vc</sub> versprochen<sub>h</sub> ] ]<sub>i</sub> [<sub>ed</sub> den Roman<sub>mf</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ] ]<sub>nf</sub> ]<sub>i</sub>
  27. [<sub>md</sub> den Roman<sub>vf</sub> hat<sub>i</sub> diesem Mann<sub>mf</sub> Peter<sub>mf</sub> [<sub>vc</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ]<sub>o</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  28. [<sub>md</sub> [<sub>ed</sub> den Roman<sub>mf</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ] ] ]<sub>vf</sub> hat<sub>i</sub> diesem Mann<sub>mf</sub> Peter<sub>mf</sub> [<sub>vc</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  36. [<sub>md</sub> den Roman<sub>vf</sub> hat<sub>i</sub> Peter<sub>mf</sub> diesem Mann<sub>mf</sub> [<sub>vc</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ]<sub>o</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
  37. [<sub>md</sub> [<sub>ed</sub> den Roman<sub>mf</sub> [<sub>vc</sub> zu lesen<sub>h</sub> ] ] ]<sub>vf</sub> hat<sub>i</sub> Peter<sub>mf</sub> diesem Mann<sub>mf</sub> [<sub>vc</sub> versprochen<sub>h</sub> ] ]<sub>i</sub>
- Bottom Status Bar:** Displays the message '144 topologies were found. Results saved in phrases.html.' and 'tree 2: penalty 0'.

Dans une étape suivante, qui est encore dans un stade expérimental, on peut écouter les phrases générées. Nous disposons pour cela d'une autre série de règles : Ces règles permettent de regrouper les différentes suites d'éléments portant la même valeur communicative en groupes prosodiques. Ensuite, en fonction de la position dans le groupe prosodique et dans la phrase dans sa totalité, on calcule le schéma prosodique correspondant à chaque mot. Ensuite, le logiciel choisit le bon schéma prosodique du mot parmi une série de fichiers par mots, qui, dans l'état actuel de l'implémentation sont encore préenregistrés. Ensuite ces fichiers sont collés ensemble et rendu audibles. Les détails de ce procédé seront le sujet d'une autre thèse (Yoo 2003, Gerdes & Yoo 2003 en préparation). Cf. section 7 pour plus de détails en ce qui concerne l'état actuel de la grammaire topologique du grec moderne.

## 2 Le problème du sujet dans un domaine enchâssé

Dans cette section nous discuterons un phénomène linguistique très délicat de la syntaxe de l'allemand : il s'agit des cas où le domaine enchâssé inclut le sujet de la phrase. Ce phénomène est très particulier et difficile à traiter, d'une part à cause du nombre très limité d'occurrences qui apparaissent dans la langue naturelle, et d'autre part, à cause de la variabilité du jugement des locuteurs natifs sur ce type d'énoncés.

Notons que les règles formelles du modèle topologique que nous avons proposées dans la section précédente ne permettent pas de traiter ce phénomène ; sans modifier de façon ad hoc les arbres de dépendance des phrases concernées.

Nous présenterons d'abord les restrictions du placement du sujet dans les domaines enchâssés en essayant de dégager les critères syntaxiques et sémantiques qui sont en jeu. L'étude de ces

critères nous amènera à distinguer des acceptabilités autonomes de types topologique, syntaxique et sémantique.

Cette étude sur le sujet dans les domaines enchâssés sera à la base d'une proposition d'un formalisme d'arbres lexicalisés basé sur le modèle topologique que nous présentons dans la section suivante.

## 2.1 Les données sur le VP fronting

Parmi les champs d'un domaine principal, le Vorfeld est le champ où on trouve le plus fréquemment et d'une manière facilement reconnaissable, un domaine enchâssé. Selon Drach 1937:21, « le syntagme qui est déplacé dans le Vorfeld – quel que soit son usage grammatical – peut être subdivisé arbitrairement. Il reste toujours une entité »<sup>201</sup>. Or si l'on observe les éléments qui peuvent se trouver dans le Vorfeld, on trouve des syntagmes verbaux contenant le sujet de la phrase, ce qui est problématique car d'un point de vue topologique, ce syntagme forme une unité alors qu'il s'agit d'une portion non connexe de l'arbre de dépendance puisque le sujet dépend du verbe fini qui est dans la parenthèse gauche. Etablir quelles sont les règles qui gouvernent la formation de ces domaines enchâssés ouverts par un verbe non fini (qu'on trouve sous le nom de VP fronting) semble non évident.

Examinons quelques exemples. Les exemples sous (254), (255) et (256) contiennent un syntagme nominal qui se joint au participe passé (ou au participe au passif) afin de former un constituant unique dans le Vorfeld. La première question qui se pose est de savoir qu'elle est l'entité qui arrive dans le Vorfeld dans des phrases comme exemples sous (254) et (255).

(254) a. Den Roman gelesen hat Peter bisher nicht.

Le roman (acc) lu a Pierre jusqu'à\_ici non.  
'Pierre n'a pas lu encore le roman.'

b. Ein berühmter Geiger geworden wäre er gerne.<sup>202</sup>

Un fameux violoniste (nom) devient, serait il avec\_plaisir.  
'Il aurait aimé devenir un violoniste connu.'

(255) a. Ein Linguist angekommen ist (\*sind) bisher nicht.<sup>203</sup>

un linguiste (nom) arrivé est (\*sont) jusqu'à\_ici non.  
'Aucun linguiste n'est encore arrive.'

b. Solche schönen Geschenke gemacht wurden (\*wurde) mir noch nie.

de\_tels beaux cadeaux (nom.) faits étaient (\*était) à\_moi encore jamais  
'Je n'ai jamais reçu des cadeaux aussi beaux'.

c. Von Grammatikern angeführt werden auch Fälle mit dem Partizip intransitiver Verben.<sup>204</sup>

Par grammairiens (dat) cités sont aussi cas avec des participes de\_intransitif verbes.  
'Des cas avec des participes des verbes intransitifs sont aussi cités par des grammairiens.'

<sup>201</sup> Texte original « das ins Vorfeld verlegte Satzglied – gleichviel, wie es grammatisch verwendet sei – kann beliebig untergliedert werden. Immer jedoch bleibt es ein Ganzes » (ma propre traduction)

<sup>202</sup> Le contraste entre les exemples (254)b. et (256)a. qui ont tous les deux des constituants topicalisés avec des arguments nominatifs montre que la distinction entre le terme de *sujet* et d'*argument au nominatif* est indispensable, bien que nous utilisons souvent *nominatif*, de manière imprécise, pour parler de l'argument portant ce cas qui a aussi la fonction de sujet.

<sup>203</sup> Nous empruntons cet exemple à Haider 1985.

<sup>204</sup> Nous empruntons cet exemple à Müller 1999.

- (256) a. ?\* Ein Linguist geschlafen hat bisher nicht.  
Un linguiste (nom) dormi a jusqu' \_ici non
- b. ?\* Dieser Frau unterlaufen ist ein Fehler noch nie.  
à\_cette femme (dat) échappé est un erreur encore jamais.

## 2.2 Interprétation prosodique et communicative des données

Une première possibilité de réponse à la qualité de groupes constitués dans le Vorfeld est qu'ils constituent une unité prosodique : les mots qui se trouvent dans ce champ forment un groupe caractérisée par une courbe mélodique particulière qui dépend du contexte dans lequel la phrase a été énoncée. Par ailleurs, on ne peut trouver de pause entre ces mots qui forment le groupe. Ainsi, comme réponse à la question (257), la phrase (254)a. n'est possible qu'accompagnée d'un contour mélodique descendant. Le contexte fait que le Vorfeld contient le rhème de la phrase et ce contour mélodique descendant est spécifique à une partie rhématique du discours<sup>205</sup>. De la même manière, si pour la même phrase on fabrique un contexte où *den Roman gelesen* 'le roman lu' obtient un caractère rhématique, ce qui est le cas de la phrase (257)b., le Vorfeld sera accompagné d'un contour mélodique plat (ce qui caractérise les thèmes non focalisés) ou un contour mélodique montant (dans le cas où le thème est focalisé). La même analyse est valable pour la phrase (254)b. comme réponse aux questions sous (258).

- (257) a. Was hat Peter noch nicht getan?  
'Qu'est-ce que Pierre n'a pas encore fait ?'
- b. Hat Peter den Roman gelesen?  
'Pierre a-t-il déjà lu le roman ?'
- (258) a. Was wäre er gerne?  
'Qu'est-ce qu'il aimerait devenir ?'
- b. Wollte er ein berühmter Geiger werden?  
'Voulait- il devenir un violoniste connu ?'

Ces données montrent clairement que le groupement des éléments dans le Vorfeld n'est pas seulement caractérisé par un contour prosodique spécifique, mais aussi que ce groupement joue un rôle communicatif spécifique, puisqu'il représente une entité thématique ou rhématique. Nous utiliserons la répartition tripartite de la structure communicative<sup>206</sup> (que nous avons introduite dans la section I.6.3) pour expliquer les différents groupements qui peuvent avoir lieu dans le Vorfeld. Ainsi, le Vorfeld peut contenir des éléments qui sont soit rhématiques (ou encore un focus), soit thématique (un topique dans le cas thèmes proéminents ou un thème dans le cas des thèmes non proéminents)<sup>207</sup>.

---

<sup>205</sup> Voir Gibbon 1998.

<sup>206</sup> Voir également Choi 1999 pour une distinction similaire, proposée pour l'analyse de l'allemand et du coréen.

<sup>207</sup> On trouve également chez Drach 1937 une caractérisation du Vorfeld en terme de valeurs prosodiques et communicatives. L'auteur note que le Vorfeld peut être la *position expressive* (*Ausdruckstelle*) soit pour des « mots sémantiquement non vides ayant une valeur d'émotion ou de volonté » soit pour « une information mineure ou une connexion avec une information donnée ». Par contre, il n'explique pas la possibilité de remplir cette *position expressive* par une information donnée, c'est-à-dire le topique.

### 2.3 Le rôle de la structure communicative dans le groupement des mots

Nous avons donc vu que le constituant qui se place dans le Vorfeld est en quelque sorte muni d'une certaine valeur communicative, ce qui lui donne également un contour prosodique spécifique. La question qui se pose est de savoir où cette structure communicative se matérialise. Dans la Théorie Sens-Texte où on fait clairement la distinction entre les différents niveaux de représentations, la structure communicative apparaît comme structure autonome dès le niveau de la représentation sémantique. Si l'on considère que le groupement des mots dans le Vorfeld est aussi motivé par la structure communicative, alors le niveau de représentation sémantique peut aussi influencer sur la linéarisation finale, et la possibilité (ou l'impossibilité) de regrouper des mots dans le champ du Vorfeld. Or la structure topologique qui contient l'ordre linéaire, est étroitement liée à l'arbre de dépendance syntaxique de surface. Ainsi, le fait d'avoir des constituants topicalisés dans le Vorfeld comme les phrases (254) et (255) mais pas comme dans les phrases (256) dépend de restrictions qui sont imposées avant le passage de l'arbre syntaxique de surface à la structure topologique.

En principe, tous les constituants sauf les sujets comme dans la phrase (256)a. peuvent occuper la place du Vorfeld, ce qui nous amène à considérer que les verbes non finis de l'allemand forment des syntagmes verbaux (VP). Cependant, il existe des exceptions à cette règle : en effet, certains syntagmes nominaux qui portent un autre cas que le nominatif sont également difficiles à grouper avec le verbe, comme nous pouvons le voir dans l'exemple (256)b., où le syntagme nominal au datif et son gouverneur ne peuvent pas former une entité communicative. On trouve ce genre de complications avec des arguments qui n'ont pas de rôle agentifs tandis que les verbes ergatifs (exemple (255)b.) et les verbes à la voix passive (exemple (255)c.) semblent accepter d'être accrochés au sujet (et à l'objet profond). La difficulté de toutes les approches basées sur la structure syntagmatique telle que HPSG est que la linéarisation, l'accord et la structure prédicat-argument sont tous basés sur les syntagmes. Ainsi, pour les phrases (255)a. et b., on doit expliquer où et comment se fait l'accord entre le sujet et le verbe. Est-ce que le VP topicalisé dans le Vorfeld porte la valeur de l'accord ou est-ce que les « constituants -fantômes »<sup>208</sup> portent cette information vers le VP topicalisé ? Inversement, le syntagme prépositionnel optionnel de l'exemple (254)c. doit recevoir la théta-rôle de l'agent du verbe (voir Müller 2000).

Finalement, comme l'observe également Weibelhuth 1985, la relation sémantique qui existe entre le prédicat et le nom semble jouer un rôle beaucoup plus important dans les restrictions portant sur le *VP fronting* que le marquage casuel des noms. En fait, lorsque le locuteur décide du groupement communicatif de son acte de parole, les restrictions s'appliquent sur l'information sémantique. L'exemple (254)c. montre qu'il ne suffit pas de bloquer simplement le groupement communicatif entre l'agent et le prédicat. Nous avons plutôt besoin d'une règle spécifique qui s'applique sur l'interface syntaxe – sémantique. Nous définissons cette règle d'interface de la manière suivante :

**L'absence d'argument agentif ainsi que le groupement communicatif de l'argument agentif avec son prédicat déclenche la construction passive<sup>209</sup>.**

Sans rentrer dans le détail de l'interface entre sémantique et syntaxe dans la Théorie Sens-Texte (Mel'čuk 1988), nous considérons donc que les restrictions qui portent sur la formation du *VP fronting* est d'ordre sémantique et nous essayerons de montrer par la suite que ce rôle supplémentaire que nous donnons à ce module n'est pas très difficile à assumer.

<sup>208</sup> Meurers 1999: *Raising Spirits (and assigning them case)*.

<sup>209</sup> La construction passive peut bien sûr être déclenchée par d'autres phénomènes, tel que la continuité du discours et la pertinence des actants dans l'événement décrit, que nous ne traitons pas dans ce présent travail.

## 2.4 *Que reste-t-il pour le module syntaxique ?*

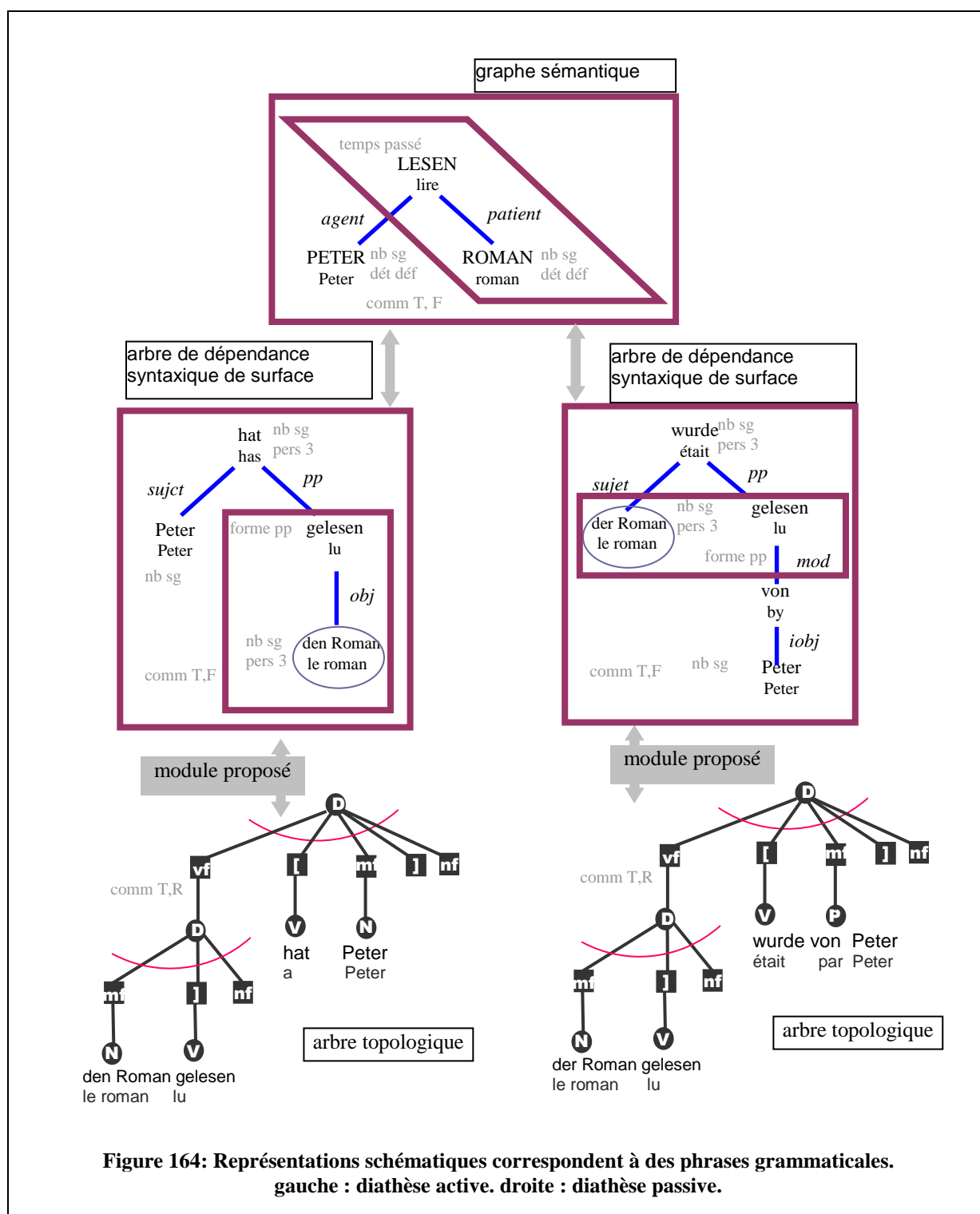
Dans l'arbre de dépendance syntaxique de surface où les éléments apparaissent de manière non ordonnée, l'accord peut être porté indépendamment de la linéarisation à laquelle on va aboutir. Ainsi, notre module syntaxique, qui a comme rôle de lier la dépendance syntaxique de surface à la hiérarchie topologique, n'a pas à se soucier des restrictions sur la formation des syntagmes verbaux enchâssés. En génération, le module génère les linéarisations possibles pour un arbre de dépendance donné, qui transforme les groupements dans des structures de domaines enchâssés ; en analyse, les groupes rencontrés sont transférés dans une structure syntaxique de surface. La tâche de refuser des phrases agrammaticales comme (256) revient donc au module sémantique, ce qui correspond bien à notre intuition que l'agrammaticalité de ces phrases est d'une nature différente que les exemples de conflits de cas dans les variantes agrammaticales des phrases (255)a. et b. (indiqués entre parenthèses et précédés d'un astérisque). L'agrammaticalité des phrases (256) est moins claire d'un point de vue syntaxique ; il semble plutôt qu'il est difficile de décoder ou de comprendre ce que le locuteur veut dire, et donc on peut dire que l'agrammaticalité est d'ordre sémantique. Il apparaît donc souhaitable de distinguer plusieurs niveaux de bonne formation des phrases.

Si l'analyse est liée au module syntaxique où l'accord a été vérifié avec succès, comme c'est le cas des phrases sous (256) nous considérons ces phrases comme « syntaxiquement bien formées » et « sémantiquement défectueuses ». De la même manière, nous avons pu assigner une structure topologique aux versions agrammaticales des phrases (255)a. et b. ; le conflit survient en fait quand l'accord est vérifié au niveau syntaxique. Nous définissons donc ces phrases comme « topologiquement bien formées » mais « syntaxiquement défectueuses ».

## 2.5 *Exemples de correspondances*

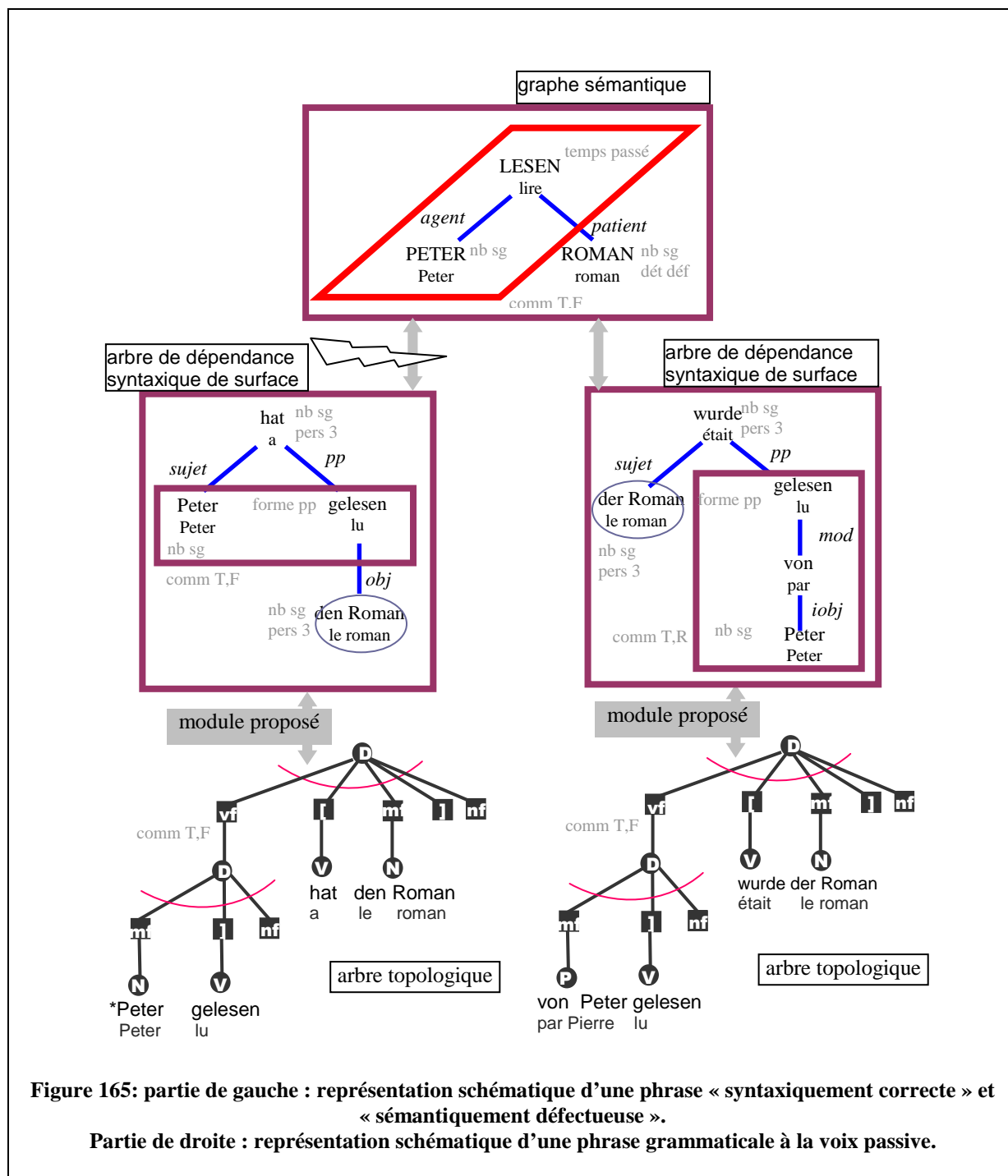
Suite à la discussion du paragraphe précédent, nous plaçons donc pour une analyse grammaticale de la phrase allemande sous forme d'une correspondance compositionnelle entre au moins trois représentations, chacune d'entre elles vérifiant des conditions de bonne formation. La Figure 164 illustre deux structures simplifiées pour deux phrases grammaticales. Ces phrases correspondent à l'analyse d'une même représentation sémantique. La raison est que notre représentation sémantique est très simplifiée et pas assez fine pour qu'elle puisse capturer l'information sémantique apportée par les choix différents de diathèse. De la même manière, notre représentation syntaxique de surface ne prend pas en compte les variations d'ordres des mots qui existent à l'intérieur d'un champ, et deux linéarisations peuvent correspondre à la même représentation syntaxique.

Comme il s'agit d'un texte écrit, nous ne trouvons aucune information sur la structure prosodique du constituant qui se place dans le Vorfeld ; ses valeurs communicatives restent non spécifiées (il peut donc être focus ou topique). Un input du module prosodique (un analyseur de la parole) spécifierait le contour mélodique et les valeurs communicatives pourraient être ajoutées.



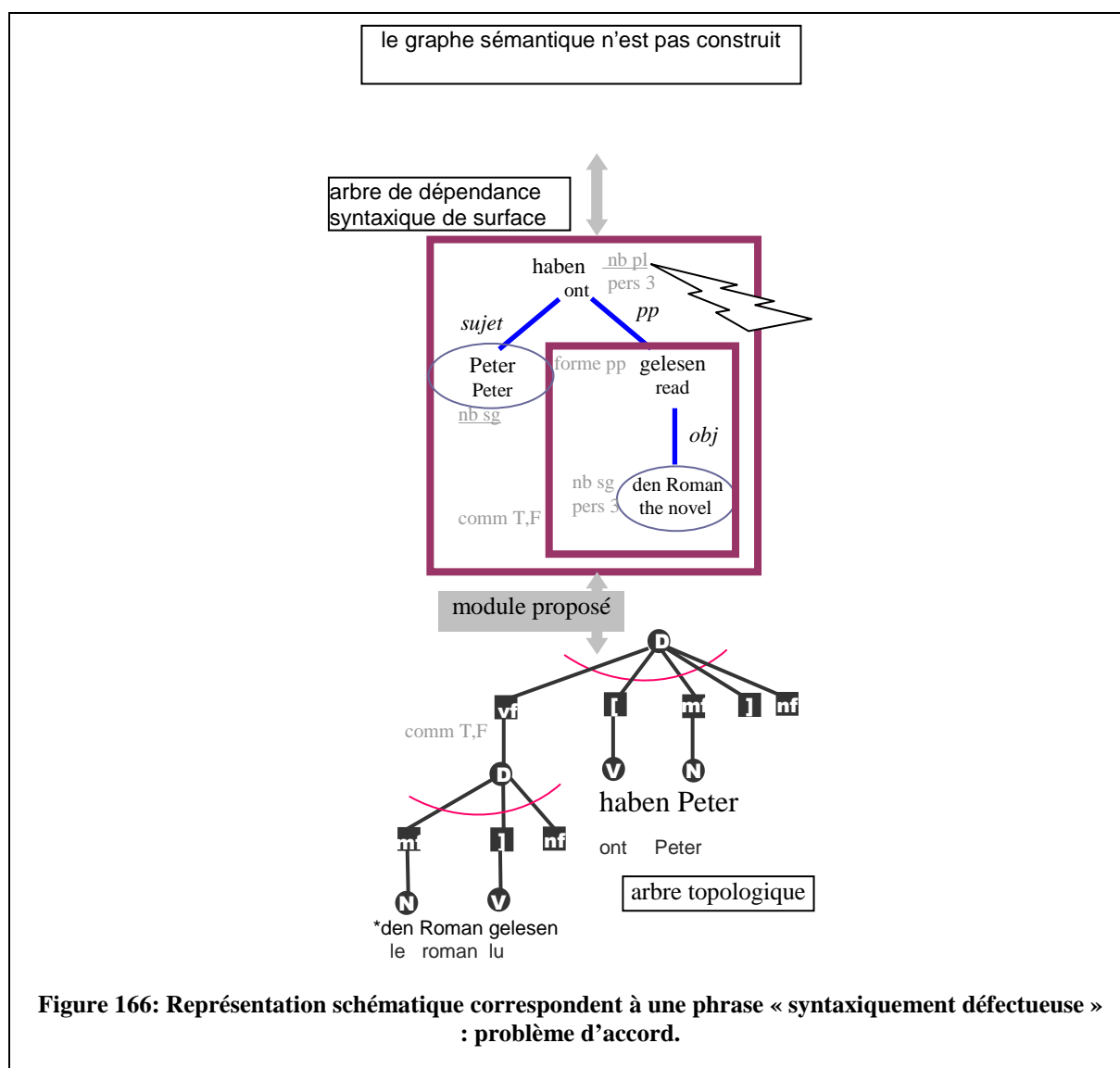
**Figure 164: Représentations schématiques correspondant à des phrases grammaticales. gauche : diathèse active. droite : diathèse passive.**

Dans la Figure 165, la dérivation de la partie de gauche représente une phrase « sémantiquement défectueuse ». En analyse, la structure topologique est constructible et peut même être traduite en une structure syntaxique. Cependant, le module sémantique n'est pas autorisé à former un groupe contenant un verbe et un patient : il échoue car il doit transférer un groupe contenant le verbe et le sujet dans un groupe de prédicat-agent, ce qui est interdit par la règle de bonne formation au niveau de l'interface sémantique–syntaxe que nous avons présenté dans la sous-section 2.3. Le groupement du prédicat avec son agent est par contre possible dans le cas de la construction passive, comme nous le montrons dans la dérivation de la partie droite de la Figure 165.



**Figure 165: partie de gauche : représentation schématique d'une phrase « syntaxiquement correcte » et « sémantiquement défectueuse ».  
Partie de droite : représentation schématique d'une phrase grammaticale à la voix passive.**

La phrase de la Figure 166 représente des phrases « syntaxiquement défectueuses ». Néanmoins, ces phrases sont topologiquement bien formées puisque le problème d'accord, qui apparaît dans la Figure 166 par exemple, n'interdit pas le module syntaxique de produire une structure topologique et même un arbre de dépendance syntaxique de surface. C'est la condition de bonne formation au niveau syntaxique qui vérifie l'accord du sujet avec le verbe.

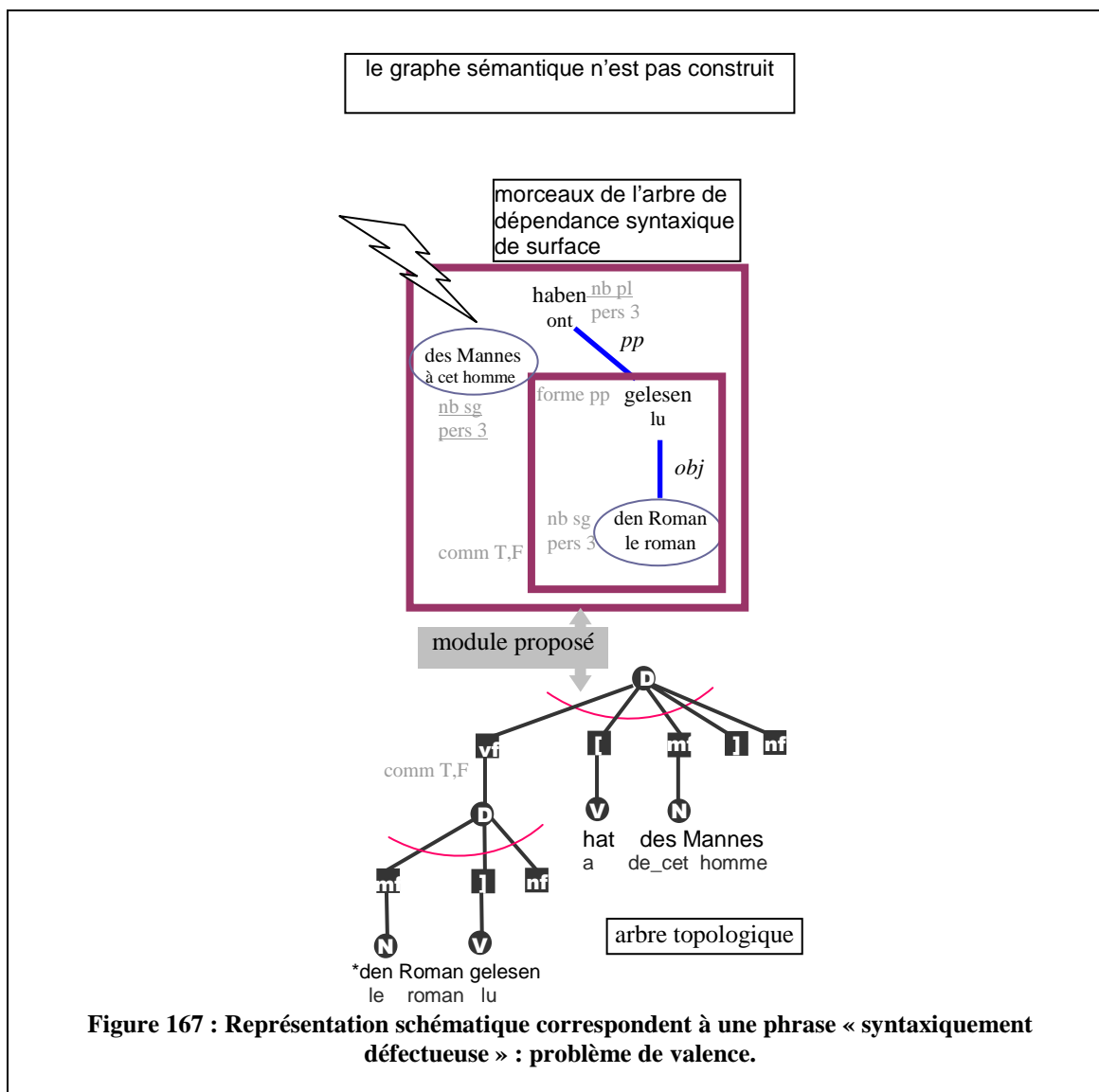


Le cas de la Figure 167 est différent : en effet, comme le syntagme nominal au génitif n'est pas un argument syntaxique de l'auxiliaire, nous ne pouvons pas établir de relation syntaxique entre l'auxiliaire et ce syntagme. Donc, soit on est obligé de dire que dans ce cas, une structure syntaxique ne peut être créée, soit on doit conclure que les parties non connexes représentent une structure syntaxique qui ne remplit pas les conditions de bonne formation de la connexion<sup>210</sup>.

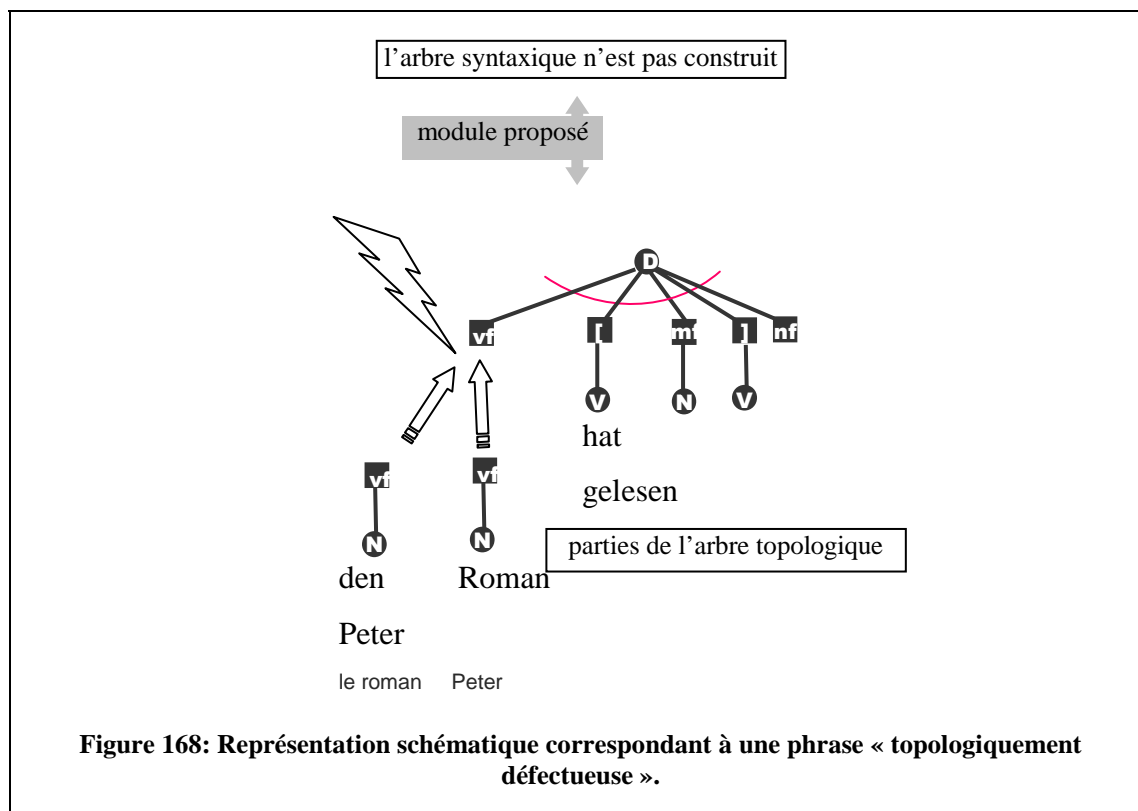
<sup>210</sup> En ce qui concerne la phrase de la Figure 167, la situation aurait été différente sans le syntagme nominal au génitif. La structure syntaxique serait connectée et bien formée puisque l'accord ne présente pas de conflit. Par contre, il y aura un problème lorsque le module sémantique remarquera l'absence d'un argument agentif pour le sémantème *lesen* 'lire'. Notre approche permet aussi une description élégante des constructions avec des verbes « sans sujet » comme dans l'exemple (i), où le sujet de l'auxiliaire *hat* 'a' reste optionnel. Le module sémantique passe un argument nominatif à une position agentive seulement s'il y a un accord en nombre et en personne.

- (i) Mir hat gegraut.  
à\_moi (dat) a eu\_horreur  
'j'ai eu horreur.'





Enfin, la Figure 168 montre une phrase « topologiquement défectueuse ». Dans cette phrase, den Roman ‘le roman’ et Peter ‘Pierre’ ne peuvent pas créer de nouveau domaine leur permettant d’atterrir tous les deux dans le Vorfeld, et donc une structure topologique connexe ne peut être construite.



Nous espérons avoir montré que les grammaticalités et les agrammaticalités des phrases peuvent être motivées par des restrictions d'ordre sémantique, syntaxique ou topologique. Ainsi, le problème du sujet se trouvant dans un domaine enchâssé qui a première vue paraît très complexe, peut être résolu facilement si on considère le fait qu'une phrase peut être syntaxiquement, sémantiquement ou topologiquement défectueuses, ce qui limite finalement le travail du module syntaxique.

Sur la base des analyses que nous avons définies dans cette section, nous allons dans la section suivante proposer une formalisation d'un modèle syntaxique, qui ne s'occupe que du lien entre topologie et syntaxe, et que laisse l'interdiction de certains groupements de mots au module sémantique.

### 3 Grammaires d'unification d'arbres (TUG)

Dans cette section, nous introduisons une nouvelle grammaire d'arbres lexicalisée dans la famille des TAG basée sur l'unification des structures d'arbres. Nous appelons ce formalisme *Tree Unification Grammar* (TUG, Grammaire d'unification d'arbres). La description que nous proposons de ce formalisme est orientée plutôt vers l'analyse, mais les règles sont bidirectionnelles, donc applicables en génération.

Dans la section précédente, nous avons défini ce que l'algorithme d'un module syntaxique est supposé réaliser : à partir d'une chaîne de mots, construire une structure de constituants topologiques, et d'une manière compositionnelle et parallèle, un arbre d'une structure syntaxique de surface. De plus, le groupement des têtes lexicales à l'intérieur d'un domaine topologique (avec éventuellement un marquage communicatif) est une information qui devrait passer au niveau de l'arbre de dépendance syntaxique. Ce n'est qu'au passage vers un niveau plus profond (peut-être sémantique) que certains groupes peuvent être interdits.

Nous aimerions arriver à accomplir la tâche de construction de la structure topologique à l'aide d'une combinaison simple de morceaux d'arbres lexicalisés, en suivant l'exemple des TAG. Cette construction doit être compositionnelle dans le sens qu'une bonne combinaison au niveau topologique de deux structures lexicalisées devra créer directement le lien syntaxique de surface entre les deux lexies qui ancrent ces deux structures lexicalisées.

TUG représente principalement une lexicalisation du formalisme de linéarisation topologique nous avons présenté dans la Section 1 (voir également Gerdes et Kahane 2001a.). Par ailleurs, nous empruntons les notions et les idées des TAG et ses extensions.

Les problèmes des dépendances longue comme le brouillage d'arguments et l'extraction *wh*-, déjà discutés par Vijay-Shanker 1987, sont repris et partiellement résolus par les DTG (*D-Tree Grammars*<sup>211</sup>, Rambow *et al.* 1995, voir section III.2.4.2). Les GAG (Grammaires d'adjonction guidées par les graphes, angl. Graph-Driven Adjunction Grammars, Kahane, 1997, Candito et Kahane 1998b) constituent une extension des DTG qui permet d'obtenir comme structure de dérivation non seulement des arbres mais aussi des graphes. L'AES d'Alexis Nasr 1996 place pour la première fois la grammaire d'arbres lexicalisée dans le cadre de la Théorie Sens-Texte et les grammaires à bulles ainsi que les GAG, tout comme les TUG, pose le problème du lien qui existe entre la grammaire de dépendance et une structure syntagmatique ordonnée. Nous renvoyons à Kahane 2001 pour une tentative de décrire d'une manière lexicalisée, toute la grammaire Sens-Texte. Le but des TUG est principalement de servir comme module de correspondances entre l'arbre de dépendance syntaxique de surface et la structure de constituants topologiques.

### 3.1 Définition des TUG

Soit  $V$  un alphabet,  $D \in V$  un symbole distinct de  $V$  et  $W$  un ensemble de mots sur  $V$ .

Nous appelons *atomes* les nœuds des arbres qui sont distingués par l'étiquette  $L$ , élément de  $V$ , et par un trait de couleur binaire. Ce trait peut prendre la valeur *pleine* (par exemple ■ ●, la distinction entre les atomes carrés et ronds étant là uniquement pour une meilleure lisibilité) ou la valeur *vide* (○). Les atomes peuvent en plus avoir une structure de traits simples non enchâssée. Les traits des nœuds vides peuvent aussi être des *traits fonctionnels*, que nous définirons plus bas. Deux atomes peuvent s'unifier si et seulement s'ils ont la même étiquette.

Un *arbre élémentaire* TUG a des atomes comme nœuds. Les atomes feuilles peuvent être lexicalisés. On les considère comme lexicalisés lorsqu'ils sont associés à un mot-forme. Les arêtes d'un arbre élémentaire peuvent porter un marquage d'*exclusion d'adjonction sœur* à leur droite ou à leur gauche.

Par exemple, l'arbre  $\alpha$  de la Figure 169 n'a pas d'exclusion d'adjonction. L'arbre  $\beta$  ne permet pas à l'atome  $b$  d'avoir une sœur à sa gauche ; pour tout arbre qui s'unifiera avec  $\beta$ , la liste des enfants de l'atome  $x$  commencera toujours avec  $b$ . De la même manière, dans l'arbre  $\gamma$ , les atomes  $a$  et  $b$  seront toujours des voisins directs, et dans l'arbre  $\delta$ ,  $a$  et  $b$  resteront toujours au début de la liste des enfants de l'atome  $x$ .

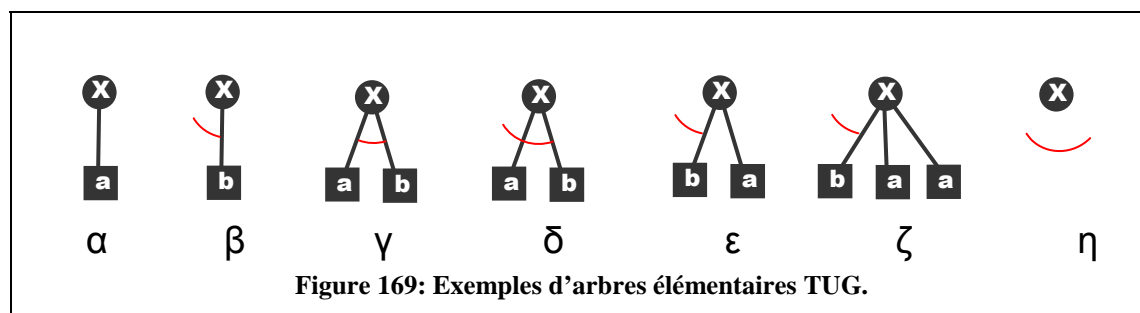


Figure 169: Exemples d'arbres élémentaires TUG.

<sup>211</sup> D vient de *domination*

Deux arbres élémentaires  $\kappa$  et  $\lambda$  peuvent *s'unifier* en un seul arbre si et seulement si la racine de l'atome  $N$  d'un des deux arbres, par exemple de l'arbre  $\kappa$ , s'unifie avec un quelconque atome  $M$  de l'arbre  $\lambda$ , et les exclusions d'adjonction de  $N$  et  $M$  permettent d'ordonner leurs enfants. Si un atome  $M'$ , enfant de  $M$ , est unifié avec  $N'$ , enfant de  $N$ , leurs enfants respectifs doivent ensuite être capables de trouver un arrangement dans leur ordre, et ainsi de suite pour leurs descendants. Nous pouvons donc voir que l'unification des arbres  $\alpha$  et  $\beta$  de la Figure 169 donnera comme résultat l'arbre  $\varepsilon$ . De la même manière, l'unification de  $\alpha$  et  $\varepsilon$  peut donner  $\varepsilon$  (si les traits des nœuds peuvent être unifiés).

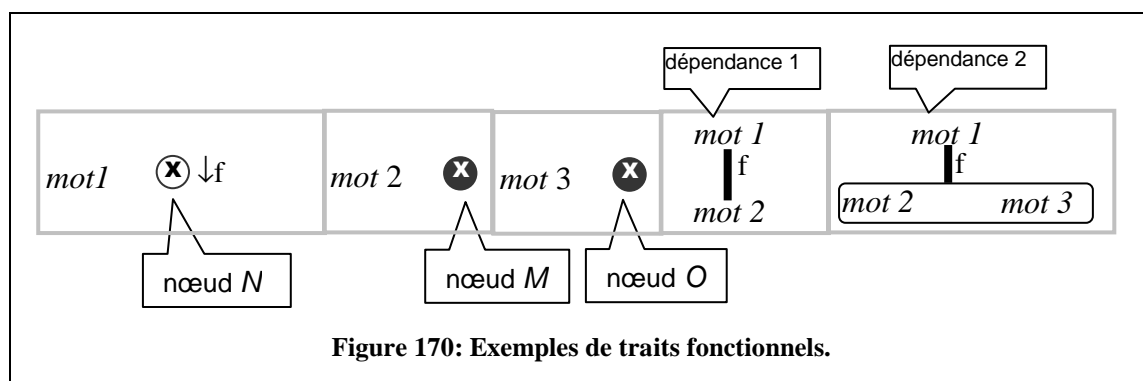
Enfin, les deux atomes  $a$  de  $\alpha$  et de  $\varepsilon$  n'ont pas à se combiner nécessairement ; ils peuvent aussi résulter en une structure comme  $\zeta$  puisqu'il n'y a pas d'exclusion d'adjonction sœur pour l'atome  $a$  de  $\varepsilon$ . De la même manière, il est possible de bloquer un nœud et lui interdire d'avoir des enfants comme c'est le cas pour l'arbre  $\eta$ .

Une *entrée lexicale* dans TUG consiste d'un mot  $w \in W$  et d'un ensemble d'arbres élémentaires dont l'un d'entre eux est lexicalisé par le mot  $w$ . Le nœud feuille qui porte le mot  $w$  est appelé *tête lexicale*.

Dans la TAG habituelle, l'arbre de dérivation indique juste quel nœud a été substitué à, ou adjoint sur quel autre nœud (et l'adresse de Gorn de cette adjonction). Le cas de TUG est moins simple car l'unification d'atomes n'est pas directionnelle, ni unique, ni locale dans le sens qu'un arbre élémentaire peut s'unifier avec des branches qui « appartiennent » à différents arbres élémentaires. Ceci implique que la connexion entre le niveau topologique – construit par ces arbres élémentaires – et l'arbre de dépendance dérivée doit être encodée de manière spécifique.

C'est la raison pour laquelle nous introduisons dans TUG un trait spécial, que nous appelons *trait fonctionnel*, marqué par un trait et par le nom du lien de la dépendance à construire :

Soit  $\downarrow f$  un trait fonctionnel d'un atome (vide)  $N$  qui fait partie de l'entrée lexicale *mot1*. Alors, un atome  $M$  qui fait partie de l'entrée lexicale *mot2* peut s'unifier avec  $N$  si et seulement si le lexème *mot1* est connecté à un ensemble qui contient *mot2* par un lien de dépendance appelé  $f$ .



De cette manière, les atomes  $N$  et  $M$  s'unifient si et seulement si nous obtenons la **dépendance 1**. De la même manière, l'unification des trois atomes  $N$ ,  $M$  et  $O$  dans un seul nœud correspond à la **dépendance 2** de la même figure.<sup>212</sup>

L'arbre de dépendance syntaxique de surface est équipé d'une hiérarchie simple et non intersective de groupements communicatifs des mots. Afin d'obtenir les correspondances entre les domaines du niveau topologique et les groupements au niveau syntaxique, nous devons stipuler en plus le fait suivant :

<sup>212</sup> Il est également possible de représenter ce cas de double dépendance par deux liens de dépendance syntaxiques du même nom  $f$ .

Tous les têtes lexicales qui sont directement dominées par un même atome  $D$  se retrouvent groupées ensemble au niveau syntaxique, et ce groupement est étiqueté par la structure de traits de l'atome  $D$ .

Enfin, une dérivation topologique d'une chaîne de mots  $P$  est complète (et  $P$  donc une phrase topologiquement correcte) si et seulement si<sup>213</sup> :

- une structure de champs topologiques sous un nœud  $D$  peut être créée sur  $P$ , en utilisant tous les arbres élémentaires des entrées lexicales en question
- tous les atomes vides sont superposés par des atomes pleins
- les traits fonctionnels ont créé un arbre  $T$  de dépendance syntaxique connexe sur l'ensemble des nœuds.

### 3.2 Où se trouvent les conditions de bonne formation ?

Dans la présentation du formalisme TUG, la correspondance entre la structure topologique et l'arbre de dépendance syntaxique de surface pourrait paraître asymétrique car seule la structure topologique possède des conditions de bonne formation. C'est ainsi que le module syntaxique TUG pourra construire l'arbre de dépendance pour la phrase (259) et la tâche de calculer les conditions de bonne formation revient à l'étape suivante de l'analyse (dans le cas de notre exemple, la phrase sera rejetée et sera considérée agrammaticale à ce niveau de l'analyse). Par contre, dans le cas de l'exemple (260) qui est aussi syntaxiquement déficiente (nous reprenons ici l'exemple donné dans la Figure 167), on n'aura pas d'arbre de dépendance syntaxique correspondant car le syntagme *des Mannes* 'de l'homme' reste non connexe aux autres parties de l'arbre de dépendance syntaxique.

(259) \*Den Roman gelesen haben Peter.

le roman (acc.) lu ont Pierre

'\*Le roman ont lu Pierre.'

(260) \*Den Roman gelesen hat des Mannes.

le roman lu a de l'homme (gén.)

'\*Le roman est lu de l'homme.'

Il serait possible de rendre le module symétrique en ajoutant aux règles de bonne formation, la condition d'une dérivation correcte. Cependant, nous avons décidé de ne pas le faire pour les raisons suivantes. Dans le Modèle Sens-Texte où l'on veut intégrer le module proposé, il existe une chaîne de représentations (R) munies de conditions de bonne formation, suivies de modules de transfert (T), puis de représentations (R) munies de conditions de bonne formation et ainsi de suite. Si nous voulons découper la chaîne RTRTRTRTR... en des parts égales, nous choisirions RT ou TR, et toutes les deux sont asymétriques. Supposons en plus que nous commençons le processus de génération de langage par une représentation conceptuelle bien formée. Dans ce cas, nous devons commencer le module de transfert et ajouter un module de bonne formation pour la représentation qui suit. Dans la direction de l'analyse, ceci donnerait un module qui construit une représentation bien formée et qui transfère cette représentation en une représentation plus profonde, tout en laissant les règles de bonne formation au niveau suivant. Et c'est ainsi que nous avons choisi pour TUG de mettre les règles de bonne formation exclusivement sur la structure topologique.

---

<sup>213</sup> Ceci est équivalent à : une linéarisation topologique d'un arbre de dépendance syntaxique de surface  $T$  est complète si et seulement si...

### 3.3 Une grammaire jouet pour l'allemand

#### 3.3.1 Description de la grammaire

A présent, nous allons présenter une TUG simple avec les entrées lexicales de la phrase (261) nous permettant d'obtenir tous les ordres des mots possibles.

- (261) Dem Lehrer zu lesen versprochen hat den Roman niemand.  
 au professeur (dat.) de lire promis a le roman (acc.) personne.  
 'Personne n'a promis au professeur de lire le livre.'

L'auxiliaire *hat* 'a' a deux entrées lexicales différentes selon que son dépendant ouvre ou non un nouveau domaine, ou autrement dit, selon qu'il est construit de manière cohérente ou incohérente dans la terminologie de Bech 1955. Dans la première entrée, l'auxiliaire ouvre un domaine principal et le verbe qui occupe seul la parenthèse gauche de ce domaine est sa tête lexicale.

Nous représentons les entrées pour les noms comme des arbres atomiques puisque la structure interne des syntagmes nominaux n'est pas traitée dans le cadre de TUG. La Figure 171 illustre les différentes entrées lexicales de notre TUG de l'allemand.

Décrivons les champs apparaissant dans l'arbre de *hat* un par un. L'atome vide du Vorfeld indique que cette place doit être occupée par un élément quelconque mais qu'on n'a encore aucune information sur les propriétés de l'objet qui doit être placé. Les contraintes d'adjonction sur la parenthèse gauche font que le champ de parenthèse gauche ne peut être occupé que par l'auxiliaire. Le Mittelfeld est vide et n'a aucune restriction sur l'adjonction. La parenthèse droite peut prendre le dépendant verbal de l'auxiliaire, c'est-à-dire le participe passé. La contrainte de l'adjonction par la droite qui s'applique à la parenthèse droite oblige le participe passé à occuper la position la plus à droite du champ. Le trait fonctionnel  $\downarrow$ pastp va créer un lien de dépendance dans la structure syntaxique entre le mot *hat* 'a' et le participe passé qui prendra cette place de parenthèse droite. Seules les éléments ayant un trait *forme* (indiquant la forme du verbe) avec la valeur *pastp* pourront s'unifier avec ce nœud. Le Nachfeld est, comme le Mittelfeld, c'est-à-dire accessible à toute sorte d'unification. L'entrée lexicale de *hat* 'a' comporte encore un autre arbre élémentaire qui n'est pas connecté à la tête lexicale : il s'agit de l'arbre élémentaire qui attend le sujet<sup>214</sup>. Le nœud racine de cet arbre est marqué par *xf*, qui indique la disjonction de Vorfeld, Mittelfeld ou Nachfeld et cet arbre ouvre donc une place pour le sujet dans l'un de ces champs majeurs d'un domaine quelconque.

---

<sup>214</sup> Voir note 210 pour une discussion expliquant pourquoi cet arbre élémentaire doit être considéré comme optionnel pour la forme de *hat* 'a'.

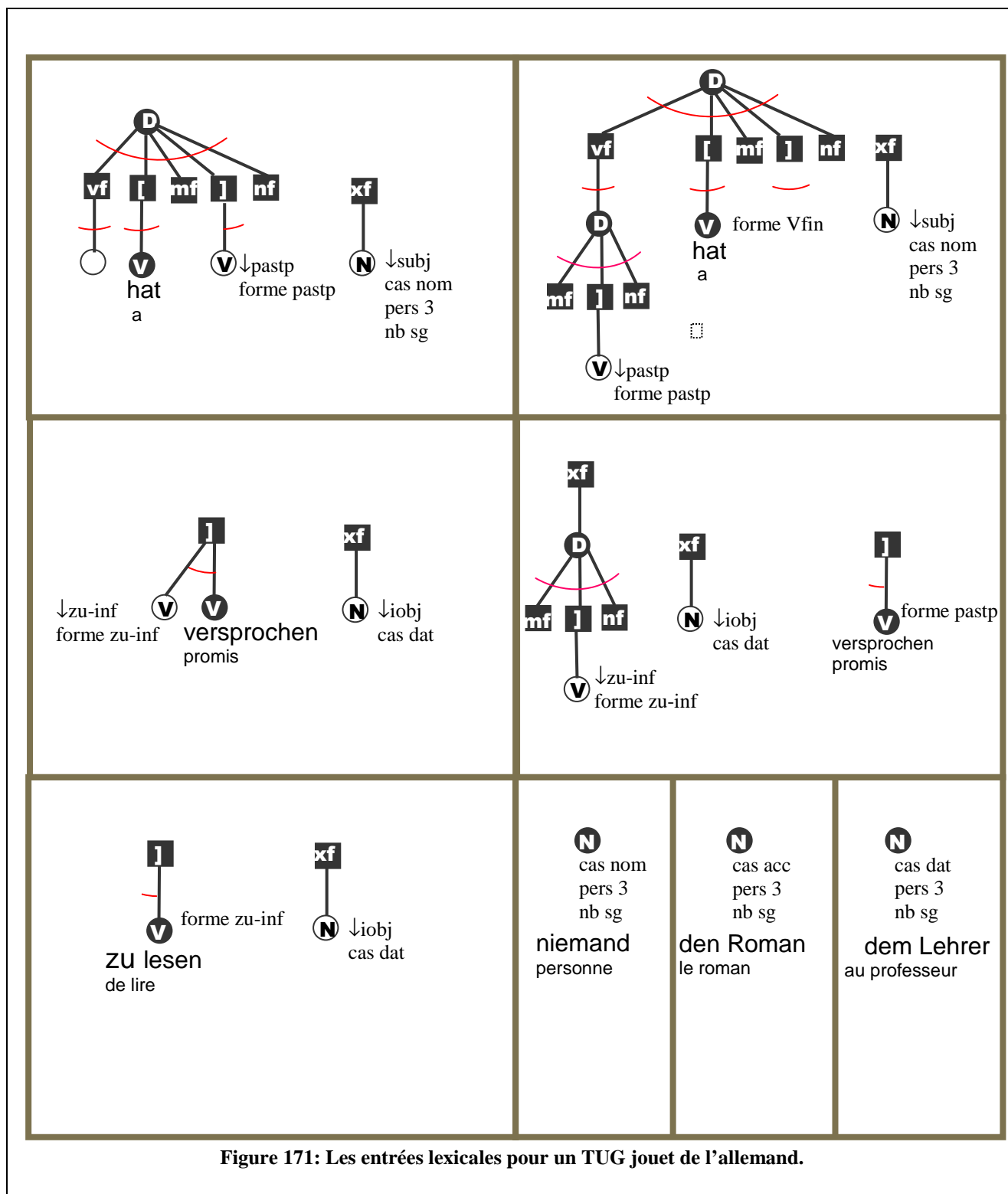


Figure 171: Les entrées lexicales pour un TUG jouet de l'allemand.

On applique la deuxième entrée lexicale possible lorsque le participe passé ouvre son propre domaine. Comme ce dernier ne peut ouvrir de domaine enchâssé que dans le Vorfeld, nous pouvons déjà prévoir la structure de ce domaine enchâssé qui assigne une place au participe passé. Dans ce cas, la parenthèse droite est complètement bloquée car seul un dépendant verbal direct de *hat* 'a' peut occuper cette place, les dépendants verbaux indirects devant rejoindre le domaine de leur mère ou ouvrir leur propre domaine.

Le participe passé *versprochen* 'promis' a encore deux entrées lexicales possibles, la première pour le cas où son dépendant verbal se place dans la parenthèse droite, et la deuxième pour le cas où son dépendant crée son propre domaine. Comme *versprochen* 'promis' est un verbe non fini, il se place toujours dans la parenthèse droite. Dans le premier cas, son dépendant verbal (ici, l'infinitif en *zu lesen* 'de lire') doit se joindre à lui dans la parenthèse droite tandis que dans le deuxième cas, il doit être l'élément le plus à gauche de la parenthèse droite et son dépendant verbal aura son propre domaine ailleurs.

Enfin, *zu lesen* 'de lire' n'a qu'une seule entrée lexicale car il ne gouverne aucun élément verbal qui pourrait ouvrir un nouveau domaine.

### 3.3.2 L'analyse avec TUG

Voyons maintenant comment une phrase peut être analysée en utilisant une TUG telle que nous l'avons définie dans les sous-sections précédentes. Les flèches<sup>215</sup> de la Figure 173 indiquent comment les atomes des entrées lexicales concernées s'unifient pour former un arbre topologique connecté. Cela donne comme résultat la correspondance désirée que nous obtenons dans la Figure 172. Les carrés dans la dépendance de la Figure 172 indiquent les groupements d'éléments correspondant aux domaines dans la structure topologique ; les ovales représentent une simplification de la structure de dépendance du syntagme nominal qui ne joue pas un rôle important ici (cf. IV.6).

TUG peut bien sûr analyser des phrases qui sont syntaxiquement et sémantiquement défectueuses car elle n'est concernée que par la bonne formation topologique par le transfert vers la structure syntaxique de surface.

Inversement, TUG permet de générer tous les ordres des mots de phrases grammaticales pour un arbre de dépendance syntaxique de surface donné, qui lui-même correspond à un graphe sémantique bien formé. Par contre, si l'arbre de dépendance syntaxique de surface n'est pas bien formé ou provient d'un graphe sémantique défectueux, les phrases qui résultent peuvent être agrammaticales car TUG n'a pas de critères lui permettant de tester la bonne formation de la représentation syntaxique.

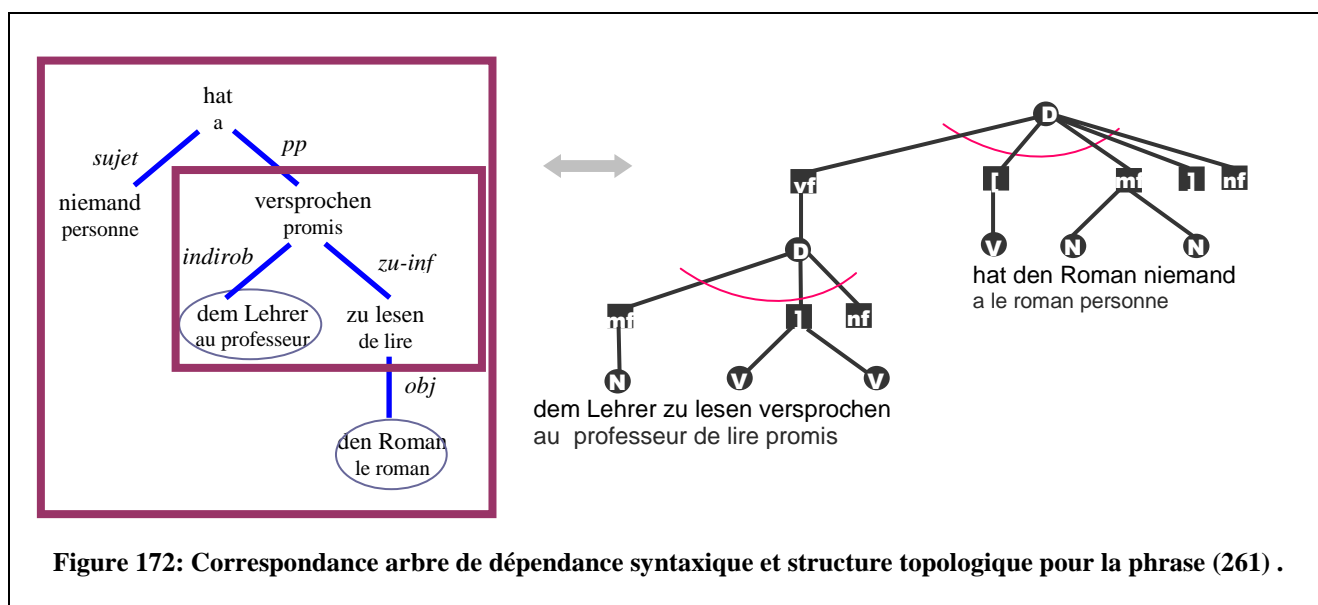
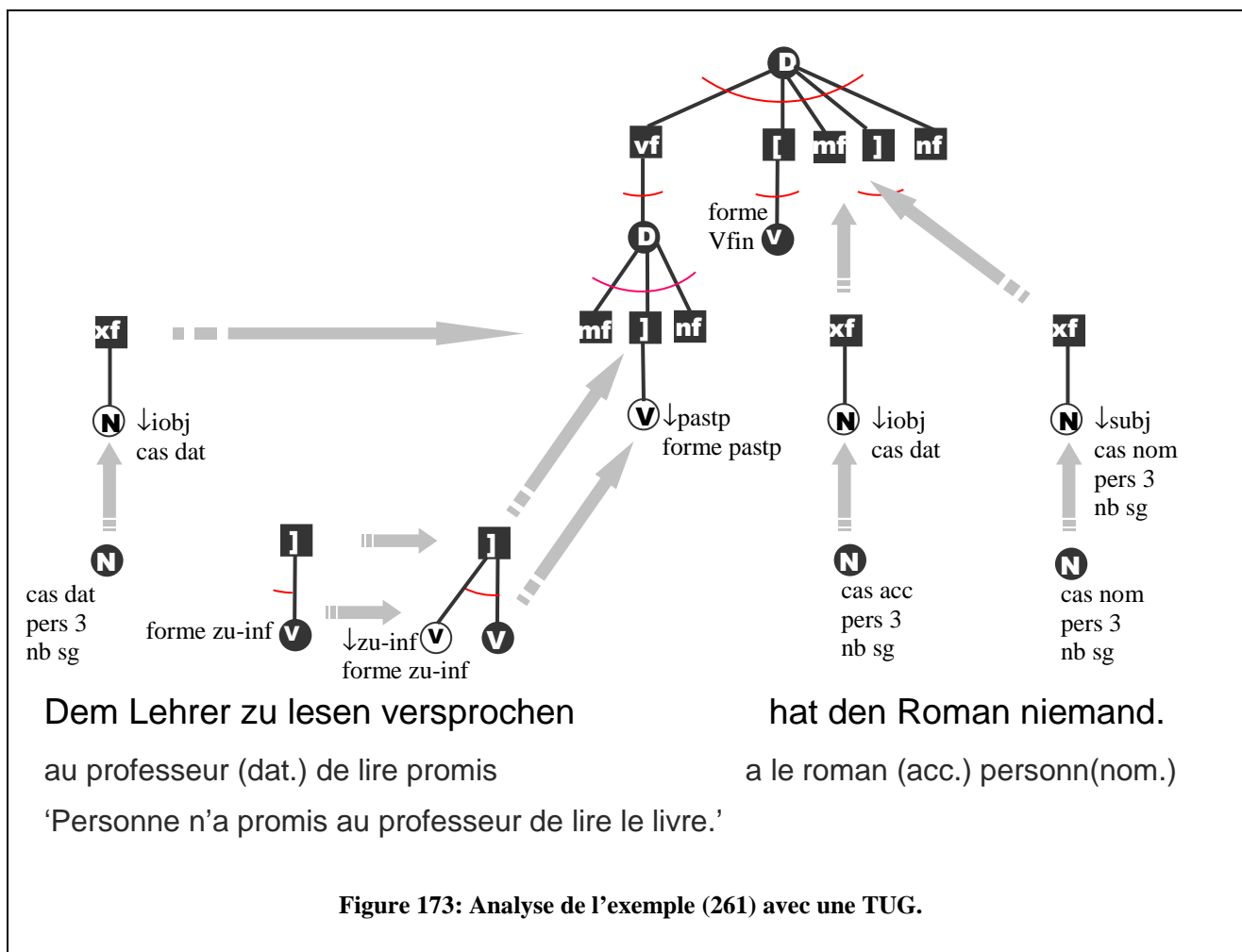


Figure 172: Correspondance arbre de dépendance syntaxique et structure topologique pour la phrase (261) .





Ainsi, TUG est un formalisme qui permet de dériver en parallèle la structure topologique et son arbre de dépendance syntaxique de surface correspondant. Elle constitue un modèle de correspondance qui peut s'intégrer dans le Modèle Sens-Texte. En fait, malgré le grand nombre de possibilité d'ordre pour une sous-catégorisation donnée, les mots-formes nécessitent très peu d'entrées lexicales différentes, ce qui représente un grand avantage du formalisme que nous présentons par rapport aux TAG.

Notre but était de présenter une TUG jouet de l'allemand afin d'illustrer le formalisme sous-jacent. Nous n'avons donc pas traité des phénomènes syntaxiques plus complexes comme le pied-piping verbal ou les variations d'ordre dans le complexe verbal, mais à l'instar du module topologique non lexicalisé, leur description est très simple et élégant dans une TUG.

Pour l'instant, TUG n'est pas implémenté mais les travaux sont en cours pour développer l'algorithme d'analyse. Par ailleurs, bien que le nombre d'arbres élémentaires d'une TUG est beaucoup moins élevé que pour TAG, le fait d'être lexicalisé rend intéressant sa gestion par un outil de type métagrammaire. Nous nous intéressons à la traduction des hiérarchies proposées pour les TAG au formalisme TUG, mais cela ne va pas de soi car beaucoup de classes de la métagrammaire ont été introduites justement pour capturer les phénomènes d'ordres des mots qui sont présentés d'une manière beaucoup plus compacte dans TUG.

Enfin, à mon avis, le processus en soi du développement de TUG est d'une importance fondamentale.

Je considère que la première étape dans la compréhension des langages naturels doit être constituée par l'observation et la description des phénomènes observés. Ce n'est qu'à partir de

<sup>215</sup> Bien sûr l'unification des atomes n'est pas directionnelle. Les flèches essaient d'indiquer la direction de la dépendance et n'ont aucune signification algorithmique.

cette étape qu'on devrait commencer la formalisation. En effet, la formalisation peut avoir un effet sur comment on voit la langue mais ne pourra pas la déterminer. Beaucoup de formalismes, comme d'ailleurs les TAG, fonctionnent de la manière inverse : ainsi, les TAG représentent d'abord des descriptions mathématiques simples et intéressantes qui ont été pensées dans un deuxième temps comme utiles pour les langues naturelles. Le résultat est qu'on se retrouve souvent avec un système qui fonctionne bien pour des phénomènes de base mais qu'il est difficile d'améliorer à cause de son manque de puissance expressive pour les descriptions linguistiques. Contrairement à une telle approche, TUG est construite afin d'obtenir précisément les structures désirées qui sont nécessaires pour notre description des phénomènes d'ordre des mots en allemand.

Après avoir étudié comment le modèle topologique peut nous aider à construire un formalisme d'arbre lexicalisé de type TAG, nous allons voir un tout autre formalisme sous un point de vue topologique, le formalisme de Grammaire lexicale fonctionnelle.

## 4 LFG

Dans cette section, nous verrons comment la structure topologique peut s'intégrer dans le formalisme de Grammaire lexicale fonctionnelle (angl. *Lexical Functional Grammar* ou LFG) et donner une grammaire d'une grande simplicité qui peut traiter des phénomènes syntaxiques souvent considérés comme très complexes.

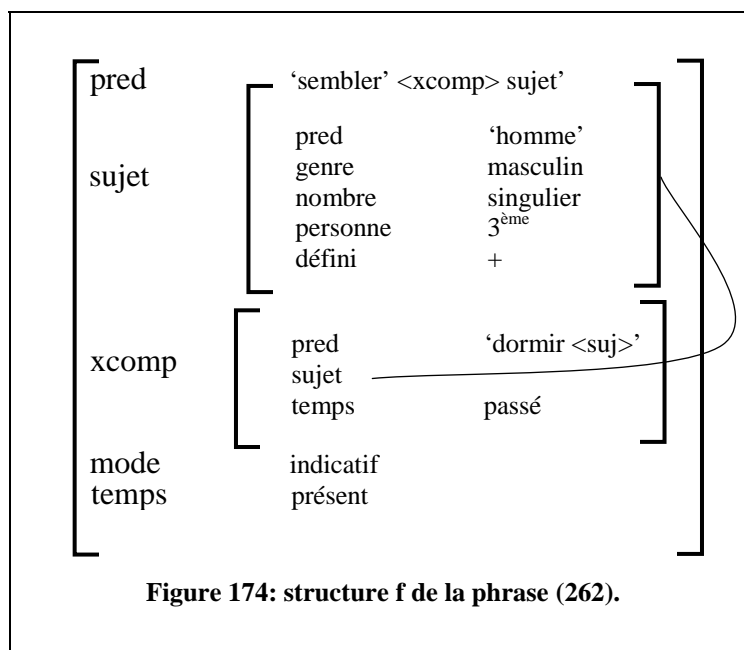
La grammaire lexicale fonctionnelle a été proposée à la fin des années 70 par Bresnan & Kaplan (Bresnan 1982, voir aussi Abeillé 1993) comme une alternative à la grammaire transformationnelle de Chomsky. Leur but était de présenter une grammaire ayant une capacité générative plus restreinte au vu d'un traitement automatique plus efficace que ce que permet le modèle transformationnel. D'un point de vue linguistique, Bresnan et Kaplan veulent « réhabiliter les descriptions de surface et remplacer le composant transformationnel par un niveau 'lexical-fonctionnel' : la structure de la phrase est décrite non seulement par des représentations arborescentes mais aussi par des structures de traits codant directement les différentes fonctions grammaticales, qui sont conçues comme des primitives » (Abeillé 1993, pages 42, 43). Le lien que propose LFG entre un niveau de constituants et un niveau de fonctions grammaticales, basé sur un système de traits non ordonné, semble similaire au lien entre topologie et dépendance syntaxique que propose notre modèle topologique. Dans cette section, nous allons voir qu'il est en fait possible de décrire le modèle topologique dans le formalisme LFG, à condition de revoir des interprétations (et conditions) habituelles des deux structures qui sont mises en relation. L'interprétation originale des deux structures est plus proche de notre propre interprétation que le développement tardif qu'ont pris les descriptions linguistiques faites avec l'outil LFG : d'une part la structure de constituants de LFG s'approche dans quelques travaux d'une structure de surface transformationnelle (R. Kaplan, communication personnelle ; cf. par exemple Choi 1999 et l'intégration en LFG de la notion de « tête fonctionnelle » de Bresnan 2001) et d'autre part nous verrons dans la suite que l'interprétation de la structure des fonctions grammaticales a acquies un caractère plus sémantique qu'à l'introduction du formalisme. Notre analyse sera donc beaucoup plus proche des premières analyses en LFG que des travaux actuels.

### 4.1 Quelques ajustements de notation

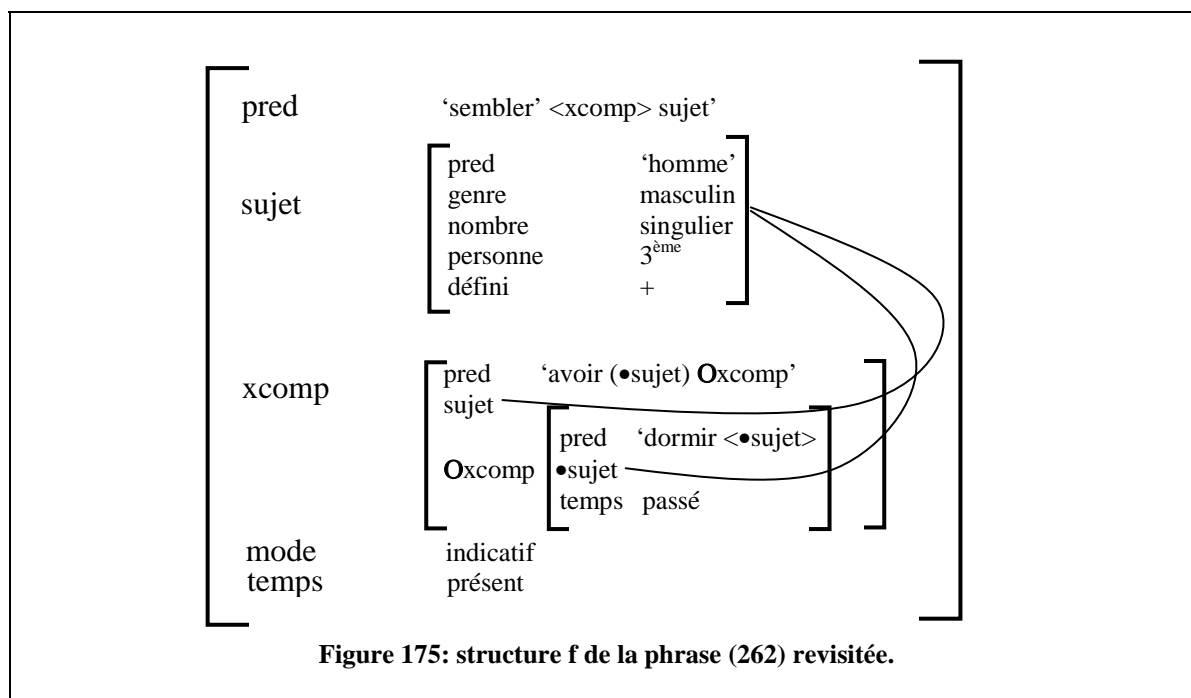
Dans le cadre de ce travail, qui a été mené en collaboration avec Lionel Clément et Sylvain Kahane (Clément, Gerdes, Kahane 2002), nous avons été amenés à reconsidérer la notation traditionnelle utilisée dans la LFG sur deux points qui, à notre avis, posent problème pour une représentation correcte des phénomènes. Le premier concerne la distinction des actants

sémantiques et des arguments syntaxiques (cf. section I.2 pour une discussion détaillée sur la nécessité de séparer ces deux niveaux) et le deuxième concerne la représentation de l'auxiliaire. La structure fonctionnelle a souvent été considérée comme étroitement liée aux arbres classique de dépendance syntaxique, tels qu'ils apparaissent chez Tesnière 1959 ou Mel'čuk 1988. Pourtant la structure f n'est généralement pas un arbre puisqu'elle renvoie à une « structure de traits minimale qui satisfait la description fonctionnelle de l'arbre syntagmatique » (Abeillé 1993:46) et qu'elle contient des informations de différents types, notamment de type syntaxique et sémantique. La Figure 174 suivante illustre la structure f de la phrase (262).

(262) L'homme semble avoir dormi.



Dans sa forme traditionnelle, le trait *pred(icat)* fait la distinction entre les actants sémantiques (qui portent les theta-rôles) et les arguments purement syntaxiques en notant les premiers à l'intérieur des parenthèses des prédicats, et les deuxièmes à l'extérieur. Par contre, il n'y a pas de notation particulière pour un actant qui n'est motivé que pour raisons sémantiques. Dans notre exemple, le statut *sujet* de l'infinitif n'est pas syntaxique mais est encore une relation de type plus profond. Nous avons choisi de coder explicitement cette différence en marquant les actants qui sont exclusivement sémantiques par un point, noté juste avant la fonction grammaticale.



Un. deuxième point problématique concerne les éléments qui n'apparaissent qu'au niveau syntaxique et qui n'ont pas de correspondant au niveau sémantique. On peut illustrer ce cas avec le statut de l'auxiliaire. Dans l'analyse habituelle dans une LFG, l'auxiliaire n'a pas de statut propre puisqu'on ne lui assigne pas de fonction prédicative indépendante<sup>216</sup>. La tâche de l'auxiliaire est réduite à vérifier les traits d'accord en genre, nombre et personne avec le sujet tout en ajoutant des morphèmes grammaticaux à la structure des traits du verbe principal. Or un tel traitement de l'auxiliaire semble contraignant pour une analyse satisfaisante des phénomènes linguistiques. En effet, nous sommes d'accord qu'au niveau sémantique, il faut analyser les formes verbales composées (« a dormi, aura dormi,... ») au même titre que les formes simples (« dort, dormira, dort,... »). Par contre au niveau syntaxique, nous considérons que les auxiliaires se comportent comme un verbe plein, que l'on doit décrire avec sa propre sous-catégorisation verbale.

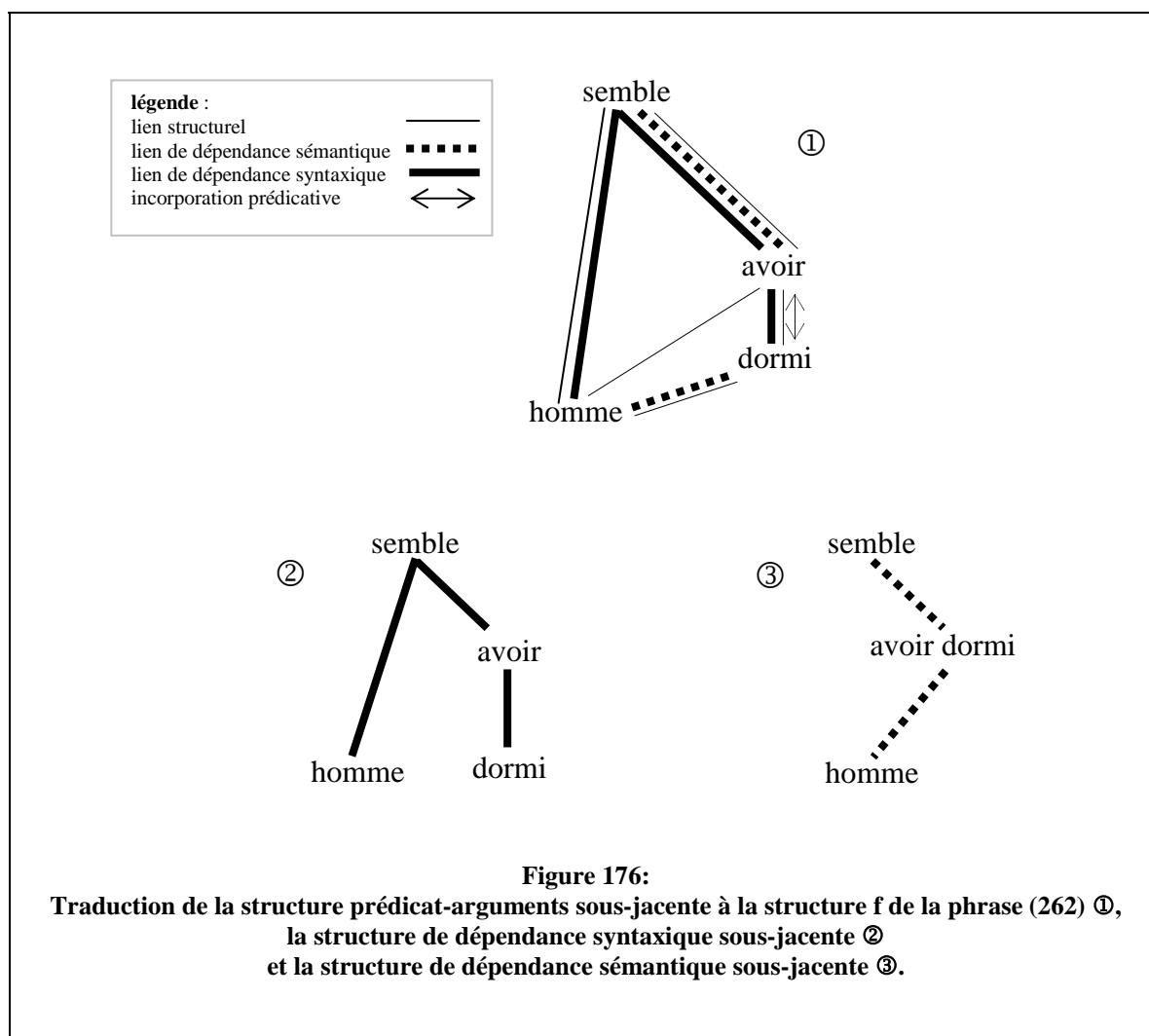
Prenons comme exemple le verbe composé *geschlafen haben* « avoir dormi » de l'allemand. Les réalisations d'ordre des mots de cette forme verbale composée sont identiques que la construction du verbe à contrôle *schlafen wollen* « vouloir dormir ». Si l'on veut capturer le comportement en commun qu'ont ces deux verbes, nous devons introduire les auxiliaires comme des prédicats, au même titre que les verbes à contrôle ou à montée. La prédication est indiquée comme étant purement syntaxique, sans sous-catégorisation sémantique, d'où la non utilisation des parenthèses sémantiques. Néanmoins, il est pourtant possible de garder la structure sémantique en indiquant dans la structure f que l'auxiliaire forme un nœud sémantique unique avec le participe passé. C'est pourquoi nous avons choisi de donner une notation spéciale à la relation *xcomp* entre l'auxiliaire et son participe: cette dernière est notée par le symbole **O**, indiquant que le prédicat qui porte la structure 'cerclée' et le prédicat de la structure 'cerclée' forment une seule unité sémantique. Nous voulons souligner le fait que ces modifications au niveau de la notation sont complètement compatibles avec le formalisme en soi et qu'en plus, elles renforcent l'analyse traditionnelle en LFG. Par contre, la puissance de ces nouvelles

<sup>216</sup> Notons que dans Bresnan 1982, les auxiliaires avaient encore une fonction prédicative mais ils ne l'ont pas préservé dans le développement de LFG par la suite. Ceci montre que la structure a acquis progressivement un caractère beaucoup plus sémantique.

notations se trouve dans sa capacité à distinguer d'une manière explicite le niveau sémantique du niveau syntaxique.

Dans la Figure 175, nous redonnons la structure f de la phrase (262) en utilisant notre propre notation : ainsi, le sujet de « dormir » qui est juste un actant sémantique est noté •sujet. Pour l'auxiliaire, une fonction *xcomp* (précédée par le symbole **O**) est introduite.

Ainsi, la structure prédicat-argument sous-jacente à la structure f (① de la Figure 176) se trouve partagée en deux structures différentes, une structure de dépendance syntaxique (② de la Figure 176) et une structure de dépendance sémantique (③ de la Figure 176). Nous notons les structures argument-prédicat de la structure f, en marquant les liens sémantiques par une ligne pointillée et les liens syntaxiques par une ligne en caractères gras. La flèche indique la fonction grammaticale qui est incorporée dans la structure avec un autre prédicat dans la construction du graphe sémantique. Notons par ailleurs que le lien du sujet entre l'auxiliaire et « l'homme » n'est ni syntaxique ni sémantique, et ne fait qu'assurer le passage de la valeur du sujet syntaxique de « sembler » au sujet sémantique du verbe « dormir » et ce lien n'apparaît donc dans aucune des deux structures sous-jacentes.



## 4.2 Formalisation

Nous présentons dans cette section une formalisation de l'analyse topologique dans le cadre d'une grammaire lexicale fonctionnelle. Comme nous l'avons vu dans les chapitres précédents,

le modèle topologique distingue les constituants des champs : les champs ont comme rôle de proposer un « gabarit phrastique » où les différents constituants peuvent se placer, évitant ainsi d'encoder les différents ordres possibles pour une structure de dépendance donnée, par de nombreuses règles de réécriture. Les constituants sont considérés comme des boîtes qui peuvent s'enchâsser, les champs formant ainsi les « compartiments » de ces boîtes.

Dans notre grammaire, cette dichotomie topologique est reflétée par deux types d'étiquettes, d'une part les constituants qui se comportent de la même manière que les constituants traditionnels, et d'autre part les champs, qui laissent passer automatiquement tout type d'information fonctionnelle. Par ailleurs, les champs ne sont pas soumis à la contrainte qui force la structure  $f$  associée à un constituant à porter le trait *pred*. Nous adoptons la notation du parseur XLFG développée par Clément (Clément & Kinyon 2001) et nous faisons précéder les champs par un trait de soulignement «   ». Nous présentons notre grammaire dans la Figure 177 :

<b>MD</b>	→	<b>_VF_LB (_MF) (RB) (_NF)</b>
<b>ED</b>	→	<b>(_MF) _RB (_NF)</b>
<b>_VF</b>	→	<b>ED</b> (↑ <i>xcomp</i> ) = ↓
<b>_XF</b>	→	<b>ADVP</b> (↑ <i>xcomp* adjunct</i> ) ∃ ↓
<b>_LB</b>	→	<b>V</b> ↑ = ↓
<b>_MF</b>	→	<b>_XF*</b>
<b>_RB</b>	→	<b>VC</b> (↑ <i>xcomp</i> ) = ↓
<b>_NF</b>	→	<b>_XF*</b>
<b>_XF</b>	→	<b>ED</b> (↑ <i>xcomp</i> <sup>+</sup> ) = ↓
<b>_XF</b>	→	<b>NP</b> (↑ <i>sujet</i> ) = ↓
<b>_XF</b>	→	<b>NP</b> (↑ <i>xcomp* {obj, objindirect}</i> ) = ↓
<b>VC</b>	→	<b>(_O) _H (_U)</b>
<b>VB</b>	→	<b>(_O) _H</b>
<b>_H</b>	→	<b>V</b> ↑ = ↓
<b>_O</b>	→	<b>VB</b> (↑ <i>xcomp</i> <sup>+</sup> ) = ↓ <sup>217</sup>
<b>_U</b>	→	<b>VC</b> (↑ <i>xcomp</i> ) = ↓ (↑ <i>type</i> ) = <i>auxiliaire</i> (↓ <i>mode</i> ) = + (↓ <i>ersatz</i> ) = + (↓ <i>xcomp</i> <sup>+</sup> ) ≠ <b>infinitif en zu</b>

Figure 177: Règles de la structure des constituants pour l'allemand<sup>218</sup>

Il s'agit d'une grammaire simplifiée : elle ne couvre pour l'instant qu'une part des phénomènes d'ordre de l'allemand, laissant de côté notamment la structure interne des syntagmes nominaux, les phrases relatives et les complémenteurs.

Le domaine principal (MD de l'anglais *main domain*) constitue le symbole initial de cette grammaire. Le domaine enchâssé (ED de l'anglais *embedded domain*) prend la place des syntagmes verbaux dans le cas des infinitifs et des participes passés. Les complexes verbaux (VC de l'anglais *verb cluster*) et la boîte verbale (VB de l'anglais *verbal box*) sont les deux autres

<sup>217</sup> Si nous n'ajoutons pas le signe + dans cette ligne, il nous est possible d'exclure le phénomène de *Zwischenstellung*, qui n'est pas accepté par tous les locuteurs natifs de l'allemand (voir section II.3).

<sup>218</sup> Nous supposons que les cas (nominatif, accusatif et datif) sont imposés par des règles de sous-catégorisation dans le lexique, bien qu'il soit possible de les considérer comme des règles grammaticales. Dans ce cas, il faudrait ajouter des règles d'ordre (le nominatif pour le sujet, l'accusatif pour l'objet et le datif pour l'objet indirect). Rappelons cependant que certains verbes peuvent avoir un sujet au génitif ou deux compléments à l'accusatif.

projections possibles du verbe (qui ne peuvent pas prendre des syntagmes nominaux). Les champs qui peuvent rester vides sont considérés comme optionnels. Dans un souci de généraliser le comportement similaire des trois champs majeurs – le *Vorfeld* (\_VF), le *Mittelfeld* (\_MF) et le *Nachfeld* (\_NF) – nous sommes obligés, dans le formalisme LFG, d'introduire une étiquette supplémentaire, appelée \_XF, bien que cette dernière n'ait pas un statut réel de champ.

Les dépendances de longue distance sont prises en compte par « l'incertitude fonctionnelle », un ajout au formalisme de départ effectué par Kaplan & Zaenen 1989 qui permet de représenter les phénomènes d'extraction d'une manière purement fonctionnelle. Nous arrivons ainsi à traiter les phénomènes de topicalisation (le placement dans le *Vorfeld*, incluant le phénomène de *VP-fronting*) et de *brouillage d'arguments* (*scrambling*).

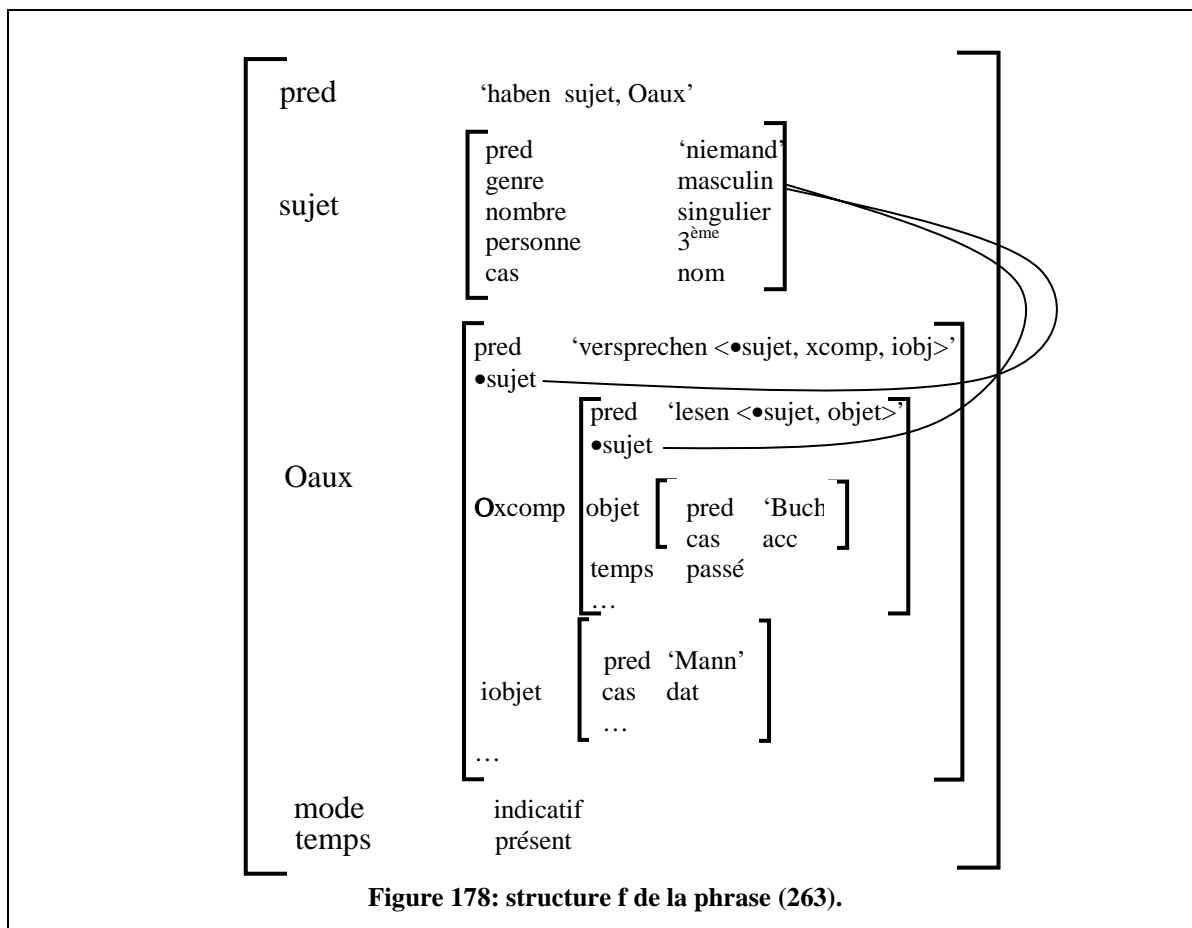
Le placement d'une chaîne subordonnée verbale (une chaîne *xcomp*) dans le domaine topologique suit le principe simple suivant : les verbes sont placés dans l'ordre descendant. Ainsi, la parenthèse gauche (\_LB) est prise en charge en premier, puis dans l'ordre la parenthèse droite (\_RB), ensuite l'*Oberfeld* (\_O) ou l'*Unterfeld* (\_U) à l'intérieur de la parenthèse droite, et ainsi de suite. Dans le cas des domaines enchâssés qui ne comportent pas de parenthèse gauche, le même principe est appliqué sauf que le placement des éléments commence avec celui de la parenthèse droite. On constate qu'aucune structure syntagmatique basée sur un formalisme LFG (ou encore HPSG) ne permet d'exprimer cette généralité et par conséquent, nous avons deux règles en ce qui concerne la parenthèse droite, l'une pour la parenthèse droite du domaine principale où la structure f du verbe prend la position *xcomp* de la structure f principale, l'autre pour la parenthèse droite du domaine enchâssé où le verbe représente la tête du domaine.

Par ailleurs, l'analyse du complexe verbal permet d'obtenir toutes les linéarisations possibles, notamment l'*Oberfeldumstellung* et le *Zwischenstellung*. Il est facile d'exclure la deuxième possibilité, qui n'est acceptée que par un certain nombre de locuteurs de l'allemand, en supprimant le signe « + » dans l'équation fonctionnelle de la boîte verbale (VB) de la règle du champ *Oberfeld* (\_O), lequel autorise une dépendance de longue distance et ainsi l'émancipation hors de la boîte verbale.

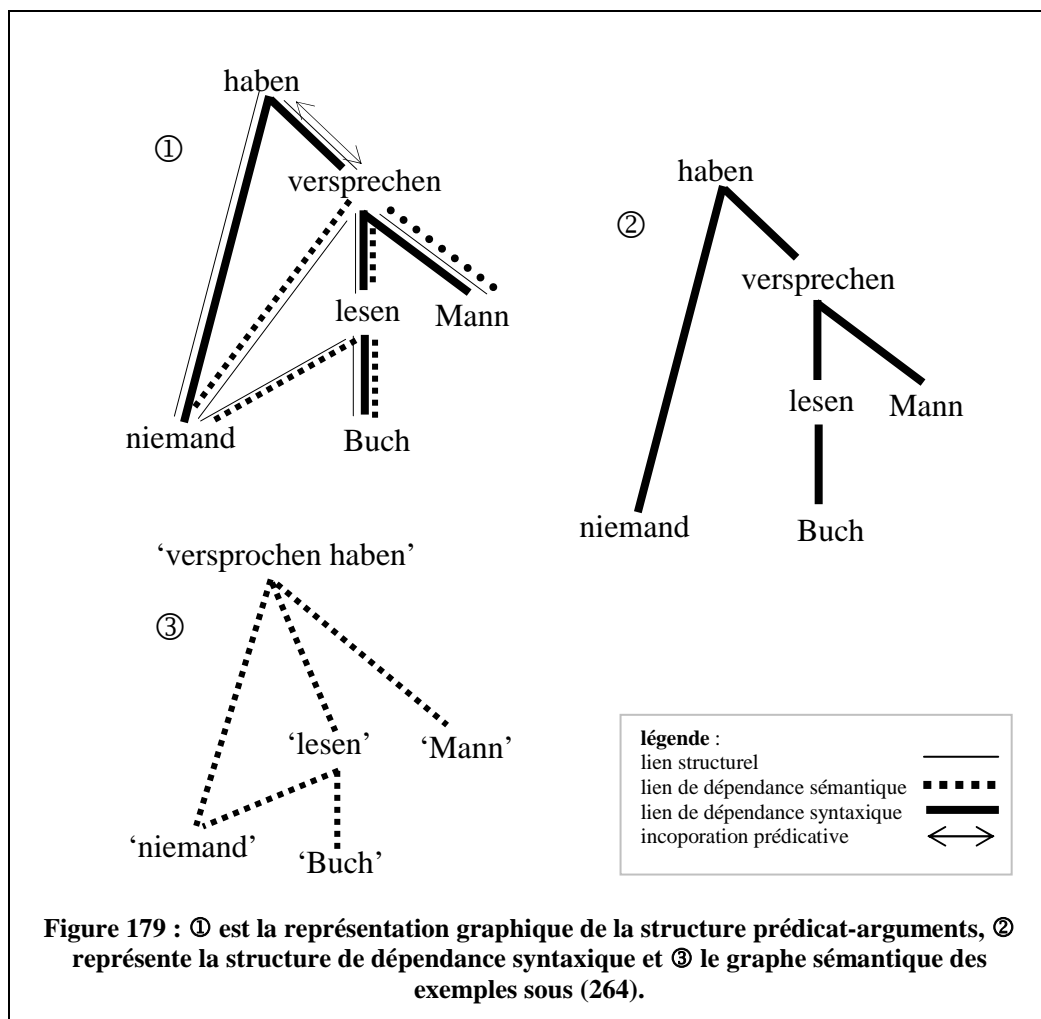
Reprenons l'exemple analysé dans la section I.6.6 que nous reproduisons sous (263)

(263) Niemand hat diesem Mann das Buch zu lesen versprochen.  
 personne<sub>nom</sub> a cet homme<sub>dat</sub> le livre<sub>acc</sub> à lire promis  
 'Personne n'a promis à cet homme de lire le livre.'

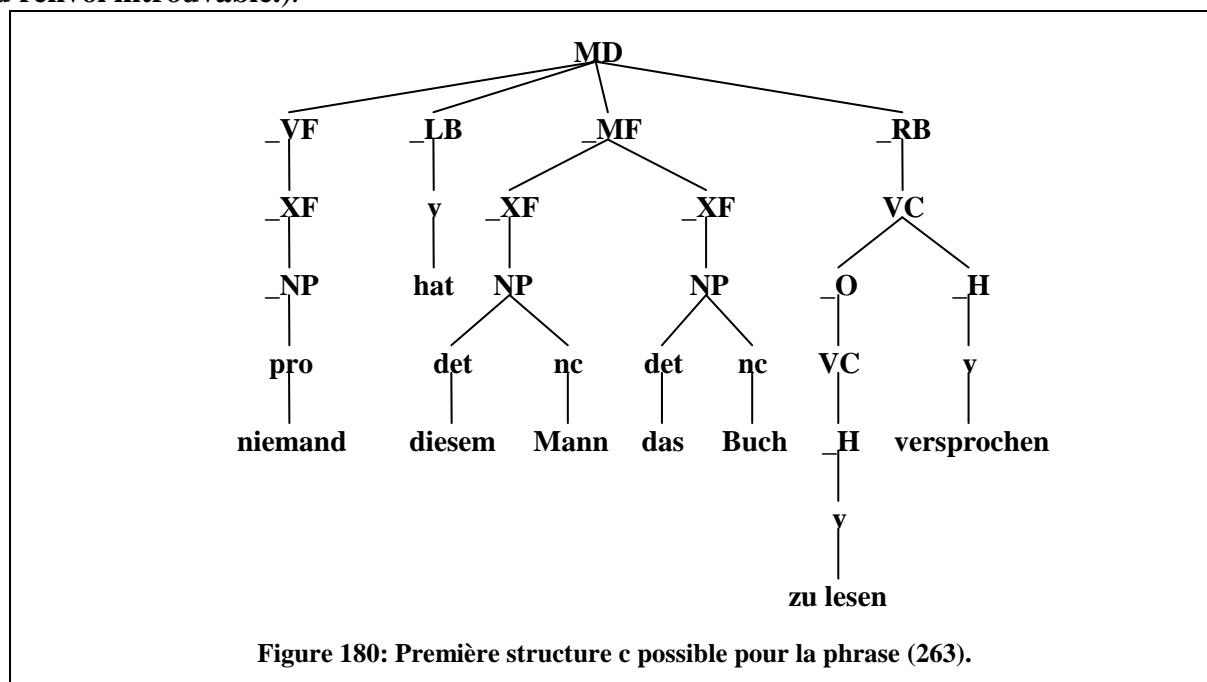


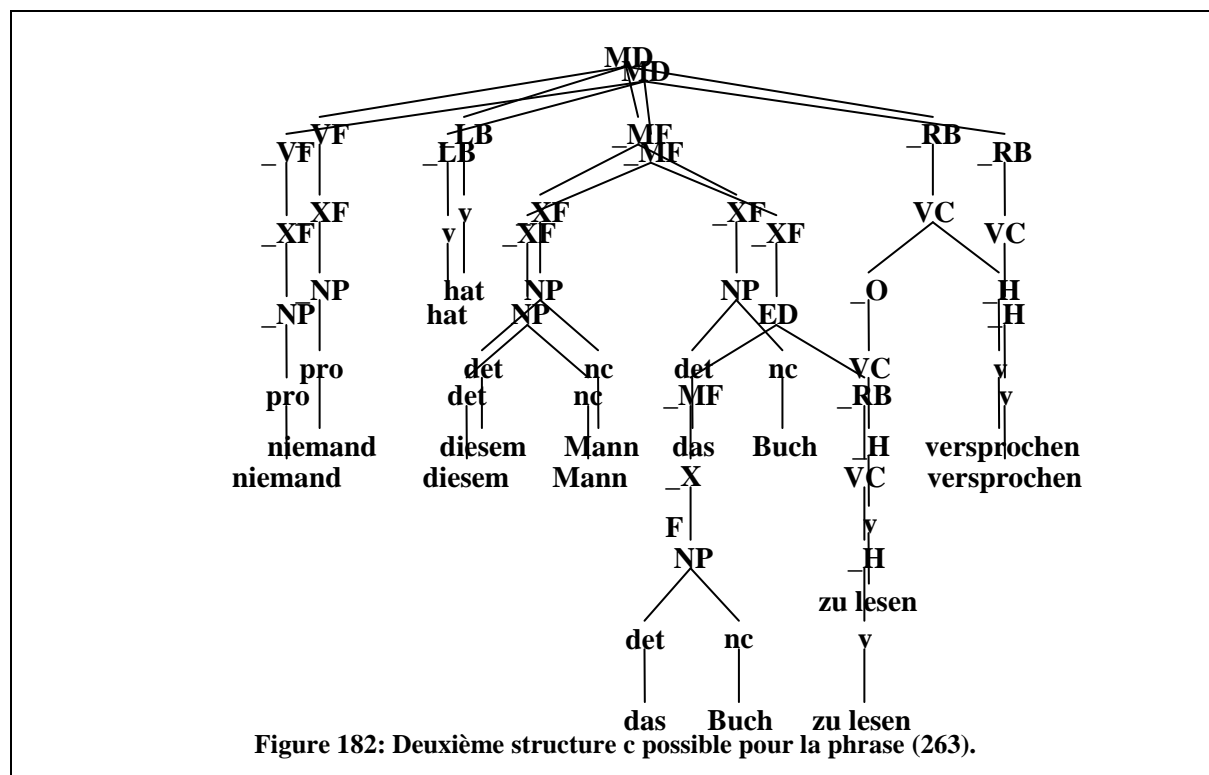


Nous donnons la structure f de la phrase (264) dans la Figure 178 et la représentation graphique des fonctions prédicatives accompagnées de l'arbre de dépendance syntaxique et du graphe sémantique dans la Figure 179.



Les règles que nous avons établies nous permettent d’obtenir la structure f de la phrase (263) (Figure 178) à l’aide de trois structures c différents (Figure 180, Figure 182 et **Erreur ! Source du renvoi introuvable.**).





Dans la première analyse, *zu lesen versprochen* forme un complexe verbal (cas de la phrase (263) ou (264)b.). Ce groupement correspond à la possibilité pour les deux verbes de constituer un seul groupe prosodique, où seul la première syllabe du verbe principal est accentuée. La seconde analyse reflète le fait que les domaines enchâssés peuvent se placer dans le *Mittelfeld* comme dans l'exemple (264)b., où l'analyse des verbes en un seul complexe verbal est impossible. Le domaine enchâssé forme une unité prosodique à part. Quant à la troisième structure c, elle correspond à une possibilité additionnelle où le domaine enchâssé ne contient plus aucun de ses compléments verbaux (phrases (264)c. et d.). Nous illustrons dans les phrases (264), ces différents exemples avec des linéarisations alternatives de (263) :

- (264) a. Diesem Mann hat das Buch niemand zu lesen versprochen.  
 b. Diesem Mann hat, das Buch zu lesen, niemand versprochen.  
 c. Zu lesen hat diesem Mann das Buch niemand versprochen.  
 d. Das Buch hat niemand diesem Mann versprochen zu lesen.

Ainsi, nous avons essayé de montrer dans cette section qu'il était possible de créer une Grammaire Lexicale Fonctionnelle de l'allemand en se basant sur le modèle topologique. Cela revient à dire que dans les analyses LFG habituelles, le formalisme lie la structure f, de caractère syntaxique profonde ou même sémantique et de vocation universelle, à la structure c qui reprend pour beaucoup d'analyses les notions X-barre. Mais ce même formalisme LFG peut aussi, avec quelques aménagements mineurs de l'analyse, servir comme outil pour la description du lien qui existe entre la structure de dépendance syntaxique de surface et la structure topologique. La grammaire que nous avons obtenue a l'avantage de pouvoir traiter un grand nombre de phénomènes linguistiques, tout en étant d'une grande simplicité. Elle est implémentée dans l'analyseur XLFG (version 3.4.3) développée par Clément, mais cette implémentation est encore

au stade du développement puisqu'elle se base sur une grammaire encore partielle et un lexique minimal.

Nous terminons cette section avec une remarque sur le statut théorique des deux structures principales de LFG, que nous avons été amenés à modifier lors du transfert du modèle topologique dans un formalisme LFG. D'abord, la structure *f* peut différencier les dépendances syntaxiques des dépendances sémantiques, comme nous l'avons montré *supra*. Quant à la structure *c*, elle encode l'ordre des mots et les groupements des mots (prosodiques et informationnels). Ainsi, nous obtenons une structure *c* qui est complètement libérée du fardeau fonctionnel, ancien héritage de la syntaxe X-barre.

Il apparaît donc clairement qu'une distinction nette entre les niveaux topologique, syntaxique et sémantique est utile non seulement pour une description linguistique adéquate, mais aussi une économie dans la formalisation. Encore une fois, la structure topologique fait ses preuves en montrant son utilité (voire sa nécessité) mais aussi son adaptabilité dans d'autres formalismes tels que LFG.

## 5 L'ordre dans le Mittelfeld – un phénomène topologique ?

Jusqu'à présent nous sommes partie de l'idée que le module de linéarisation crée tous les ordres possibles pour un arbre de dépendance syntaxique donné, sans prendre en compte l'information communicative. Un ordre est une linéarisation correcte d'un arbre syntaxique quand il existe au moins une structure communicative qui la rend possible. C'est dans ce sens, que toutes les phrases que notre grammaire génère sont grammaticales. C'est ce qui explique aussi pourquoi certaines linéarisations sont difficilement acceptées par certains locuteurs, car la structure communicative qui la requiert est très complexe – mais imaginable dans des contextes rares. Il est envisageable de ne pas considérer la structure communicative au niveau de la linéarisation : le module topologique ne fait que passer l'information communicative présente sur l'arbre de dépendance au module suivant (un module prosodique, par exemple). C'est à ce module que reviendrait la tâche de choisir parmi les linéarisations proposées, celles qui correspondent à la structure communicative donnée, et de leur attribuer la ou les courbes prosodiques correspondantes.

Cependant, rien ne nous empêche d'envisager de prendre en compte l'information communicative dans les règles topologiques. Si c'était le cas, il est clair qu'on aboutirait à beaucoup moins de structures topologiques (cf. I.6), et le module prosodique peut directement procéder à la génération de la courbe mélodique, car chaque linéarisation sortante d'une telle approche sera « prononçable ». La prise en considération de la structure communicative ne concernerait pas les parenthèses (où un élément se place sous conditions entièrement syntaxiques) mais les champs majeurs, en particulier le Mittelfeld.

L'objectif de cette section est d'explorer les possibilités d'organiser l'ordre dans le Mittelfeld en présentant une légère complexification du modèle que nous avons présenté dans le deuxième chapitre. Le débat sur l'ordre dans le Mittelfeld est actif depuis au moins 100 ans et nous n'aspirons ni à présenter toutes les propositions qui ont été faites, ni à avancer une solution globale. Il s'agit juste de montrer que les descriptions d'ordre à l'intérieur du Mittelfeld, quelque soit leur nature, peuvent s'intégrer dans un module de linéarisation d'un arbre de dépendance syntaxique basé sur le modèle topologique.

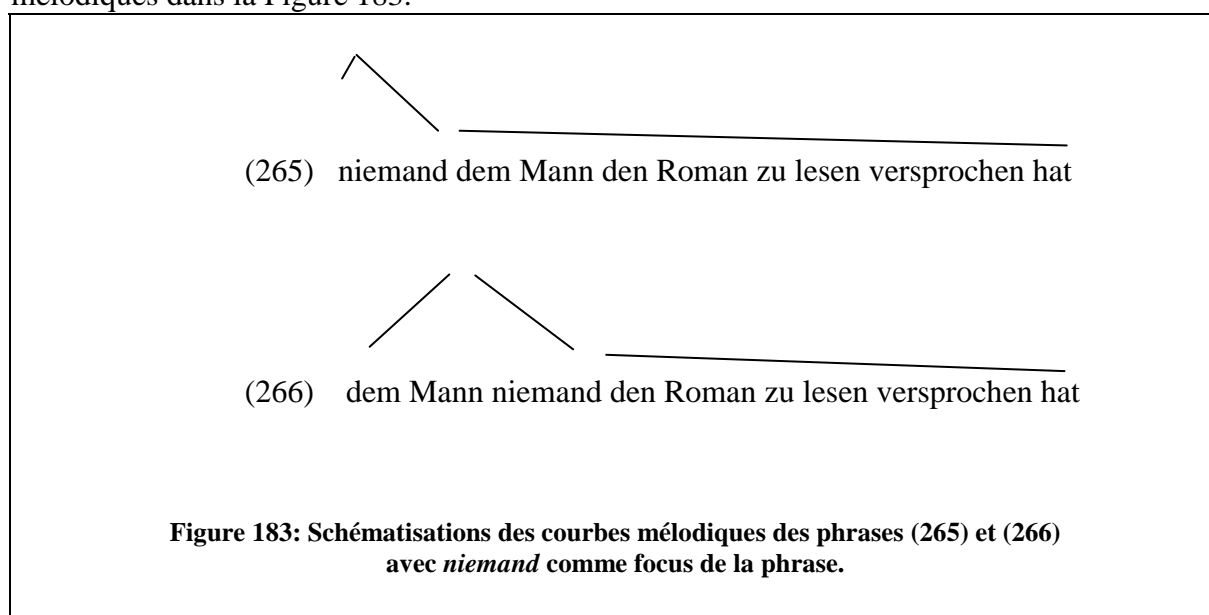
### 5.1 L'organisation communicative et prosodique du Mittelfeld

Dans les premiers chapitres de cette thèse, nous avons déjà vu beaucoup d'exemples où la même structure de dépendance syntaxique donne lieu à différents ordres pour les éléments qui se

placent dans le Mittelfeld. Les phrases (265) et (266) représentent deux linéarisations différentes de la même structure de dépendance.

- (265) (Ich glaube, dass) niemand dem Mann den Roman zu lesen versprochen hat.  
 (je pense, que) personne au homme le roman de lire promis a.  
 '(Je pense que) personne n'a promis à l'homme de lire le roman.'
- (266) (Ich glaube, dass) dem Mann niemand den Roman zu lesen versprochen hat.  
 (je pense, que) au homme personne le roman de lire promis a.  
 '(Je pense que) personne n'a promis à l'homme de lire le roman.'

Si les phrases (265) et (266) partagent le même arbre de dépendance syntaxique, il semble clair qu'elles diffèrent quant à leur structure communicative. Par sa nature, le pronom *niemand* 'personne' ne peut pas renvoyer à quelque chose qui a déjà été introduit dans le discours. Il constitue donc une entité focus (ou éventuellement tail s'il existe une autre entité qui est plus proéminent dans la phrase). Si nous supposons que *niemand* constitue à lui seul le focus de la phrase<sup>219</sup>, alors *dem Mann* 'à l'homme' fait partie du tail dans la phrase (265) et du topique dans la phrase (266) (voir section I.6.3 pour une discussion sur la structure communicative envisagée dans le cadre de ce travail). D'un point de vue prosodique, nous retrouverons les contours correspondant à chaque entité communicative, avec notamment une courbe descendante sur le focus (ce qui suit le focus, c'est-à-dire le tail, est marquée par une désaccentuation et une mélodie plate) et une courbe montante sur le topique. Nous avons schématisé les courbes mélodiques dans la Figure 183.



La combinaison des éléments prosodiques n'est pas arbitraire. Dans le cas de notre exemple, nous avons rencontré l'implication suivante : si *niemand* 'personne' est le seul élément focal de la phrase, alors il est précédé d'un élément topique et il est suivi d'un élément tail ; aucune autre configuration n'est possible<sup>220</sup>. Une telle implication lie l'information communicative avec l'information prosodique sans que la syntaxe, dans son sens classique, n'entre en jeu.

<sup>219</sup> Nous rappelons qu'une phrase ne peut contenir qu'un seul focus, mais ce focus peut être constitué de plusieurs éléments groupés ensemble.

<sup>220</sup> Rossi 1999 a montré que la phrase déclarative française s'organise de la même manière, autour d'un élément « noyau », le focus. Notons également que malgré les différences prosodiques internes aux langues (langue à accent, langue à intonation etc.), les courbes mélodiques qui s'attachent aux différentes unités communicatives sont souvent très semblables entre les langues. Ainsi, en français (mais aussi dans

Dans le Mittelfeld allemand, nous rencontrons également des règles d'ordres qui semblent plutôt liées au module syntaxique. Il s'agit par exemple des pronoms faibles – souvent des pronoms personnels monosyllabiques comme *er* 'il', *ihr* 'à elle', *es* 'le' (accusatif ou nominatif neutre) etc.– qui ont tendance à se placer à gauche du Mittelfeld, à côté donc du verbe en position seconde ou du complémenteur. Traditionnellement, on appelle cette position la *position de Wackernagel*<sup>221</sup> et les pronoms qui y prennent place sont appelés les *pronoms de Wackernagel*. Cependant, le problème vient du fait que les pronoms de Wackernagel peuvent aussi apparaître plus tard dans le Mittelfeld sous des conditions prosodico-communicatives particulières. Ainsi, bien que la phrase (267) ait sans doute l'ordre standard (voir la section I.6 pour la définition que nous donnons à ce terme), le pronom de Wackernagel peut aussi apparaître après un nom plein, comme on le voit en (268)<sup>222</sup>.

- (267) (Ich glaube, dass) ihm Peter den Roman zu lesen versprochen hat.  
 (je pense, que) à\_lui Pierre le roman de lire promis a.  
 '(je pense que) Pierre lui a promis de lire le roman.'
- (268) (Ich glaube, dass) Peter ihm den Roman zu lesen versprochen hat.  
 (je pense, que) Pierre à\_lui le roman de lire promis a.  
 '(je pense que) Pierre lui a promis de lire le roman.'

D'un point de vue prosodique, le deuxième ordre des mots n'est possible que si le pronom *ihm* 'à lui' est focalisé, avec donc une courbe mélodique descendante sur ce pronom. On constate même qu'avec une telle prosodie, la linéarisation en (268) est légèrement meilleure que celle en (267). S'agit-il dans ce cas toujours d'un pronom de Wackernagel qui a changé de place ou avons-nous affaire à un homonyme qui n'est pas un pronom de Wackernagel ?

La réponse semble difficile. De manière générale, on constate que la recherche portant sur l'ordre des mots dans le Mittelfeld s'avère souvent frustrante car les règles basées sur des critères syntaxiques sont souvent contredites par un travail sur corpus (en particulier des corpus oraux) et les règles basées sur la structure communicative semblent souvent circulaires.

Pour illustrer cette difficulté, Müller 1999 (section 11.4) liste différentes restrictions syntaxiques qui ont été proposées dans la littérature, de l'ordre des mots dans le Mittelfeld pour les contredire une par une par des exemples, souvent très naturels, tirés des corpus. Müller se retrouve finalement avec un seul critère démuné de contre-exemple : un élément focalisé doit suivre un élément non focalisé. Puis il pose la question suivante : « d'où sort l'information sur le focus si on ne dispose pas d'information sur l'intonation de la phrase ? Sans une telle information basée sur le contexte [de l'énonciation], on peut justifier presque tous les ordres des constituants en leur attribuant le trait FOCUS+<sup>223</sup> » (Müller 1999:171). Enfin il conclut résigné : « A cause de

des langues comme le coréen, le grec et dans une certaine mesure l'anglais), un élément focal est marqué par une courbe mélodique descendante, un topique par une courbe mélodique ascendante souvent suivie de pause, et un tail par une courbe mélodique plate dans les basses fréquences. Voir entre autres Di Cristo & Hirst 1998 et Yoo 2000.

<sup>221</sup> D'après le linguistique Jacob Wackernagel 1853-1938. Cette position est importante pour l'explication des processus de cliticisation, qui peuvent aboutir par exemple à des suffixes de verbe.

<sup>222</sup> On trouve une différence (qui est légère à mon avis) entre les pronoms au nominatif et les autres pronoms, comme le remarque Lenerz 1993, les derniers étant plus faciles à sortir de la position de Wackernagel.

<sup>223</sup> Texte original : « Wo kommt die Fokusingformation her, wenn keine Informationen über die Information des Satzes vorliegen? Ohne Fokusingformation aus dem Kontexte kann man fast jede Konstituentenreihenfolge durch die Vergabe eines FOCUS+ rechtfertigen. »

tous ces problèmes, je n'applique pas de règles de précedence linéaire pour l'ordre dans le *Mittelfeld*<sup>224</sup> » (ibid.)

Bien que le formalisme HPSG a un caractère bidirectionnel inhérent, le système dans lequel est implémenté la grammaire HPSG de Müller n'est utilisable qu'en analyse donnant donc des analyses pour des phrases marginales, et l'auteur considère qu'il faudra écrire une grammaire à part pour la génération (Müller 1999:172). Il semble donc clair que l'amélioration d'une telle grammaire passe par la formalisation de la structure communicative et de son influence sur l'ordre des mots.

Pourtant, si nous voulons prendre la structure communicative en compte, nous nous retrouver avec le fait qu'il est parfois très difficile de définir les entités communicatives. Bien souvent, les entités communicatives sont définies par leur caractérisation prosodique, et inversement, la structure prosodique est analysée selon les entités communicatives, créant parfois une analyse circulaire du phénomène.

Sans rentrer dans le détail des théories prosodiques et communicatives existantes, nous allons ici nous intéresser juste à la question de savoir comment l'organisation communicativo-prosodique interagit avec la structure topologique. Notre approche est donc très délimitée et simplifiée : nous supposons que l'arbre de dépendance syntaxique est augmenté d'un marquage de la structure communicative, c'est-à-dire que les éléments portent l'un des trois traits communicatifs, (*Top* pour topique, *Foc* pour focus et *Tail* pour tail). Ces traits renvoient à un patron mélodique correspondant (qui peut varier par rapport à la place dans la phrase et les entrées lexicales)<sup>225</sup>. La tâche de créer cette information et de veiller sur la bonne formation de la structure communicative revient à un module antérieur à la structure topologique : c'est ce module qui veillera par exemple à ce que la phrase ne contienne qu'un seul focus, une phrase avec zéro ou deux focus donnant des ordres des mots impossibles.

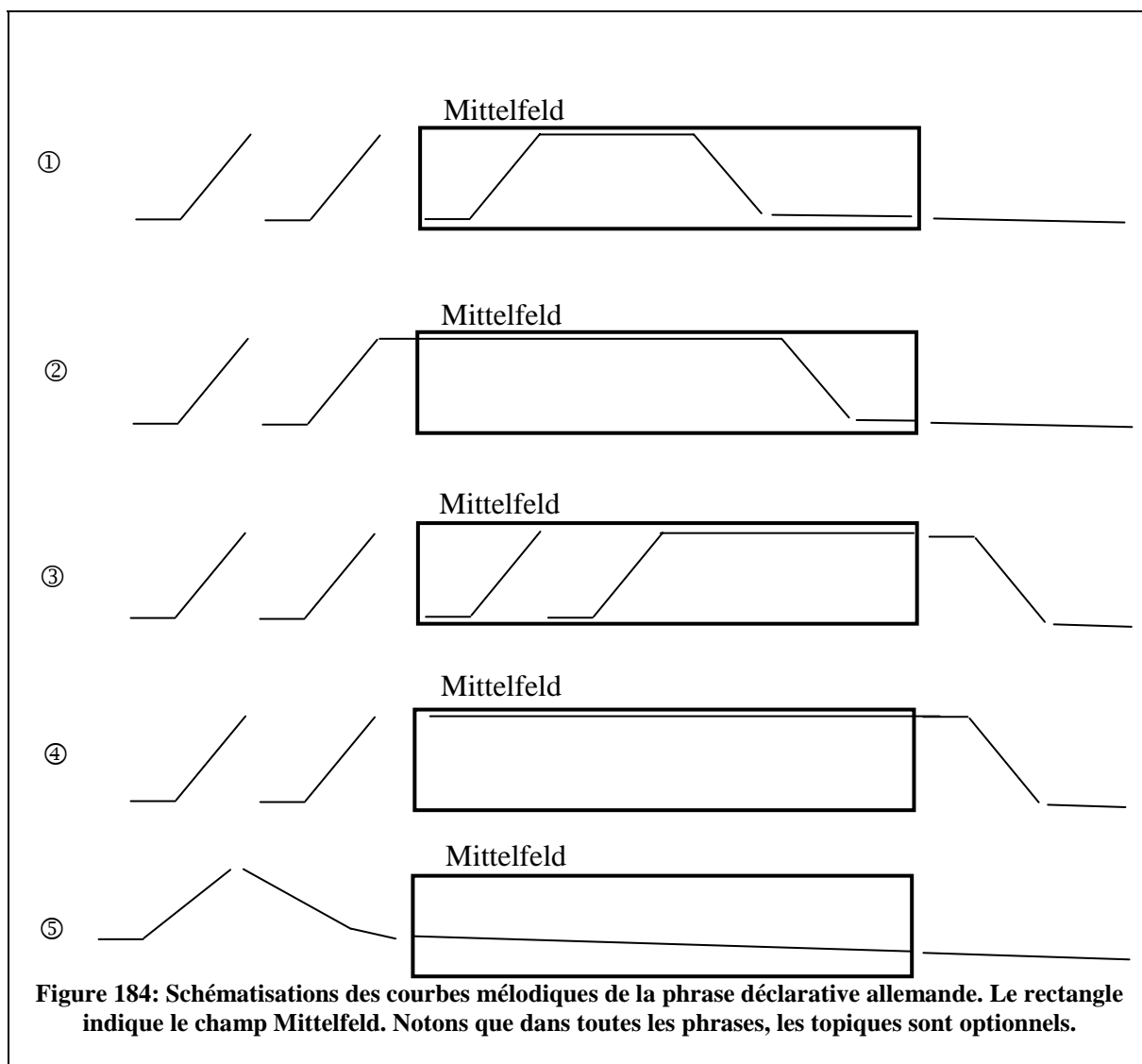
En allemand, les critères prosodiques aident beaucoup à déterminer les éléments communicatifs, car contrairement au français où l'on peut tester par la dislocation gauche ou la focalisation, on se retrouve souvent seulement avec les outils prosodiques pour distinguer ces éléments communicatifs. Nous proposons donc de procéder d'une manière inverse et de considérer d'abord les possibilités prosodiques du *Mittelfeld* pour voir à quelle organisation communicative elles pourraient correspondre. Nous présentons ces différentes possibilités dans la Figure 184.

---

<sup>224</sup> Texte original : « Wegen all dieser Probleme wende ich in meiner Implementation der HPSG keine LP-Regeln für die Reihenfolge der Komplemente im *Mittelfeld* an. »

<sup>225</sup> La grammaire que nous avons développée pour le grec moderne (dernière section de ce chapitre) repose sur les mêmes principes. L'ordre des mots du grec moderne qui est beaucoup plus libre que celui de l'allemand semble plutôt contraint par la structure communicative que syntaxique.

En général, le focus apparaît très souvent dans le Mittelfeld : ainsi, sur cinq configurations mélodiques possibles, seule la phrase ⑤ n'a pas de partie focale dans le Mittelfeld : dans ce cas, c'est le Vorfeld ou la parenthèse gauche qui accueille le focus de la phrase. Une question du type *Wer hat ihm den Roman zu lesen versprochen?* 'Qui lui a promis de lire le roman ?' permet à l'exemple (269) d'avoir une telle configuration mélodique. Les configurations ① et ② correspondent aux contours mélodiques des phrases (265) et (266). Pour obtenir les phrases ③ et ④, il faut trouver un contexte où le focus de la phrase se trouve juste après le Mittelfeld, par



exemple dans la parenthèse droite. Ainsi, l'exemple (269) peut avoir une prosodie comme la phrase ③, si *zu lesen versprochen* 'de lire promis' qui se trouve dans la parenthèse droite, est le focus de la phrase (il s'agit souvent d'un focus contrastif).

- (269) Peter hat ihm den Roman jedenfalls zu lesen versprochen.  
 Pierre a à\_lui le roman sûrement de lire promis.  
 'Pierre lui a sûrement promis de lire le livre.'

L'adverbe de phrase *jedenfalls* 'sûrement' aide à obtenir la configuration mélodique recherchée et représente un test possible pour déterminer le focus d'une phrase, mais sa présence n'est pas obligatoire. Les phrases ③ et ④ ne se distinguent finalement que par l'organisation de la partie topicale de la phrase : la phrase ③ contient un topique dans le Mittelfeld tandis que dans, la phrase ④, le topique se trouve uniquement dans le Vorfeld.



## 5.2 Analyse du Mittelfeld

Il semble donc clair que l'organisation du Mittelfeld dépend en partie de la structure communicative de la phrase. Nous suivons Frey 2000 et Abraham 2001 en posant que le Mittelfeld peut être réorganisé en des « sous-champs », comprenant une place thématique, suivie d'une place pour les adverbiaux de phrase puis d'une place rhématique.

En ce qui concerne ces adverbiaux de phrase, il s'agit d'une certaine classe d'adverbes qui expriment une évaluation faite par l'énonciateur sur les faits qui sont exposés dans la phrase. On trouve donc des adverbiaux évaluatifs (*erstaunlicherweise* 'de manière étonnante', *zum Glück* 'heureusement'), des adverbiaux d'évidence (*offensichtlich* 'évidemment', *jedenfalls* 'en tout cas', *prinzipiell* 'par principe') et des adverbiaux épistémiques (*wahrscheinlich* 'probablement', *vermutlich* 'vraisemblablement', *anscheinend* 'vraisemblablement'). Frey montre d'une manière très détaillée que la place devant ces adverbiaux est exclusivement réservée aux éléments thématiques et que le Vorfeld n'est pas concerné puisqu'il peut accueillir aussi des éléments focaux (par exemple, la phrase ⑤ de la Figure 184). Entre autres, Frey montre que la détermination d'un contexte thématique implique que la partie thématique du Mittelfeld doit précéder l'adverbial de phrase. Ceci est valable pour un sujet qui est thème (phrase (270)) ou pour un objet qui est thème (phrase (271)). Nous reprenons ci-dessous les exemples de Frey 2000:140-41 pour illustrer cela.

(270) *contexte* : Ich erzähl dir mal was von Otto.

'Je vais te raconter quelque chose sur Otto.'

a. Nächstes Jahr wird Otto wahrscheinlich seine Kollegin heiraten  
prochaine année va Otto probablement sa collègue marier.

'L'année prochaine, Otto se mariera probablement avec sa collègue.'

b. #Nächstes Jahr wird wahrscheinlich Otto seine Kollegin heiraten  
prochaine année va probablement Otto sa collègue marier.

'L'année prochaine, probablement Otto se mariera avec sa collègue.'

(271) Ich erzähl dir mal was von Otto.

'Je vais te raconter quelque chose sur Otto.'

a. Nächsten Monat wird den Otto erfreulicherweise die Botschafterin von Norwegen  
heiraten  
prochain mois va le Otto (acc.) heureusement l'ambassadrice de Norvège marier.

'Le mois prochain, heureusement, l'ambassadrice de Norvège se mariera avec Otto.'

b. #Nächsten Monat wird erfreulicherweise den Otto die Botschafterin von Norwegen  
heiraten

prochain mois va heureusement le Otto (acc.) l'ambassadrice de Norvège marier.

Frey remarque plus loin que la place devant l'adverbial de phrase n'est pas accessible à des expressions non référentielles, comme par exemple *niemand* 'personne'. Ainsi, contrairement à la phrase (272) qui est complètement naturelle, la phrase (273) peut être considérée comme agrammaticale, quel que soit la prosodie qui l'accompagne car aucun contexte ne permet d'obtenir une telle phrase.

(272) Ich glaube dass Peter wahrscheinlich dem Mann den Roman zu lesen versprochen hat.

je pense que Pierre probablement à l'homme le roman de lire promis a.

'Je pense que Pierre a probablement promis à l'homme de lire le livre.'

(273) <sup>\*/??</sup> Ich glaube dass niemand wahrscheinlich dem Mann den Roman zu lesen versprochen hat.

je pense que personne probablement à l'homme le roman de lire promis a.

'?? Je pense que personne a probablement promis à l'homme de lire le livre.'

L'approche que nous adoptons est une simplification de Frey 2000 et de Abraham 2001 car nous ne distinguons pas de critères pour le placement d'un élément thématique dans le Vorfeld et celui dans sa place spécifique du Mittelfeld. En effet, dans une structure transformationnelle qui est à la base de leurs approches, la place devant le V2 est d'une nature différente que celle après le V2. Nous ne rejetons pas l'idée qu'une gradation entre les différents topiques en fonction de leur position peut exister, le topique contenu dans un Vorfeld étant plus proéminent qu'un topique qui est placé dans le Mittelfeld, mais dans le cadre de l'approche simplifiée que nous présentons, cette distinction ne nous semble pas être d'une grande importance.

### 5.3 Réorganisation du champ Mittelfeld

Nous considérons que beaucoup de phénomènes d'ordre dans le Mittelfeld peuvent être expliqués à l'aide de la tripartition des valeurs communicatives en topique, focus et tail. Dans la séparation du Mittelfeld en sous-champs de Frey 2000, les places sont considérées comme fixes et il ne pose pas clairement les restrictions de placement par rapport aux autres unités communicatives de la phrase. Par contre, Abraham 2001 considère une distribution des places de thèmes et de rhèmes qui varie en fonction de l'organisation communicative globale de la phrase. D'une manière similaire, nous posons aussi différentes suites de champs, mais dans notre formalisme de linéarisation (Cf. II.1), c'est au moment où un élément ouvre un domaine qu'il faut décider des places que le domaine contient. Pour le domaine principal de la phrase, c'est donc la racine de l'arbre de dépendance qui décide sur la forme du domaine principal, car c'est elle qui ouvre ce domaine. Le domaine principal qu'ouvre la racine de l'arbre syntaxique, comporte toujours un Vorfeld, une parenthèse gauche et droite et un Nachfeld ; par contre l'organisation du Mittelfeld varie fortement en fonction de la valeur communicative de la racine. Nous décrivons cas par cas les places que propose un tel gouverneur en fonction de sa valeur communicative. L'ancien Mittelfeld est réorganisé en quatre sous-champs : le TopF (*topicfield*) où des éléments topicalisés peuvent se placer, l'AdvF (*adverbialfield*) pour les adverbiaux de phrase, le FocF (*focusfield*) pour le focus et le tailF (*tailfield*) pour un élément communicativement marqué tail<sup>226</sup>.

Lorsque le gouverneur syntaxique de la phrase est topical, les quatre sous-champs peuvent être disponibles entre parenthèse gauche et parenthèse droite, au lieu de l'ancien Mittelfeld. Pourtant, pour que la phrase soit linéarisée correctement et les valeurs communicatives des éléments ne soient pas en contradiction avec la courbe mélodique globale de la phrase, il faut poser deux conditions : il faut que le Vorfeld soit lui aussi occupé par un élément topical et il faut que la parenthèse droite et le Nachfeld, qui suivent une place pour le tail, soit eux-aussi tail, car après un tail, on ne peut rencontrer ni topique ni focus. Il est facile d'exprimer de telles conditions dans notre formalisme : il suffit d'exprimer cette capacité d'accueil des champs avec des noms des champs différents. Nous aurons donc un Vorfeld avec une restriction d'accès limitée aux éléments topiques, appelé « VF-T ». Chaque règle de correspondance qui place un élément dans ce champ, doit aussi vérifier la valeur communicative de l'élément à placer ; seul un topique peut y aller (cf. section 1.2). De la même manière nous aurons une parenthèse droite et un Nachfeld spécialisés dans l'accueil des éléments tails : « )-tail » et « NF-tail ». Pour un gouverneur syntaxique topical, nous obtenons donc un domaine principal avec des champs restreints par traits communicatifs :

---

<sup>226</sup> Notons qu'il serait plus correct d'ajouter un champ WackerF (*Wackernagelfield*) qui pourrait être occupé par un élément topique ou tail. Un tel champ devrait se trouver entre la parenthèse gauche et le TopF, permettant ainsi à tous les pronoms de Wackernagel qui ne sont pas focus d'apparaître au début du Mittelfeld. Cependant, nous laissons de côté cette distinction afin de maintenir une description simple.

VF-T	(-T	TopF	AdvF	FocF	tailF	)-tail	NF-tail
------	-----	------	------	------	-------	--------	---------

Une telle configuration correspondra à la fin à la configuration mélodique ① de la Figure 184.

Pourtant, ce n'est pas le seul domaine principal qu'un gouverneur syntaxique topical peut ouvrir. Afin de permettre une réalisation dans le cas où son dépendant verbal dans la parenthèse droite est focal, il faut empêcher un élément tail de se placer devant le focus, car cet ordre est prosodiquement impossible (sauf cas de pronoms de Wackernagel). La façon plus simple façon d'assurer un tel placement est de faire disparaître le champ tail :

VF-T	(-T	TopF	AdvF	FocF	)-F	NF
------	-----	------	------	------	-----	----

Notons que nous ne spécifions plus la valeur communicative du Nachfeld, car il peut être occupé par un élément focus ou par un élément tail. Pour qu'aucun topique n'aille dans le Nachfeld, il suffit que chaque règle de correspondance plaçant un élément dans le Nachfeld exige qu'il s'agisse d'un élément focal ou bien d'un élément tail. De la même manière nous posons que chaque règle de correspondance qui place un élément dans le Vorfeld, vérifie qu'il s'agit bien d'un focus ou d'un topique et jamais d'un tail.

Le troisième et dernier choix de valeur communicative de la parenthèse droite est *topique*. Lorsque parenthèse gauche et droite sont toutes les deux topiques, aucun autre champ que topique (ni adverbe, focus ou tail) ne peut apparaître entre les deux parenthèses :

VF-T	(-T	TopF	)-T	NF-F
------	-----	------	-----	------

Remarquons que nous pouvons donner la valeur focus au Nachfeld, car chaque phrase contient au moins un élément avec le trait communicatif *focus*. Un tel cas de focus dans le Nachfeld peut arriver avec l'extraposition habituelle de complétive (ou d'infinitifs) dans le Nachfeld, mais aussi avec l'extraposition nominale. Un exemple d'une telle phrase peut être trouvée chez Müller 1999, où l'auteur cite plusieurs exemples réels d'extraposition nominale dans le Nachfeld (Müller 1999:13.1.1.3). En voici un (parenthèse gauche et droite en gris) :

(274) Unter denen des alten Indien **muten** uns am fremdartigsten **an** die Kapitel über Nägelwunden, das Beißen mit den Zähnen und die Anwendung von Schlägen.

Parmi ceux de l'ancienne Inde, donnent\_l'impression\_de nous[dat.] le plus\_étrange [particule de donner l'impression] les chapitres sur les blessures d'ongles, les morsures avec les dents et l'application de coups.

'Les chapitres sur les blessures par ongles, les morsures par dents et l'application de coups nous semblent les plus bizarres parmi [les pratiques] de l'Inde antique.'

Notons que la prosodie de cette phrase a obligatoirement une montée très marquée sur la particule *an* qui se trouve à la fin d'une série d'éléments classés comme topiques.

Remarquons que ces trois structures précédentes correspondent aux configurations mélodiques ③ ou ④ de la Figure 184.

Passons maintenant aux cas où la parenthèse gauche est un élément focal. Ici, les choix sont beaucoup plus restreints, car aucun élément topique ne peut suivre la parenthèse gauche :

VF	(-F	AdvF	FocF	tailF	)-tail	NF-tail
----	-----	------	------	-------	--------	---------

Nous n'avons plus besoin de spécifier l'accès au Vorfeld, car il peut être topique ou focus.

Dans le cas où la parenthèse droite est occupée par un élément focal, aucun tail ne peut apparaître devant elle et nous n'avons plus besoin de spécifier le type communicatif du Nachfeld, car il peut aussi bien être occupé par un focus que par un tail :

VF	(-F	AdvF	FocF	)-F	NF
----	-----	------	------	-----	----

Un seul schéma peut apparaître en tant que domaine principal au cas où la parenthèse gauche est déjà tail. En ce cas, elle est suivie de rien que des éléments tail et, comme toute phrase a un focus, le Vorfeld sera occupé par l'élément focus :

VF	(-tail	tailF	)-tail	NF-tail
----	--------	-------	--------	---------

Il est clair que cette analyse topologique dépend de la structure communicative qu'il obtient en tant qu'entrée. Il est envisageable que le traitement de la structure communicative ne donne pas de valeur communicative à chaque élément ou encore qu'il marque quelques éléments comme des éléments neutres. Par exemple, si l'information sémantique du temps fait apparaître un auxiliaire au niveau syntaxique, il n'est pas obligatoirement marqué d'une valeur communicative<sup>227</sup>. Il sera au module sémantique de décider s'il met une valeur communicative à un tel nœud ou non. De manière générale, il nous semble en plus possible qu'un verbe fini ayant le marquage de tail puisse apparaître dans la parenthèse droite entouré de topiques ou de focus. Pour obtenir ces ordres, nous pouvons considérer que les six gabarits que nous avons rencontrés permettent également le placement d'un élément neutre ou tail dans la parenthèse gauche.

Ces six gabarits constituent les seuls domaines qui peuvent apparaître en tant que domaine principal. Bien que cela amène une complexification considérable par rapport au schéma unique d'une phrase déclarative comme nous l'avons considéré dans la formalisation sans valeurs communicative, un tel traitement reste d'une complexité raisonnable en vue du résultat qu'il fournit : avec un tel traitement, le nombre de linéarisations se réduit énormément, souvent à une ou deux possibilités par arbre de dépendance donné. Seul les phrases thétiqes (qui ne sont constitués que d'éléments rhématiques) gardent toute leur liberté de placement.

#### 5.4 Illustration des nouveaux champs introduits

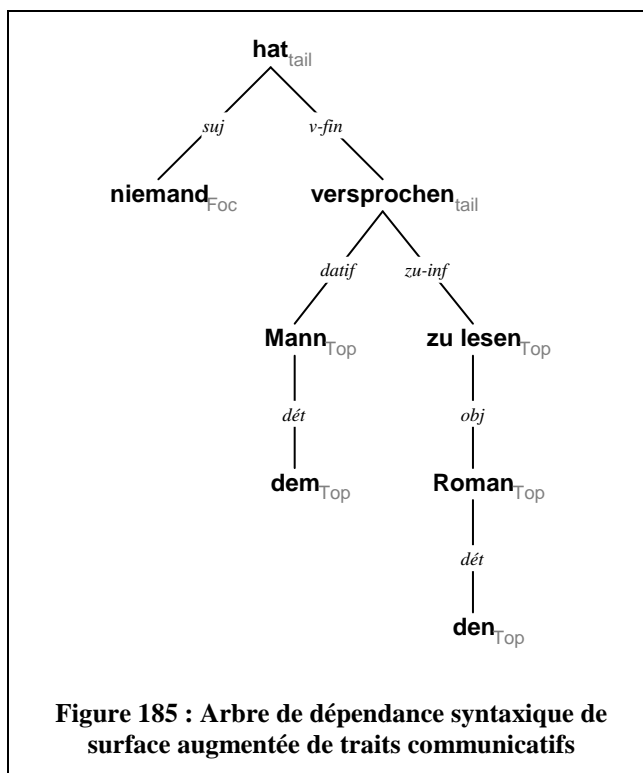
Nous terminons cette section avec l'illustration des règles de placement en prenant en compte les nouveaux champs que nous avons introduit.

Nous partons comme d'habitude d'un arbre de dépendance syntaxique mais qui, cette fois-ci, est augmenté d'un marquage communicatif (Figure 185). Dans l'exemple choisi, l'arbre de dépendance syntaxique a deux parties non connexes qui ont la même valeur communicative.

---

<sup>227</sup> Bien entendu, un auxiliaire peut porter un marquage communicatif comme dans l'exemple (i), c'est-à-dire avec interprétation contrastive. Pour une discussion de ces phénomènes appelés *Verum-Fokus*, voir Höhle 1986.

- (i) Peter wird den Roman nicht lesen, er HAT ihn schon gelesen.  
 Pierre va le roman non lire, il A le déjà lu.  
 'Pierre ne va pas lire le roman, il l'A déjà lu.'



Le gouverneur syntaxique *hat* ‘a’ ouvre un domaine principal et se place dans la parenthèse gauche :

VF-T	(-T ou tail	TopF	AdvF	FocF	tailF	)-tail	NF-tail
	hat						
	a						

Figure 186: La tête syntaxique *hat* ‘a’ ouvre un domaine principal qui comporte ces champs et se place

VF-T	(-T ou tail	TopF	AdvF	FocF	tailF	)-tail	NF-tail
	hat			niemand		versprochen	
	a			personne		promis	

Figure 187: Placement des dépendants de *hat* ‘a’ : le dépendant verbal se place dans la parenthèse droite et son dépendant nominal dans FocF.

Son dépendant verbal *versprochen* ‘promis’ se place dans la parenthèse droite qui est marquée tail. Comme certains éléments de l’arbre de dépendance syntaxique sont marqués comme topiques, et que la parenthèse droite est occupée par un tail, le dépendant nominal de la tête syntaxique *niemand* ‘personne’ n’a pas d’autre choix que de placer dans le champ FocF.

Enfin, les dépendants de *versprochen* ‘promis’ sont beaucoup plus libres. Comme nous n’avons pas indiqué de hiérarchie entre les deux dépendants, ils peuvent au choix aller soit dans le Vorfeld, soit dans le TopF : aucun des deux ne peut aller dans le Nachfeld qui ne peut recevoir que des éléments tail car la parenthèse droite est prise par un élément tail (sinon, il y aurait violation de la bonne formation de la structure communicative). Par ailleurs, dans les deux cas, le verbe *zu lesen* ‘de lire’ ouvre un nouveau domaine. En effet, il ne peut pas rejoindre son gouverneur dans la parenthèse droite car ces deux éléments n’ont pas la même valeur communicative (il y aurait encore violation de la bonne formation de la structure communicative).

Un marquage communicatif sur l’arbre de dépendance syntaxique permet donc de réduire drastiquement le nombre des linéarisations qu’un arbre de dépendance syntaxique peut donner.

Il nous semble important de souligner que la structure topologique même munie de l’information

<i>VF-T</i>	<i>(-T</i>	<i>TopF</i>	<i>AdvF</i>	<i>FocF</i>	<i>tailF</i>	<i>)-tail</i>	<i>NF-tail</i>
den Roman zu lesen	hat	dem Mann		niemand		versprochen	
le roman de lire	a	à cet homme		personne		promis	

<i>VF-T</i>	<i>(-T</i>	<i>TopF</i>	<i>AdvF</i>	<i>FocF</i>	<i>tailF</i>	<i>)-tail</i>	<i>NF-tail</i>
dem Mann	hat	den Roman zu lesen		niemand		versprochen	
à cet homme	a	le roman de lire		personne		promis	

Figure 188: Les deux possibilités de placement des dépendants de *versprochen* ‘promis’.

communicative ne représente pas la structure prosodique en soi. La structure topologique représente une interface permettant de s’approcher à la production finale de la phrase. En effet, pour une analyse complète, il aurait fallu prendre en compte les spécificités phonétiques de l’allemand (au niveau du mot et du groupement des mots). C’est également dans un niveau intermédiaire que sera calculé la courbe mélodique du verbe de la parenthèse gauche et ce en fonction de la structure communicative des éléments voisins (et donc de la courbe mélodique correspondante de ces éléments).

Le modèle topologique semble donc être une bonne interface pour la structure prosodique. L’implémentation du formalisme en DepLin permet un développement concret, avec une sortie son des résultats et nous montrerons une application de cette procédure, notamment avec le grec moderne, dans la dernière sous-section de ce chapitre.

## 6 La topologie du syntagme nominal allemand

Dans cette section, nous montrons comment une approche topologique peut être utilisée pour décrire l’ordre interne au group nominal en allemand. Eisenberg 1999:400 remarque que « la topologie du syntagme nominal a été beaucoup moins souvent traité que celle de la phrase dans la littérature » et que souvent, elle est s’appuie sur la topologie de la phrase. Nous suivons

partiellement cette approche bien que nous soyons conscient du danger de la projection des propriétés de la phrase sur le syntagme nominal peut amener. »<sup>228</sup>

Malgré cette mise en garde, nous procédons à un élargissement de la topologie au syntagme nominal car il est important de montrer l'aptitude du modèle topologique à tous les phénomènes d'ordres de la langue. De plus, nous pensons qu'il est possible de faire des rapprochement entre la topologie de la phrase et celle du syntagme nominal : par exemple, le comportement de la parenthétique verbale est parallèle à celui du lien qui existe entre le déterminant et le nom (Eisenberg 1999:401) et il serait dommage de ne pas rechercher les similarités qui existent.

Nous proposons dans un premier temps que le syntagme nominal puisse être placé dans un domaine du syntagme nominal qui comporte 6 champs : le champ du déterminant (DétF)<sup>229</sup>, le champ des adjectifs (AdjF), le champ du nom (NomF), le champ du génitif (GenF), le champ du syntagme prépositionnel (SPF) et le champ de la relative (RelF)<sup>230</sup>.

<i>DétF</i>	<i>AdjF</i>	<i>NomF</i>	<i>GenF</i>	<i>SPF</i>	<i>RelF</i>

Figure 189: Domaine du syntagme nominal.

Il est envisageable d'affiner la structure interne du champ des adjectifs (AdjF) de la même manière que le Mittelfeld, car il semble qu'il existe des ordres plutôt fixes parmi différents types d'adjectifs, mais, pour une présentation simplifiée, nous allons garder un AdjF unique.

Dans une analyse habituelle du syntagme nominal, on considère, comme l'indique son nom, que le nom est la tête du syntagme. Si on considère donc la structure habituelle de dépendance syntaxique de surface d'un groupe nominal, le nom est modifié par les adjectifs, les syntagmes prépositions et les relatives de la même manière qu'il est déterminé par l'article. Bien qu'au niveau sémantique la prédominance du nom semble claire, au niveau syntaxique, nous ne connaissons pas de bonne raison pour considérer que la tête syntaxique doit être le nom. Dans une description qui se place dans le cadre de la Théorie Sens-Texte, nous pouvons bien imaginer que pendant la correspondance sémantique–syntaxe, le déterminant apparaît et se place en tant

<sup>228</sup> texte original : « Die Topologie der NGr [Nominalgruppe] ist in der Literatur viel weniger ausführlich behandelt worden als die des Satzes, und meist wird sie im Anschluß an die des Satzes und unter Bezug darauf behandelt. Wir folgen diesem Vorgehen zum Teil, wollen uns aber bewusst bleiben, daß damit die Gefahr einer Projektion von Eigenschaften des Satzes auf die NGr verbunden ist. » (ma traduction)

<sup>229</sup> Ce champ du déterminant peut aussi accueillir un génitif saxon au lieu du déterminant.

<sup>230</sup> Ce champ de la relative peut aussi accueillir d'autres postpositions phrastiques au nom, comme par exemple les (très rares) participes (ou adjectives) épithètes postposés.

- (i) Das schöne Auto meines Bruders, gestohlen am letzten Dienstag, ist wieder aufgetaucht.  
La belle voiture de mon frère, volée mardi dernier, est réapparue.

Lorsque un syntagme nominal comprend un participe postposé et une relative, la phrase devient moins acceptable et il est difficile d'évaluer une éventuelle préférence d'ordre entre les deux :

- (ii) ?? Das schöne Auto meines Bruders, gestohlen am letzten Dienstag, das er mir geliehen hatte, ist wieder aufgetaucht.

La belle voiture de mon frère, volée mardi dernier, qu'il m'avait prêtée, est réapparue.

- (iii) ?? Das schöne Auto meines Bruders, das er mir geliehen hatte, gestohlen am letzten Dienstag, ist wieder aufgetaucht.

?? La belle voiture de mon frère, qu'il m'avait prêtée, volée mardi dernier, est réapparue.

En défaut d'une analyse plus approfondie de ce phénomène, nous considérons donc tout simplement que l'épithète postposé et la relative occupent le même champ.

<b>die</b>	<b>alte(n)</b>	<b>Männer</b>	<b>(mit Hüten)</b>
les	vieux	hommes	avec des chapeaux
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
#			

Figure 190 : Tableau de Netter 1994.

que gouverneur du nom. Il est difficile de déterminer la tête du SN, au moins juste par des faits purs de distribution, comme le montre Netter 1994.

Reprenons d'abord le tableau de Netter 1994 que nous reproduisons ci-dessous (Figure 190) : les lignes indiquent un syntagme nominal complet. Les  $2^3-1=7$  combinaisons du déterminant, adjectif et nom sont toutes permises. C'est seulement s'il n'y a aucun d'entre eux qu'on n'obtient pas de syntagme nominal (donné à la dernière ligne). Le syntagme prépositionnel *mit Hüten* 'avec chapeaux' n'est pas obligatoire pour la grammaticalité du syntagme ; le déterminant *die* sans modification peut aussi former un syntagme nominal, mais dans un langage écrit, on préférerait le démonstratif *diese* 'ces', mais avec le SP, *die* 'les' convient bien. S'il est précédé par le déterminant *die*, l'adjectif doit suivre la déclinaison faible (*alten*), sinon forte (*alte*) 'vieux'.

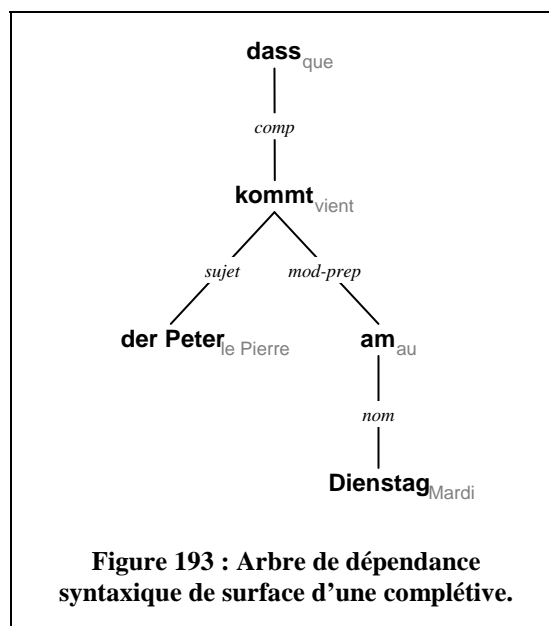
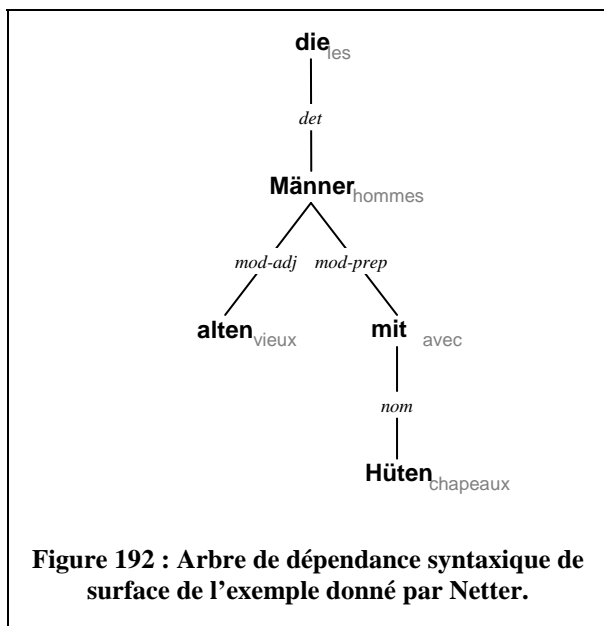
Depuis l'hypothèse DP d'Abney 1987 dans la théorie X-barre, on trouve beaucoup d'analyses qui placent le déterminant en tête du syntagme nominal. Cela est souvent fait pour exprimer des généralités qui existent entre le déterminant et le complémenteur. De la même manière (mais sans que cela soit obligatoire pour le traitement topologique que nous présentons), nous considérons que le déterminant, quand il est présent, est la tête du groupe nominal, même quand nous suivons l'hypothèse DP. Les autres éléments du groupe nominal dépendant du nom. Pour faciliter d'expression et en tenant compte de la structure sémantique, nous parlerons à partir de maintenant de groupe nominal. En fait, une telle analyse nous permet de faire un parallèle entre le positionnement du déterminant et celui du complémenteur<sup>231</sup>.

Prenons comme point de départ l'arbre de dépendance que nous présentons dans la Figure 192. Nous considérons que le déterminant ouvre le domaine nominal et se place dans le champ du déterminant. Son dépendant nominal se place dans le champ du nom et l'adjectif et le syntagme prépositionnel dans les champs de leur nom. Nous obtenons ainsi la structure topologique donnée dans la Figure 191.

<i>DétF</i>	<i>AdjF</i>	<i>NomF</i>	<i>GenF</i>	<i>SPF</i>	<i>RelF</i>
die	alten	Männer		mit Hüter	
les	vieux	hommes		avec chapeaux	

Figure 191: Linéarisation possible de l'arbre de dépendance syntaxique de surface de la Figure 192.





Notons que les éléments se place dans les deux cas d'une manière très similaire et dans le même ordre. L'arbre de dépendance, mais aussi la structure topologique, est très semblable à une structure avec complémenteur, comme nous le montrons dans les Figure 193 et Figure 194. On observe une seule différence : en ce qui concerne la phrase complète, le placement du syntagme prépositionnel *am Dienstag* 'au jeudi' dans le Nachfeld n'est évidemment pas obligatoire ; il peut aussi bien, en fonction de sa valeur communicative, se placer dans le Mittelfeld.

Deux observations rendent cette analyse topologique du syntagme nominal intéressante :

D'une part, nous constatons que les syntagmes prépositionnels et les relatives peuvent

<i>CF</i>	<i>Mittelfeld</i>	<i>]</i>	<i>Nachfeld</i>
dass	der Peter	kommt	am Dienstag
que	le Pierre	vient	mardi

**Figure 194: Linéarisation possible de l'arbre de dépendance syntaxique de surface de la Figure 193**

s'émanciper du domaine de leur gouverneur (le groupe nominal) et se place dans le domaine du verbe du gouverneur et peuvent à nouveau aboutir dans la place qui suit la parenthèse droite : le Nachfeld (l'autre possibilité est le placement dans le Vorfeld du domaine supérieur).

D'autres part, on constate un parallélisme étonnant entre la formation du complexe verbal et les noms composés.

Prenons comme exemple de ce dernier phénomène, un « gros » mot de l'allemand :

- (275) das neue Verkehrs<sub>5</sub>infrastruktur<sub>4</sub>finanzierungs<sub>3</sub>gesellschafts<sub>2</sub>gesetz<sub>1</sub><sup>232</sup>  
 la nouvelle circulation-infrastructure-financement-société-loi  
 la nouvelle loi de société de financement d'infrastructure de la circulation

Nous remarquons que l'ordre dans « le complexe nominal » est identique à celui d'un complexe verbal : après que l'élément qui gouverne (syntaxiquement) le syntagme nominal, le déterminant, a pris position dans son champ qui correspond au champ du complémenteur, son dépendant, le

<sup>231</sup> Ce rapprochement n'est ni étonnant vu l'interaction diachronique qui existe entre les deux catégories, ni original et limité dans le cadre d'une analyse topologique.

<sup>232</sup> Journal taz Nr. 6584 du 26/10/2001, page 20.

nom le plus haut dans la structure de dépendance (*Gesetz* 'loi') prend place dans le champ du nom. On peut considérer que les règles topologiques fonctionnent de manière similaire à celles de la parenthèse droite : le nom ouvre un domaine, parallèle à la boîte verbal, contenant une place à sa gauche pour ses dépendants nominaux.

On peut même constater un parallélisme en ce qui concerne la forme des mots qui entrent dans cette sorte de constructions « amalgamées ». En effet, dans l'exemple suivant (276), nous constatons que les trois éléments supérieurs du complexe sont de type verbal. Ils peuvent aussi apparaître dans des constructions purement verbales comme dans (277). Le nom *Dreckwäsche* 'linge sale' sans déterminant a un caractère ambigu en ce qui concerne sa catégorie fonctionnelle. Son placement dans la phrase (276) est exclu s'il est muni d'un déterminant (cf. phrase (278)). Dans la phrase (279), il apparaît sans déterminant juste devant le verbe. Il est possible de l'analyser en tant que nom habituel, car les génériques de l'allemand peuvent apparaître sans article. Mais on peut aussi le considérer comme une sorte de particule, car il montre deux caractéristiques des particules verbales : prosodiquement, *Dreckwäsche* et le complexe verbal qui suit peuvent se prononcer comme une seule unité mélodique, et syntaxiquement, la phrase (280) montre qu'il peut apparaître dans les constructions de Oberfeldumstellung (cf. II.3), où, normalement, seuls des éléments de caractère verbal (comme verbes et préfixes) peuvent apparaître.

- (276) Das ist klasse zum Dreck<sub>5</sub>wäsche<sub>4</sub>rum<sub>3</sub>liegen<sub>2</sub>lassen<sub>1</sub><sup>233</sup>  
 cela est la\_classe pour\_le saleté-linge-autour-posser-laisser  
 C'est pratique si on laisse traîner autour de soi son linge sale.
- (277) Ich will meine Dreckwäsche überall rum<sub>3</sub>liegen<sub>2</sub>lassen<sub>1</sub>  
 Je veux mon saleté-linge partout autour-posser-laisser.  
 Je veux laisser traîner partout autour de moi mon linge sale.
- (278) \* Das ist klasse zum die Dreck<sub>5</sub>wäsche<sub>4</sub>rum<sub>3</sub>liegen<sub>2</sub>lassen<sub>1</sub>  
 cela est la\_classe pour\_le saleté-linge-autour-posser-laisser  
 C'est pratique si on laisse traîner autour de soi son linge sale.
- (279) Ich will Dreckwäsche rum<sub>3</sub>liegen<sub>2</sub>lassen<sub>1</sub>  
 Je veux saleté-linge autour-posser-laisser.  
 Je veux laisser traîner mon linge sale.
- (280) Sie war böse weil sie nicht überall hat Dreck<sub>6</sub>wäsche<sub>5</sub> rum<sub>4</sub>liegen<sub>3</sub>lassen<sub>2</sub>dürfen<sub>1</sub>.  
 elle était fâchée parce que elle non partout a saleté-linge-autour-posser-laisser-  
 avoir\_permission.  
 Elle se fâchait car on lui n'a pas permis de laisser traîner partout autour d'elle du linge  
 sale.

Une analyse plus détaillée de ces phénomènes dépasse le cadre de cette thèse, mais il semble d'une part évident que le modèle topologique peut servir dans la linéarisation du syntagme nominal (le domaine du nom) comme de la proposition (le domaine du verbe), et d'autre part qu'il existe des parallèles étonnantes entre la topologie verbale et la topologie nominale.

---

<sup>233</sup> Ma parole spontanée à propos des avantages pratiques d'un mezzanine, remarqués (et notée) par Stefan Müller.

## 7 Topologie comparative : coréen, grec, français

Jusqu'à présent, nous avons exploré les atouts d'une analyse topologique pour différents phénomènes de l'allemand. Nous avons vu la simplicité avec laquelle beaucoup de questions de linéarisation peuvent être résolues. La question qui se pose est de savoir à quel point le modèle topologique peut servir pour d'autres langues. Dans le formalisme TAG, Rambow & Lee 1994 comparent le scrambling en coréen et en allemand. Sous une même optique, dans une analyse basée sur la théorie de l'optimalité et décrite dans le formalisme LFG, Choi 1999 montre que les restrictions d'ordres qui régissent le coréen ressemblent, sans pour autant être identiques, à celles de l'allemand. D'une manière similaire à l'allemand, le verbe principal de la phrase coréenne a une place fixe en fin de phrase, et ne peut être suivi que par des éléments non accentués (tail). Par contre, les autres dépendants nominaux du verbe se placent plus librement dans une région qui ressemble au Mittelfeld allemand. Une autre similarité entre ces deux langues concerne la formation de complexes verbaux, notamment avec des phénomènes appelés traditionnellement brouillage d'arguments (scrambling). Une étude est en cours sur des possibilités d'une analyse topologique du coréen (Gerdes & Yoo, en préparation).

Pour une langue comme l'anglais ou le français, nous pourrions envisager une analyse topologique comprenant des champs dont l'accès est pour la plus grande partie contraint par le type de dépendance syntaxique qu'entretient un mot avec son gouverneur. Par exemple, pour une phrase déclarative du français, il faudra prévoir une place pour le sujet<sup>234</sup>, des places pour les clitiques, la place du verbe, un champ un peu plus libre qui pourrait accueillir des éléments comme l'objet direct et des syntagmes prépositionnels etc. La structure topologique résultante d'une telle grammaire ressemblerait beaucoup à un arbre de constituants traditionnel. En français, la question d'inversion du sujet, un phénomène « d'ordre libre », peut amener à des analyses « topologiques » du français. Nous renvoyons aux travaux de Marandin 2001 et de Cori et Marandin 1993 pour une analyse du français basée sur l'idée d'une structure de sites ordonnés dans la phrase.

Par contre, pour les langues comme le russe ou le grec moderne, où l'ordre des mots est encore plus libre qu'en allemand (cf. définition dans la section I.6), la structure topologique semble peu restreinte par des critères syntaxiques, mais plutôt liée à la structure communicative.

Dans la suite nous explorons les possibilités d'une analyse topologique du grec moderne. Nous présentons les résultats préliminaires d'un travail effectué en collaboration avec Hi-Yon Yoo.

### 7.1 La topologie du grec moderne

Dans les grammaires traditionnelles, le grec moderne est présenté comme une langue « à ordre libre » (Holton *et al.* 1997). La justification que l'on trouve pour ce fait est que, par exemple, pour une phrase à deux arguments nominaux (SVO), les six permutations sont attestées (SVO, SOV, OSV, OVS, VSO, VOS). Or si l'on prend en considération les possibilités de variation prosodique pour un ordre donné, on s'aperçoit que le nombre de phrases attestées pour une structure argumentale augmente considérablement. Pour une phrase à deux arguments, on a à faire non plus à six phrases possibles, mais à 25 réalisations possibles, combinant les variations d'ordre des mots et les différentes réalisations prosodiques. Dire donc qu'une langue est « à ordre libre des mots », revient finalement à considérer que pour une structure argumentale donnée, beaucoup d'ordres des mots, mais pas tous, sont possibles, et ce avec différentes réalisations prosodiques.

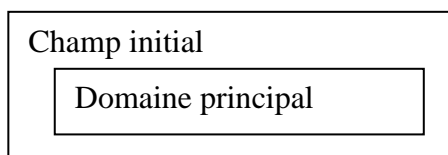
---

<sup>234</sup> Cette place peut principalement accueillir le sujet et éventuellement certains éléments lexicalement et communicativement très restreints.

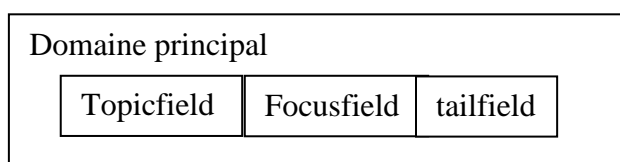
Néanmoins, une langue comme le grec moderne a un comportement bien différent de l'allemand ou encore du coréen car le placement des éléments de phrase est encore davantage régi par des contraintes d'ordre sémantico-communicatives que syntaxiques<sup>235</sup>. Contrairement à l'allemand où nous avons une place syntaxiquement déterminée pour le verbe principal de la phrase (V2), le grec moderne semble plus « libre » dans le sens où c'est essentiellement le marquage communicatif porté par les éléments et hérité du niveau sémantique qui détermine leur place linéaire. D'un point de vue topologique, la structure du grec ressemble au Mittelfeld de l'allemand (cf. IV.5).

### 7.1.1 Les règles topologique du grec

La structure topologique du grec moderne est conçue de la manière suivante. Elle consiste en un champ initial qui ouvre un domaine principal.



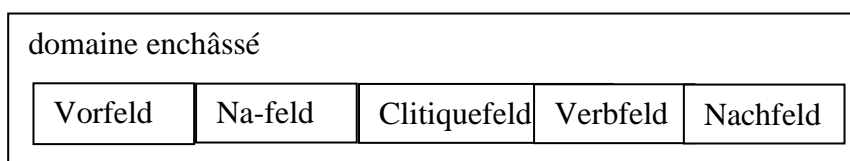
Le domaine principal ouvre trois champs majeurs : ces champs ont été nommés à partir du marquage communicatif, compte tenu du fait que les éléments ayant le même marquage, seront susceptibles d'être placés dans ces champs. Dans l'ordre, nous avons donc le TopF (champ de topique), le FocF (champ du focus) et le tailF (champ du tail). Le découpage communicatif suit les conventions introduites en section I.6.3. Nous indiquons dans les règles de description de champs sur le fait que le FocF contient exactement un élément (un domaine enchâssé), tandis que le TopF et le tailF n'ont pas de restrictions quant au nombre d'éléments qu'ils peuvent accueillir.



La tête verbale se place dans l'un des champs majeurs en fonction de sa valeur communicative : s'i elle a un marquage communicatif de focus, elle sera placé dans le champ focus et ainsi de suite.

S'appliquent ensuite les règles de création de domaines enchâssés.

La tête verbale de la dépendance syntaxique ouvre un domaine enchâssé dans le champ majeur où il a été placé.



Le domaine enchâssé ouvert par la tête verbale comporte 5 champs :

- celui de vorfeld (suivant la même dénomination que pour la topologie de l'allemand, mais contrairement à l'allemand ce champ peut contenir un nombre quelconque d'éléments)
- le Na-feld (la place pour le marquage de certaines formes verbales comme le futur et les verbes en *na* ; il contient un élément ou reste vide),

<sup>235</sup> les contraintes syntaxiques peuvent également intervenir, comme l'illustre l'exemple sur les compléments phrastiques que nous présentons dans la suite.

- le Clitiquefeld (la place pour les clitiques ; tous les clitiques qui apparaissent à ce niveau y seront placés<sup>236</sup>),
- le Verbfeld (ne contient que le verbe) et
- le Nachfeld (pour d'autres dépendants du verbe, sans restriction du nombre d'éléments qu'il peut contenir.).

La tête verbale ouvre donc ce domaine enchâssé et se place dans le champ du verbe, qui doit être obligatoirement rempli. Ensuite, les dépendants verbaux peuvent se placer. Si les dépendants partagent le même marquage communicatif, ils restent à l'intérieur du même domaine et se placent selon le rôle qu'ils jouent dans la dépendance syntaxique ou leur nature grammaticale : ainsi, un clitique se place dans le Clitiquefeld, le complément sujet dans le Vorfeld ou le Nachfeld, le complément objet direct dans le Nachfeld etc. Nous supposons que la structure communicative est une partition, c'est-à-dire que chaque élément appartient à un et un seul groupement communicatif. Afin de représenter à l'aide de traits simples la superposition d'une telle structure communicative sur l'arbre de dépendance syntaxique, nous introduisons la convention suivante : seul le gouverneur syntaxique d'un groupe porte la valeur communicative du groupe, ses dépendants étant marqués N comme « neutre ». Autrement dit, chaque élément qui porte une valeur autre que N ouvre un autre groupement communicatif.

Par contre si un dépendant a une valeur communicative différente de son gouverneur, les règles de correspondance l'obligent généralement à s'émanciper dans le champ correspondant à sa valeur communicative. Seuls l'objet direct et les dépendants verbaux connaissent des restrictions interdisant l'émancipation dans un champ correspondant à leur valeur communicative.

Un dépendant verbal ouvre toujours un domaine enchâssé indépendamment du champ où il a été placé.

Si le dépendant verbal est un complémenteur, il ouvre un domaine pour les complétives, qui contient un champ CF pour lui-même et un deuxième champ pour le verbe régi. Ce deuxième champ diffère en fonction de l'endroit où le complémenteur a ouvert son domaine : dans le TopF, le domaine complétif consiste en CF et TopF, dans les tailF de CF et tailF et dans le Nachfeld, de CF et Nachfeld. Seul quand le complémenteur se trouve dans le focF, le domaine qu'il crée est plus complexe : il contient un CF suivi des trois champs majeurs : TopF, FocF et tailF.

Les dépendants d'un gouverneur qui se trouve dans le domaine complétif doivent obligatoirement se placer à l'intérieur de ce domaine, et ne peuvent donc pas s'émanciper.

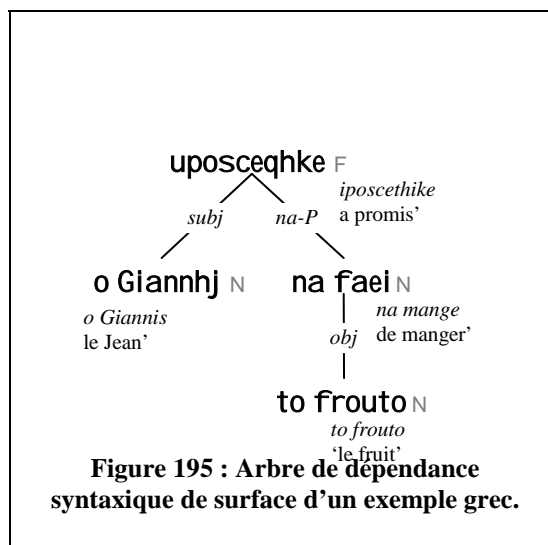
### 7.1.2 Illustration

Illustrons les règles que nous avons proposées par quelques exemples.

Comme d'habitude, notre point de départ est un arbre de dépendance syntaxique de surface augmenté d'un marquage communicatif. L'arbre de dépendance syntaxique que nous présentons dans la Figure 195 représente un verbe ayant un dépendant nominal sujet et un dépendant verbal (verbe en *-na*). La tête verbale est marquée comme focus, et les dépendants, qui forment un groupe communicatif avec leur gouverneur, sont marqués neutre.

---

<sup>236</sup> Comme en français, l'ordre de placement des clitiques n'est pas libre, et pour un traitement complet, il faudra envisager une séparation plus fine de ce champ.



Nous obtenons, après applications des règles de linéarisation pour le grec, les trois linéarisations suivantes que nous présentons en (281).

- (281) a. [ *md* [ *ed* oGiannhj<sub>vf</sub> uposceqhke<sub>v</sub> [ *ed* nafaei<sub>v</sub> tofrouto<sub>nf</sub> ]<sub>nf</sub> ]<sub>F</sub> ]<sub>i</sub>  
 b. [ *md* [ *ed* uposceqhke<sub>v</sub> [ *ed* nafaei<sub>v</sub> tofrouto<sub>nf</sub> ]<sub>nf</sub> oGiannhj<sub>nf</sub> ]<sub>F</sub> ]<sub>i</sub>  
 c. [ *md* [ *ed* uposceqhke<sub>v</sub> o Giannhj<sub>nf</sub> [ *ed* na faei<sub>v</sub> to frouto<sub>nf</sub> ]<sub>nf</sub> ]<sub>F</sub> ]<sub>i</sub>  
 'Jean a promis de manger le fruit.'

Dans la première linéarisation, la tête verbal *uposceqhke* 'a promis' se place dans le FocF et ouvre un domaine offrant des places pour ses dépendants. A cause de leur marquage communicatif N, tous les dépendants restent à l'intérieur du domaine et il n'y a pas d'émancipation. Ne pouvant se placer dans le Vorfeld, le dépendant verbal *na faei* 'de manger' ouvre un nouveau domaine dans le Nachfeld : *na* se place dans le Na-feld et le verbe dans le VerbF. Le dépendant nominal de *na faei* 'de manger' *to frouto* 'le fruit' n'aura pas d'autres choix que de se placer dans le Nachfeld de son gouverneur car il a le même marquage communicatif (donc pas d'émancipation) et qu'il ne peut pas aller dans le Vorfeld (car il est l'objet direct du verbe). Quant au dépendant sujet du verbe principal, il peut se placer dans le Vorfeld (linéarisation a.), ou dans le Nachfeld, avant le dépendant verbal (linéarisation c.) ou après le dépendant verbal (linéarisation a.).

Nous illustrons dans la Figure 196 les trois arbres topologiques obtenus par DepLin.

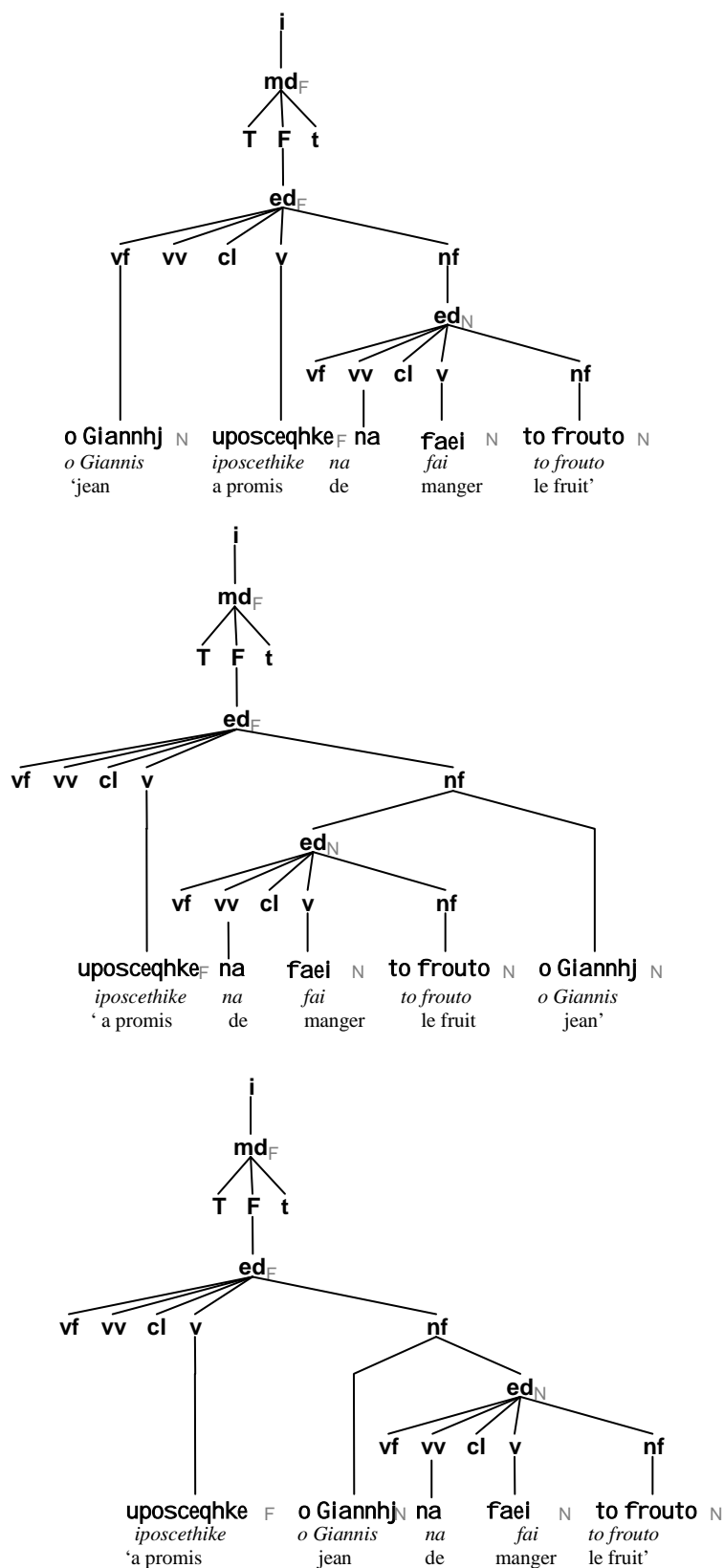
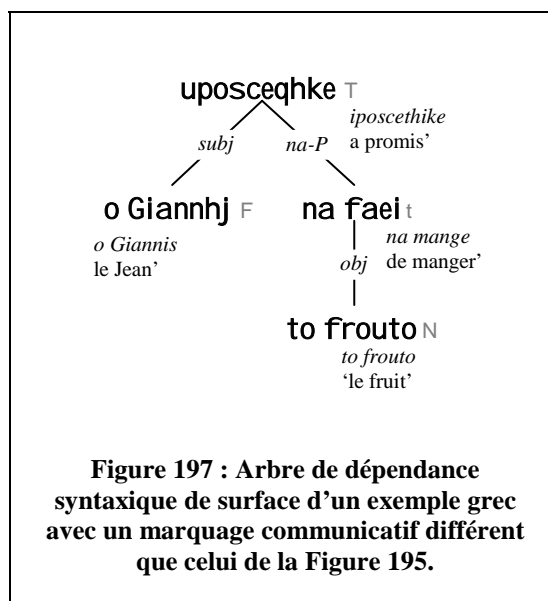
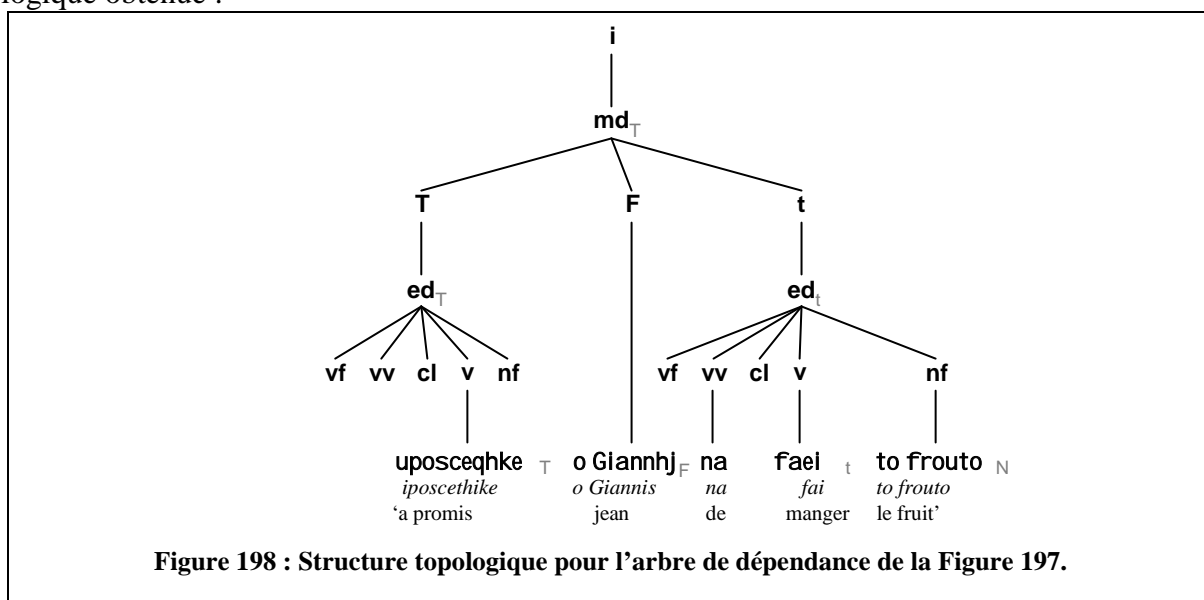


Figure 196 : Les trois structures topologiques possibles pour l'arbre de dépendance de la Figure 195.

Si l'arbre de dépendance a un autre marquage communicatif comme c'est le cas dans la Figure 197, alors les linéarisations obtenues ne sont pas les mêmes. Dans le cas de notre exemple, une seule linéarisation est possible :

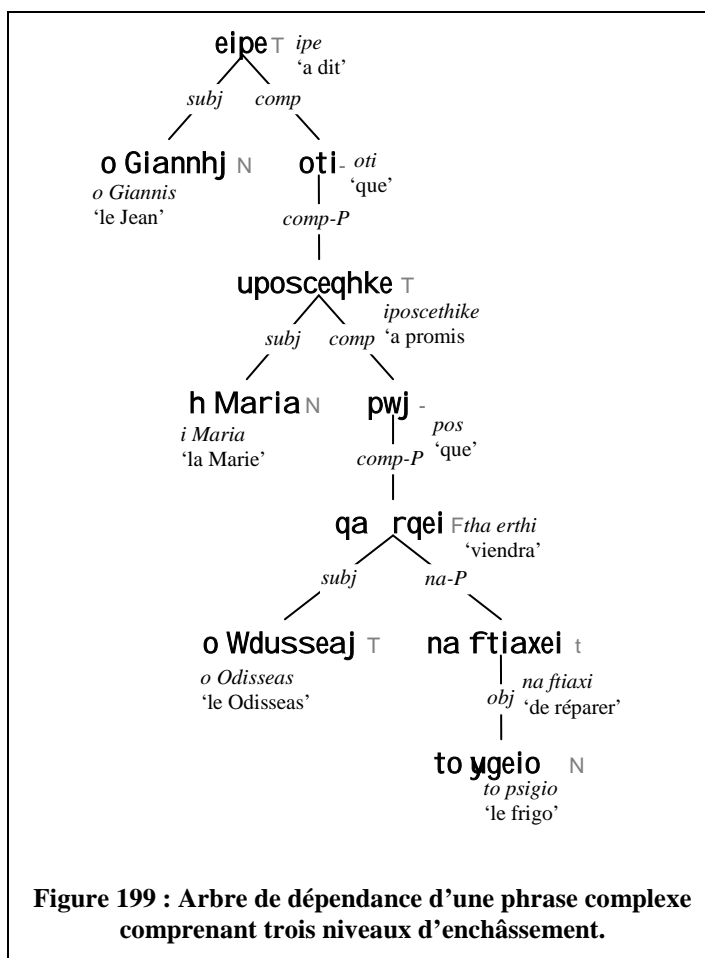


La tête verbale qui a la valeur communicative de topique se place dans le TopF et ouvre un domaine. Cependant, ce domaine ne sera rempli que par le seul verbe car ses deux dépendants vont s'émanciper : le dépendant nominal qui est marqué comme focus s'émancipe et se place dans le champs FocF tandis que le dépendant verbal (avec son dépendant nominal qui lui est neutre) ouvre un domaine dans le tailF (car il est marqué communicativement comme tail) et son dépendant nominal qui est neutre se place dans le Nachfeld. La Figure 198 illustre la structure topologique obtenue :



Nous terminons l'illustration de la topologie du grec avec une phrase un peu plus complexe mais ayant trois niveaux d'enchâssement. Dans l'arbre de dépendance que nous montrons dans la Figure 199, le focus est porté par le verbe enchâssé *qa erqei* 'viendra' qui se trouve dans une phrase complétive qui dépend du verbe *uposceqhke* 'a promis' qui lui-même est dans une complétive dépendant de la tête verbale *eipe* 'a dit'. Un contexte qui permettrait une telle linéarisation serait lorsqu'on reprend l'interlocuteur qui semble avoir mal compris notre propos : ainsi on répète en quelque sorte la situation (la partie topicale) et on contraste sur le verbe qui est au futur (avec par exemple un verbe qui serait au passé).

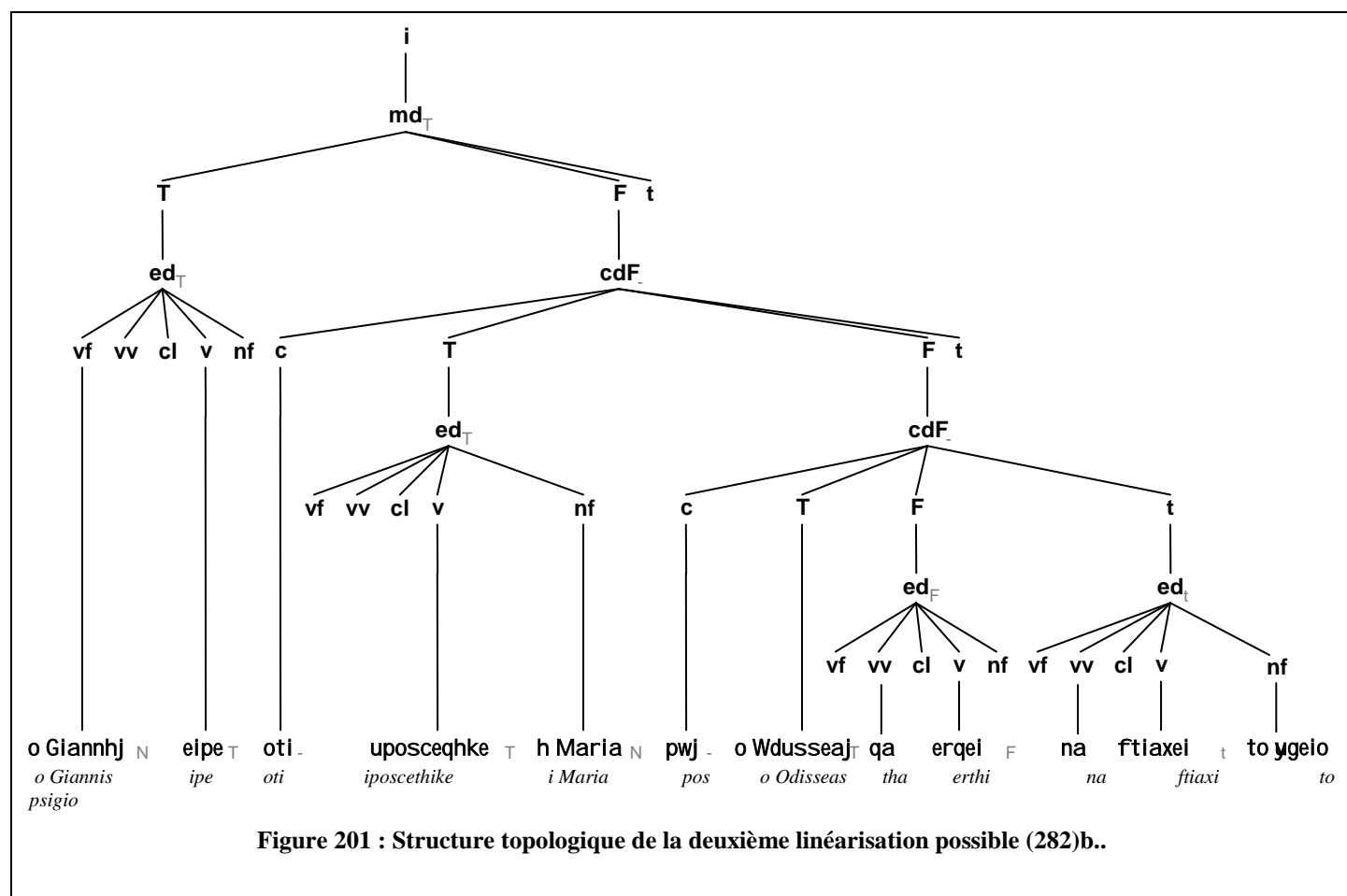
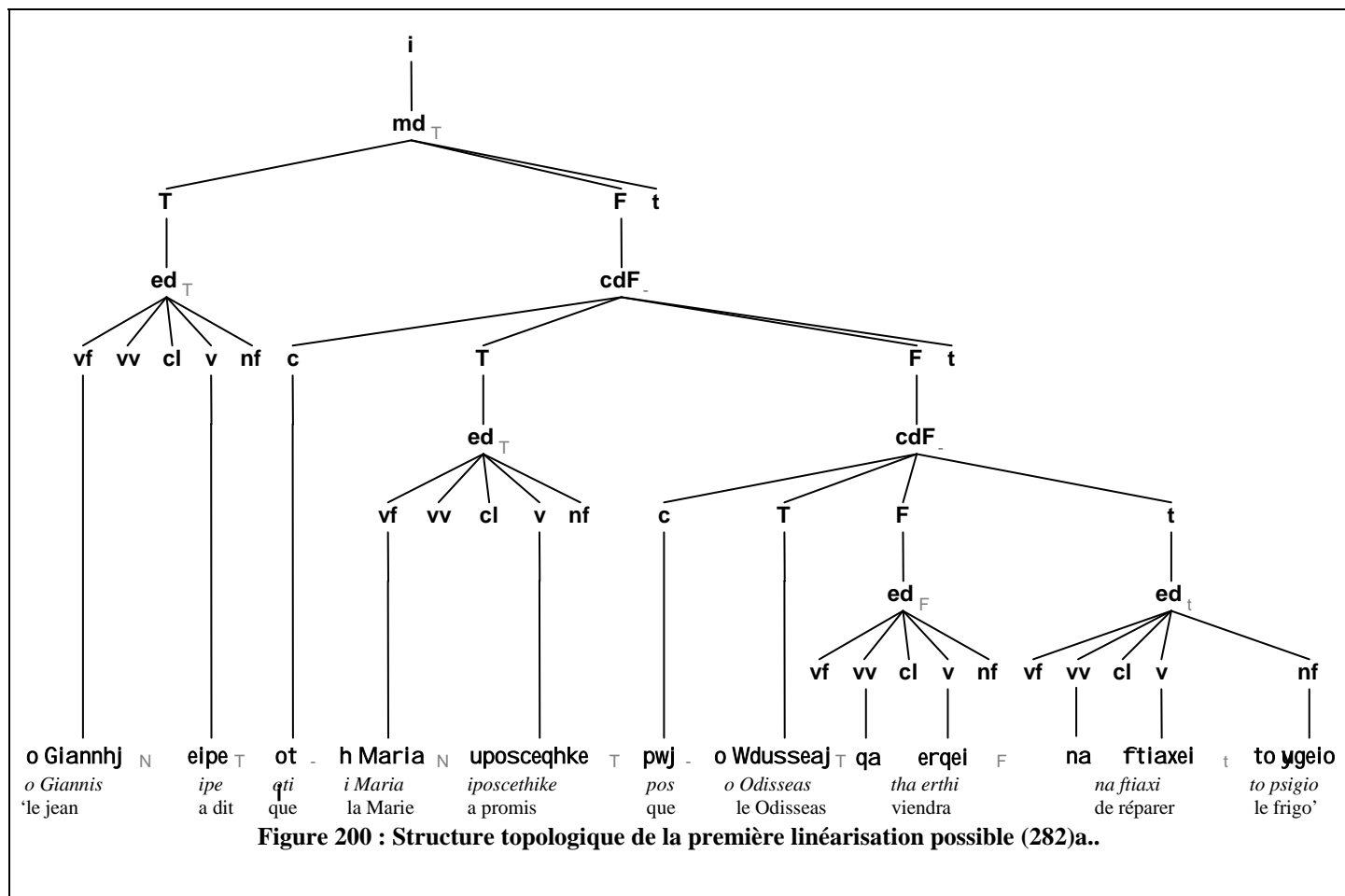




A partir de cette structure de dépendance syntaxique augmentée de son marquage communicatif, nous obtenons les quatre linéarisations possibles que nous indiquons sous (282).

- (282) a. {8} [ md [ ed o Giannhj<sub>vf</sub> eipe<sub>v</sub> ] T [ cdF oti<sub>c</sub> [ ed h Maria<sub>vf</sub> uposceqhke<sub>v</sub> ] T [ cdF pwj<sub>c</sub> o Wdusseaj<sub>T</sub> [ ed qa erqei<sub>v</sub> ] F [ ed na ftiaxei<sub>v</sub> to ygeio<sub>nf</sub> ] t ] F ] F ] i  
 b. {8} [ md [ ed o Giannhj<sub>vf</sub> eipe<sub>v</sub> ] T [ cdF oti<sub>c</sub> [ ed uposceqhke<sub>v</sub> h Maria<sub>nf</sub> ] T [ cdF pwj<sub>c</sub> o Wdusseaj<sub>T</sub> [ ed qa erqei<sub>v</sub> ] F [ ed na ftiaxei<sub>v</sub> to ygeio<sub>nf</sub> ] t ] F ] F ] i  
 c. {8} [ md [ ed eipe<sub>v</sub> o Giannhj<sub>nf</sub> ] T [ cdF oti<sub>c</sub> [ ed h Maria<sub>vf</sub> uposceqhke<sub>v</sub> ] T [ cdF pwj<sub>c</sub> o Wdusseaj<sub>T</sub> [ ed qa erqei<sub>v</sub> ] F [ ed na ftiaxei<sub>v</sub> to ygeio<sub>nf</sub> ] t ] F ] F ] i  
 d. {8} [ md [ ed eipe<sub>v</sub> o Giannhj<sub>nf</sub> ] T [ cdF oti<sub>c</sub> [ ed uposceqhke<sub>v</sub> h Maria<sub>nf</sub> ] T [ cdF pwj<sub>c</sub> o Wdusseaj<sub>T</sub> [ ed qa erqei<sub>v</sub> ] F [ ed na ftiaxei<sub>v</sub> to ygeio<sub>nf</sub> ] t ] F ] F ] i  
 'Jean a dit que Marie a promis que Ulysse viendra fixer le frigo'.

Dans toutes les linéarisations obtenues, la tête verbale qui est marquée comme topique se place dans le champ TopF et ouvre un domaine : son dépendant nominal *o Giannhj* 'Jean' partage la même valeur communicative et ne peut donc sortir du domaine ouvert par son gouverneur. Par contre il peut se placer soit dans le Vorfeld (Figure 200) soit dans le Nachfeld (Figure 202).



Le dépendant verbal de *eipe* 'a dit' est une phrase complétive qui contient des éléments topiques, focus et tail. Le complémenteur doit trouver un champ où il pourrait se placer : les complémenteurs n'ont pas de valeur communicative et peuvent donc ouvrir un domaine complétif dans n'importe quel autre champ vacant. Par ailleurs, la valeur communicative du complémenteur n'étant pas la même que celle du gouverneur, il y aura émancipation de la phrase complétive du domaine principal. Cependant, à cause de règles des descriptions des boîtes (comme par exemple, le fait qu'un focus doit se placer quelque part), le complémenteur finira par ouvrir un domaine dans le FocF (contenant un champ pour le complémenteur et les trois autres champs majeurs) et se placer dans la position CF. Son dépendant verbal qui est topique va ensuite ouvrir un nouveau domaine dans le TopF et se placer dans le VerbField. Le dépendant nominal *h Maria* 'Marie' de ce verbe qui partage le même marquage communicatif, peut soit se placer dans le Vorfeld (Figure 202) soit dans le Nachfeld (Figure 203).

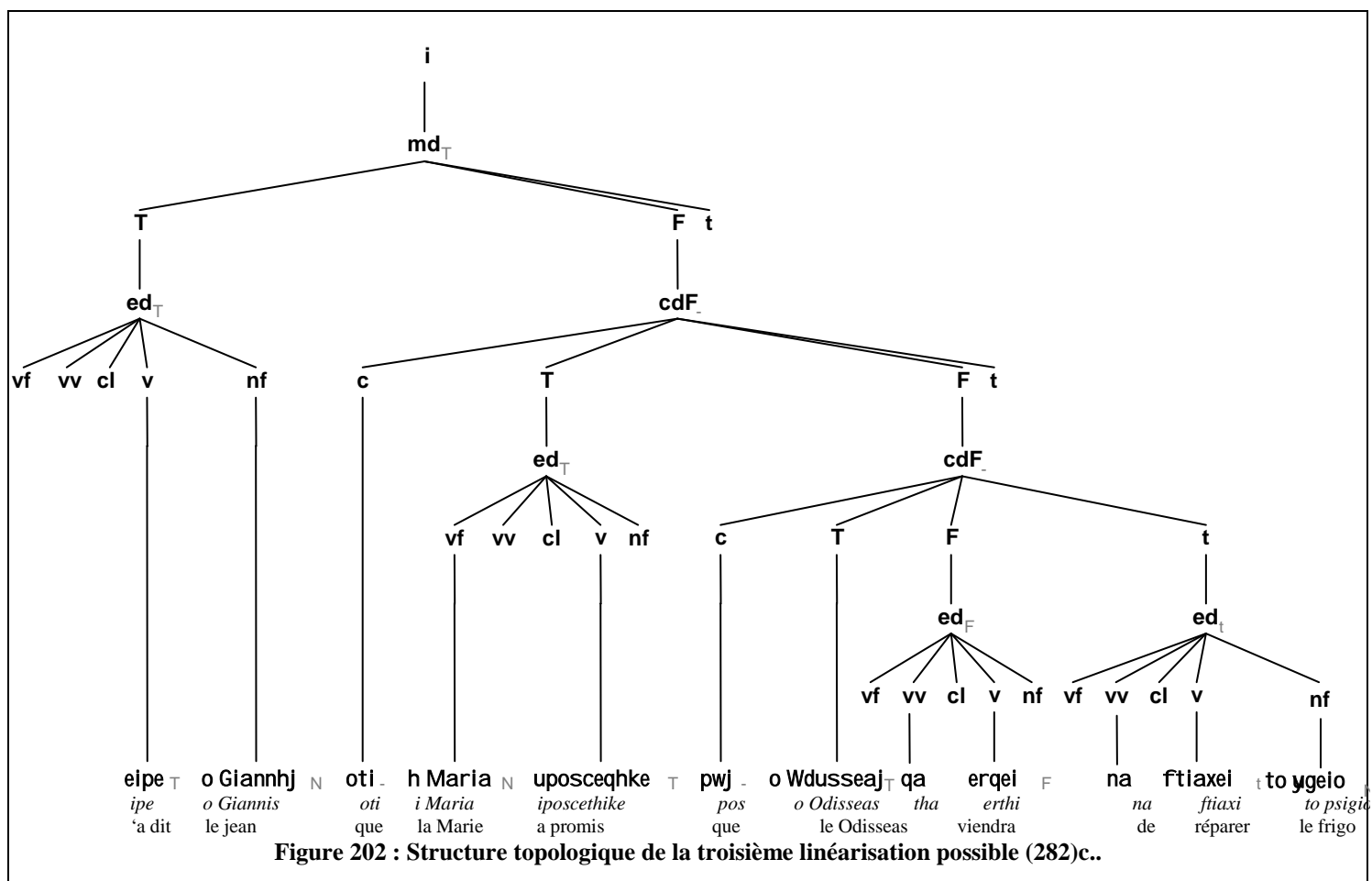


Figure 202 : Structure topologique de la troisième linéarisation possible (282)c..

Pour les mêmes raisons que précédemment, la phrase complétive qui dépend de *uposceqhke* 'a promis' ouvre un domaine complétif contenant une place pour le complémenteur et les trois champs majeurs dans le FocF de son gouverneur. Le complémenteur se place dans le CF du domaine créé et son dépendant verbal *qa erqei* 'viendra' qui porte la valeur de focus ouvre un nouveau domaine dans le FocF du domaine complétif et se place dans le VerbF de ce domaine créée (à noter que la partie *qa* indiquant la valeur future du verbe se place dans le Na-feld). Le dépendant nominal de *qa erqei* 'viendra' *o Wdusseaj* 'Ulysse' marqué comme topique s'émancipe du domaine du gouverneur et se place dans le TopF du CF qui lui est accessible (il s'agit de sa seule place possible).

Enfin, le dépendant verbal de *qa erqei* 'viendra' qui est tail va aussi s'émanciper et se placer dans le tailF du CF qui lui est accessible. Le dépendant nominal *na ftiaxei* 'de réparer' qui partage la même valeur communicative se placera dans la seule place qui lui est accessible, le Nachfeld du domaine de son gouverneur.

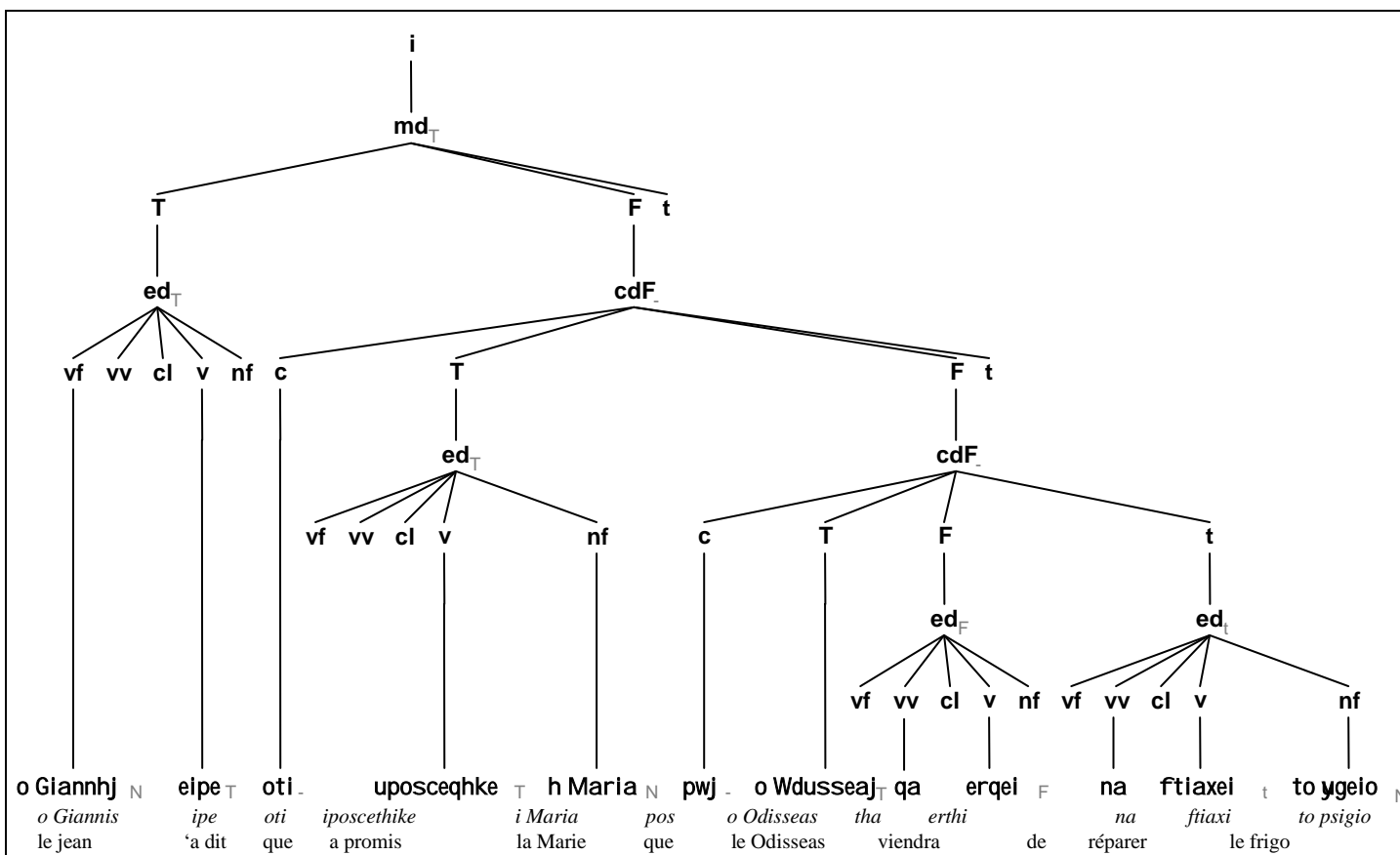


Figure 203 : Structure topologique de la quatrième linéarisation possible (282) d

# Conclusion

# Conclusion

*Mais c'est la vérité<sup>237</sup> !*

Le travail que nous avons présenté dans cette thèse a quatre facettes.

La première est descriptive : nous avons tenté de rediscuter les données sur l'ordre des mots en allemand, l'ordre des mots de l'allemand continuant à poser des problèmes importants, quel que soit le formalisme de description choisi.

La deuxième est théorique. L'aspect théorique comprend une analyse de l'allemand qui distingue l'ordre que prennent les mots et la structure de dépendance d'une phrase. Nous revisitons également la notion de constituant et nous intégrons cette formalisation dans la théorie Sens-Texte en tant que module syntaxique de surface, remplaçant ainsi le niveau habituel de morphologie profonde par une structure de constituants topologiques.

La troisième est formelle : à partir du modèle topologique, nous avons proposé une formalisation du modèle topologique qui permet une description élégante du cadre dans lequel l'ordre des mots peut varier. Nous avons montré que cette formalisation peut être directe (Gerdes & Kahane 2001, TUG) mais aussi intégrée dans des formalismes existants (LFG, TAG). En allemand, le choix entre les linéarisations permises par le modèle topologique est guidé aussi bien par la structure communicative que par les relations syntaxiques qu'entretiennent les mots. Nous montrons les possibilités et les limites pour rendre compte de ces phénomènes dans le formalisme TAG et dans la description d'une TAG à l'aide d'une méta-grammaire. Puis, tout en gardant les avantages du formalisme TAG, nous présentons un modèle de description syntaxique lexicalisé (TUG), une extension des TAG multiples, qui permet de surmonter les déficiences de TAG.

Enfin, la dernière facette de cette thèse est informatique (DepLin, méta-grammaire TAG). Nous avons implémenté une méta-grammaire des schèmes verbaux (donc phrastiques) de l'allemand. Nous avons également programmé un linéarisateur (DepLin), une implémentation directe de notre modèle topologique. Cette implémentation a permis d'affiner la grammaire de la linéarisation de l'allemand, et d'étudier l'applicabilité de ce modèle à d'autres langues. Nous avons également étudié la possibilité de traduction de cette approche dans d'autres formalismes ; ceci a abouti à une implémentation du modèle topologique en LFG.

Nous espérons que les idées défrichées dans ce travail contribueront à des grammaires de couverture plus large et des analyses linguistiques plus approfondies de certains phénomènes syntaxiques que celles proposées dans cette thèse. En ce qui concerne l'aspect descriptif, l'analyse topologique du syntagme nominal reste embryonnaire et il serait intéressant de l'explorer d'avantage. Pour l'aspect appliqué, l'implémentation du formalisme TUG permettrait aussi, en analyse, un traitement plus efficace que notre implémentation actuelle en LFG.

Si on me demandait quelle est l'idée la plus importante que j'ai retenue de cette thèse, je répondrais par *weniger ist mehr* 'moins est plus'. Cela veut dire que je prône dorénavant le minimalisme, en ce sens que je pense que les tâches les plus restreintes sont les plus intéressantes. Je n'aurais jamais pensé avant d'entamer cette thèse qu'une tâche aussi limitée que l'analyse des ordres linéaires possibles d'un arbre de dépendance syntaxique de surface dans la langue que je connais le mieux puisse me surprendre et me fasciner à ce point, et me procurer l'impression d'avoir saisi un détail sur ma langue, sur la langue.

Je suis maintenant convaincu que si on trouve deux phénomènes similaires dans deux langues, il faut commencer par comprendre le phénomène de la première langue en détail et optimiser et

---

<sup>237</sup> Argument ultime utilisé par Igor Mel'čuk dans des discussions linguistiques quand même des preuves sur l'ergativité en Lezgien n'ont pas su convaincre son interlocuteur.

formaliser son analyse avant de passer au phénomène de la deuxième langue. De même, si une structure linguistique représente deux choses à la fois, il ne faut pas se réjouir du fait qu'on puisse tout représenter, mais il faut plutôt chercher à en faire deux structures. Une bonne description d'un seul phénomène d'une seule langue est toujours universelle.





## Références bibliographiques

- Abb, B. 1994. *Die Generierung deutscher Nominalstrukturen. Eine funktionale Analysis in der HPSG-Basisgrammatik eines kognitiv orientierten Sprachproduktinssystems*, Schriften des Graduiertenkolleg des Fachbereichs Informatik der Universität Hamburg, Hamburg.
- Abeillé, A. 1991. *Une grammaire lexicalisée d'arbres adjoints pour le français : application à l'analyse automatique*, Thèse de Doctorat, Université Paris 7.
- Abeillé, A. 1993. *Les nouvelles syntaxes*, Paris, Armand Colin.
- Abeillé, A. 1994. « Syntax or semantics: MC-TAGs vs. Synchronous TAG », *Computational Intelligence*, 10:4.
- Abeillé A. 1998. « Verbes à montée et auxiliaries dans une grammaire d'arbres adjoints », Baschung K. (éd) *LINX n°39 Modèles linguistiques : convergences, divergences*.
- Abeillé A. 1999. *Théories syntaxiques, grammaire française et traitements automatiques des langues* Thèse d'habilitation, Université Paris 7.
- Abeillé A. & O. Rambow. 2000. « Tree Adjoining Grammar: An Overview » A. Abeillé; & O. Rambow (éds.) *Tree Adjoining Grammars: Formalisms, Linguistic Analyses and Processing*. CSLI Publications, Stanford, CA.
- Abeillé, A. 2002. *Une grammaire électronique du français*, CNRS Editions
- Abraham W.O.G. (éd.) 1985. *Erklärende Syntax des Deutschen*. Studien zur deutschen Grammatik, N° 25, Niemeyer, Tübingen.
- Abraham W.O.G. 1994. *Deutsche Syntax im Sprachenvergleich. Grundlegung einer typologischen Syntax des Deutschen*. Studien zur deutschen Grammatik 34, Narr, Tübingen.
- Abraham, W.O.G. 1999 « Grammaticale Miszellen für Ingerid Dal », in Askedal, J.O. (ed), *Historische germanische und deutsche Syntax. Akten des internationalen Symposiums anlässlich des 100. Geburtstages von Ingerid Dal*, Oslo, 27.9.-1.10.1995, Peter Lang, Francfort-sur-le-Main, pp. 87-132
- Abraham, W. O. G. 1999. «Grammaticale Miszellen für Ingerid Dal», in Askedal, J.O. (ed), *Historische germanische und deutsche Syntax. Akten des internationalen Symposiums anlässlich des 100. Geburtstages von Ingerid Dal*, Oslo, 27.9.-1.10.1995, Peter Lang, Frankfurt am Main, pp. 87-132.
- Abraham, W. O. G. et L. Molnárfi 2001 « German Clause Structure under Discourse Functional Weight: Focus and Antifocus » W.O.G. Abraham et C. Jan-Wouter Zwart *On Formal German(ic) Typology*, Benjamins, Amsterdam.
- Abraham, W. O. G. 2001 *VP-internal subjects as 'unaccusatives': Burzio's 'object account' vs. the 'perfectivity account'* GLAC7, Banff, Canada.
- Bech, G. 1955. *Studien über das deutsche Verbum infinitum*, 2nd edition 1983, Linguistische Arbeiten, Nr. 139, Niemeyer, Tübingen.

- Becker, T., A.K. Joshi, O. Rambow. 1991. «Long Distance Scrambling and Tree Adjoining Grammars» *Actes de ACL-Europe*. Berlin, 21-26.
- Becker, T., M. Niv, O. Rambow 1992. *The Derivational Generative Power of Formal Systems or Scrambling is Beyond LCFRS*. IRCS Rapport technique 92-38, IRCS, University of Pennsylvania.
- Becker, T., P. Lopez 2000. « The application of HPSG-to-TAG compilation techniques ». W. Wahlster (éd.) *Verbmobil: Foundations of Speech-To-Speech Translation*, LNCS. Springer, Berlin.
- Becker, T. 2002. « Practical, Template-Based Natural Language Generation with TAG ». *Actes de l'Atelier TAG+6*, Venise, Italie.
- Behagel, O. 1909. « Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern », *Indogerm. Forsch.* 25, pp.110-142.
- Berman, J. 1998 « On the Syntax of Correlative *Es* and Finite Clauses in German: an LFG Analysis », in Tibor Kiss et Detmar Meurers (eds.), *Actes de ESSLI-98 Workshop on Current Topics in Constraint-Based Theories of Germanic Syntax*, Sarrebruck.
- Bod R. 1998. *Beyond Grammar: An Experience-Based Theory of Language*, CSLI Publications, Stanford University Press.
- Bonami O. 2001. « Inversion du sujet, constituance et ordre des mots » dans Marandin J.-M. (eds.) *Cahier Jean-Claude Milner*, Verdier, pp. 99-174
- Bonhomme, P. et P. Lopez, 2000. « TagML: XML encoding of Resources for Lexicalized Tree Adjoining Grammars ». *Actes de LREC 2000*, Athènes, Grèce.
- Bresnan, J. 1982. *The mental Representation of Grammatical Relations*, MIT Press.
- Bresnan, J., R. M. Kaplan, S. Peters, A. Zaenen. 1982. « Cross-serial Dependencies in Dutch », *Linguistic Inquiry* 13(4): 613-635.
- Bresnan, J. 2001. *Lexical-Functional Syntax*. Blackwell, Oxford.
- Bröker, N. 1998. «Separating Surface Order and Syntactic Relations in a Dependency Grammar ». *Actes de COLING-ACL 98*.
- Büring, D. 1995. *The 59th Street Bridge Accent*. Thèse de doctorat. Universität Tübingen.
- Butt, M., T. H. King, M-E Nino, and F. Segond. 1999. *A Grammar Writer's Cookbook*. Stanford, CSLI Publications.
- Butt, M., H. Dyvik, T. H. King, H. Masuichi & C. Rohrer 2002 « The Parallel Grammar Project » *Proceedings of COLING-2002, Workshop on Grammar Engineering and Evaluation*, Taipei, Taiwan.
- Candito, M.-H., 1996 « A Principal Based Hierarchical Representation of LTAG », *Actes de COLING*, Copenhagen.
- Candito, M.-H., 1998. « Building Parallel LTAGs: Application to French and Italian », *Actes ACL-COLING*, Montréal, Canada.

- Candito M.-H., S. Kahane. 1998a. « Can the TAG Derivation Tree Represent a Semantic Graph? An Answer in the Light of Meaning-Text Theory », *Actes du Fourth Workshop on Tree-Adjoining Grammars and Related Frameworks (TAG+4)*.
- Candito M.-H., S. Kahane. 1998b. « Defining DTG Derivations to get Semantic Graphs », *Actes du Fourth Workshop on Tree-Adjoining Grammars and Related Frameworks (TAG+4)*.
- Candito, M.H., 1999. *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées. Application au français et à l'italien*. Thèse de Doctorat, Université Paris 7.
- Chiang, D. 2001. « Constraints on Strong Generative Power » *Meeting of the Association for Computational Linguistics*, 124-131
- Choi, H-W. 1999. *Optimizing Structure in Context – Scrambling and Information Structure*, CSLI, Stanford.
- Chomsky, N. 1957. *Syntactic Structures*, La Hague, Mouton.
- Clément L., K. Gerdes et S. Kahane 2002 « An LFG-based Grammar for German Based on the Topological Model » *Actes de LFG 2002*, Athènes.
- Coch, J., 1998. « MultiMeteo: Interactive Weather Report Generation » *Proceedings of the 9th International Natural Language Generation Workshop (INLG'98)* Niagara-on-the-Lake, Canada.
- Comrie, B. 1977. « In Defense of Spontaneous Demotion: the Impersonal Passive », Cole, P. & J.M. Sadock (eds.) *Syntax and Semantics 8 Grammatical Relations*. Academic Press.
- Cori M., J.M. Marandin 1993 « Grammaires d'arbres polychromes », *TAL*, volume 34, numéro 1, pp. 101-132.
- Drach, E. 1937. *Grundgedanken der deutschen Satzlehre*, Diesterweg, Frankfurt.
- Duchier, D., R. Debusmann 2001. « Topological Dependency Trees: A Constraint-based Account of Linear Precedence », dans *Proceedings ACL 2001*, Toulouse.
- Duden, 1984. *Grammatik der deutschen Gegenwartssprache*, volume 4, Dudenverlag, 3<sup>e</sup> édition, Mannheim, Vienne, Zurich,.
- Duden, 1995. *Grammatik der deutschen Gegenwartssprache*, volume 4, Dudenverlag, 5<sup>e</sup> édition (complètement révisée), Mannheim, Vienne, Zurich.
- Evers Arnoldus, 1975, *The Transformational Cycle in Dutch and German*. Thèse de doctorat, University of Utrecht.
- Frank, R. 1992. *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. Thèse de Doctorat, Université de Pennsylvanie.
- Frey F. 2000. « Über die syntaktische Position der Satztopiks » *ZAS Papers in Linguistics 20*, Zentrum für Allgemeine Sprachwissenschaft, Berlin, pp.137-172
- Gaiffe B., B Crabbé, A. Roussanaly 2002. « A New Metagrammar Compiler ». *Actes de l'Atelier TAG+6*, Venise, Italie.

- Genzmer, H. 1998. *Sprache in Bewegung. Eine deutsche Grammatik*, Suhrkamp, Francfort-sur-le-Main.
- Gerdes K., P. Lopez, 2000. « Shared Semantic Representations for LTAG », *Proceedings of the ATALA Workshop on Representation and Treatment of Syntactic Ambiguity in Natural Language Processing*. Paris.
- Gerdes K. 1998. *Le cas de l'allemand –Représentation du système casuel de l'allemand dans une grammaire d'arbres adjoints*, Mémoire de D.E.A, Université Paris 7/
- Gerdes K. 2001 « TAG and Topology –Problems and Proposals for German », Actes de *Formal Grammars / Mathematics of Language FG/MOL 2001*, Helsinki. (version d'article soumis)
- Gerdes, K. *DepLin: Implementation of a Dependency Linearizer Based on Topological Phrase Structure*, <http://talana.linguist.jussieu.fr/~kim/DepLin/>
- Gerdes K. 2002. « DTAG ? –Attempt to Generate a Useful TAG for German Using a Metagrammar », *Actes de TAG+6*, Venise.
- Gerdes, K., S. Kahane 2001a. « Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy », *Actes d'ACL 2001*, Toulouse.
- Gerdes K., S. Kahane 2001b. « Pas de syntaxe sans prosodie : illustration par l'allemand », *Journées Prosodie*, Grenoble
- Gerdes K., S. Kahane 2001c. « Description of German syntax based on a topological Hierarchy », *Proceedings of Seventh Germanic Linguistics Annual Conference, GLAC7*, Banff.
- Gerdes K. & S. Kahane 2002 « Phrasing it differently » *Selected Lexical and Grammatical Topics in the Meaning-Text Theory. In Honour of Igor Mel'čuk*. Benjamins, Amsterdam, Philadelphia.
- Gerdes K. & H.-Y. Yoo 2003 « The fields on the way to prosody - Alternatives to phrase structure based approaches to prosody », *Actes de International Conference on Phonetic Sciences*, Barcelone, à paraître.
- Gibbon, D. 1998. « Intonation in German », Hirst, D.J. - Di Cristo, A. (éds.). *Intonation Systems. A Survey of Twenty Languages*. Cambridge, Cambridge University Press.
- Grewendorf, G., F.Hamm, W. Sternefeld. 1989. *Sprachliches Wissen*, Francfort-sur-le-Main, Suhrkamp, 3e édition.
- Haider, H. 1985. « The Case of German », Jindřich Toman (ed.), *Studies in German Grammar, Studies in Generative Grammar, N° 21*, Foris Publications, Dordrecht, 1985. pp. 23-64
- Haider, H. 1993. *Deutsche Syntax – Vorstudien zur Theorie einer projektiven Grammatik*, Tübingen, Narr.
- Han C.-H., J. Yoon, N. Kim & M. Palmer. 2000. *A Feature-Based Lexicalized Tree Adjoining Grammar for Korean*. IRCS Report 00-04, IRCS, University of Pennsylvania.
- Harbusch, K. & G. Kempen. 2000. « Complexity of Linear Order Computation in Performance Grammar, TAG and HPSG », *Actes de l'Atelier TAG+5*, Paris, 101-106.

- Höhle, T. 1996. « The W...W Construction: Appositive or Scope Indicating? » U. Lutz and G. Müller (eds.), *Papers on Wh-Scope Marking, Arbeitspapiere des Sonderforschungsbereichs*, 340. University of Stuttgart and University of Tübingen. 249-270.
- Holton, D., P. Mackridge, I. Filippaki-Wharbuton 1997. *Greek Grammar: A Comprehensive Grammar of the Modern Language*. London, Routledge.
- Hook, P. E. & A. Manaster-Ramer. 1985. « The Verb-Second Constraint in Kashmiri and Germanic: Toward a Typology of V-2 Languages. » *Germanic Linguistics: Papers from a Symposium at the University of Chicago, 24 April 1985*. Jan Terje Faarlund, Ed. Bloomington: Indiana University Linguistics Club. pp. 46-58.
- Hudson R. 2000. « Discontinuity », S. Kahane (ed.), *Grammaires de Dépendances, T.A.L.*, 41(1): 15-56, Hermès, Paris.
- Huybregts, R. 1976. « Overlapping Dependencies in Dutch », *Utrecht Working Papers in Linguistics*, 1, 24-65.
- IDS-Grammatik : Zifonun, G. et al. 1997. *Grammatik der deutschen Sprache*, Schriften des Instituts für deutsche Sprache, de Gruyter, Berlin/New York, trois volumes.
- Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge Mass., MIT Press
- Jacobs, J. 1988. « Probleme der freien Wortstellung im Deutschen ». *Sprache und Pragmatik*. Arbeitsberichte 5. pp. 8-37
- Joshi, A., Kallmeyer, L. 1999. « Factoring Predicate Argument and Scope Semantics: Underspecified Semantics with LTAG ». *Proceedings of the 12th Amsterdam Colloquium*.
- Joshi, A., Levy, L. et M. Takahashi, 1975. « Tree Adjunct Grammars » *Journal of Computer and System Sciences* 10, Academic Press. pp.136-163
- Kahane, S. 1997. «Bubble Trees and Syntactic Representations ». *Actes du 5th Meeting of the Mathematics of the Language*, DFKI, Saarbrücken.
- Kahane, S. 2001. « A Fully Lexicalized Grammar for French based on Meaning-Text theory", *Computational Linguistics, Actes. CICLing 2001*, Mexico.
- Kahane S., A. Nasr, O. Rambow. 1998. « Pseudo-Projectivity: a Polynomially Parsable Non-Projective Dependency Grammar », *COLING-ACL'98*, Montreal, 646-52.
- Kahane S. & Mel'čuk I. 1999 « La synthèse sémantique ou la correspondance entre graphes sémantiques et arbres syntaxiques – Le cas des phrases à extraction en français contemporain » *T.A.L.* 40:2 pp.25-85
- Kahane S., M.-H. Candito & Y. de Kercadio. 2000. « An Alternative Description of Extractions in TAG », *Actes de l'Atelier. TAG+5*, Paris, 115-22.
- Kahane S. 2002 *Grammaire d'Unification Sens-Texte* Thèse d'habilitation, Université Paris 7.
- Kallmeyer, L. 1999. *Tree Description Grammars and Underspecified Representations*. Thèse de Doctorat, Universität Tübingen.

- Kallmeyer, L. 2000. « Scrambling in German and the Non-locality of Local TDGs » *actes de l'atelier TAG+5*, Paris, 129 - 134.
- Kallmeyer, L. & A. Joshi, 1999, « Factoring Predicate Argument and Scope Semantics: Underspecified Semantics with LTAG », Paul Dekker (ed.) *les Actes du 12th Amsterdam Colloquium*, December 1999, 169-174.
- Kaplan, R. & A. Zaenen 1989. « Long Distance Dependencies, Constituent Structure and Functional Uncertainty » M. Baltin, A. Kroch (éds.) *Alternative Conceptions of Phrase Structure*, University of Chicago Press.
- Kathol, A. 1995 « Verb-‘Movement’ in German and Topological Fields », A. Dainora et al. (eds.), *Papers from the 31<sup>st</sup> regional Meeting, Chicago Linguistic Society*, volume 31, p. 231-245, Chicago.
- Kathol, A., 1995. *Linearization-based German Syntax*, Thèse de Doctorat, Ohio State University.
- Kroch, A. & A. Joshi 1985. « Some Aspects of the Linguistic Relevance of Tree Adjoining Grammars », *Rapport technique*, Dept. CIS, Université de Pennsylvanie, Philadelphie.
- Kuhn, J. 2001. *Formal and Computational Aspects of Optimality-theoretic Syntax*. thèse de doctorat, IMS, Université de Stuttgart.
- Kuthy, K. de, 1998. « Splitting PP Arguments from NPs – An Argument Raising Approach and its Interaction with Lexical Semantics », in T. Kiss et D. Meurers (eds.), *Proceedings of the ESSLI-98 Workshop on Current Topics in Constraint-Based Theories of Germanic Syntax*, Sarrebruck.
- Lambrecht, K. 1994. *Information Structure and Sentence Form: Topic, Focus and the Mental Representations of Discourse Referents*, Cambridge Studies in Linguistics 71, Cambridge University Press.
- Lasnik, H. 1999 « Minimalism », R.A Wilson & F.C. Keil (eds). *The MIT Encyclopedia of the Cognitive Sciences* MIT Press, Cambridge Massachusetts.
- Lenerz J. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*, Tübingen, Narr.
- Lenerz, J. 1993. « Zur Syntax und Semantik deutscher Personalpronomina » M. Reis (ed.) *Wortstellung und Informationsstruktur* ». Tübingen, Niemeyer (Linguistische Arbeiten, 306), 117-153.
- Lombardo V. and Lesmo L. 2000. « A formal Theory of Dependency Syntax with Empty Units », Kahane S. (éd) *Grammaires de Dépendances, T.A.L.*,41.1.
- Lopez, P. 1999. *Robust Parsing with Lexicalized Tree Grammars for Spoken Dialogue Systems*, Thèse de Doctorat, Université Nancy 1.
- Marandin J.-M. 2001. « L'hypothèse des sites en syntaxe » dans Marandin J.-M. (eds.) *Cahier Jean-Claude Milner*, Verdier, pp. 175-227.
- Mel'čuk, I. 1988. *Dependency Syntax*. New-York, State University of New York Press.

- Mel'čuk, I. 2001. *Communicative Organization in Natural Language (The Semantic-Communicative Structure of Sentences)*. Amsterdam, Benjamins.
- Mel'čuk, I. (à paraître). *Do Cases Have Meanings?*
- Mel'čuk, I. & N. Pertsov. 1987. *Surface syntax of English – A Formal Model within the Meaning-Text Framework*, Benjamins, Amsterdam.
- Meurers, W. D. 1999. « Raising Spirits (and assigning them case) », *Groninger Arbeiten zur Germanistischen Linguistik (GAGL)* 43, 173-226.
- Meurers, W. D. 2000. *Lexical Generalizations in the Syntax of German Non-Finite Constructions*. Arbeitspapiere des SFB 340, N° 145. Thèse de doctorat de l'Université de Tübingen.
- Milićević, J. (à paraître) « Linear Placement of Serbian Clitics: A Description within a Dependency-Based Approach », in A. Polguère et L. Wanner (eds.), *Topics in Dependency Syntax*, Benjamins, Amsterdam.
- Miller, P. H. 1999. *Strong Generative Capacity: The Semantics of Linguistic Formalism*. Stanford, CSLI Publications.
- Müller, St. 1996 « The Babel-System – an HPSG Prolog implementation » *Proceedings of the Fourth International Conference on the Practical Application of Prolog*, London, pp. 263-277.
- Müller, St. 1997 *Head-Driven Phrase Structure Grammar für das Deutsche*, Entwurf, DFKI, Sarrebruck.
- Müller, St. 1999. *Deutsche Syntax Deklarativ: Head-Driven Phrase Structure Grammar für das Deutsche*, Linguistische Arbeiten 394, Tübingen, Niemeyer.
- Müller, St. 2000. « The passive as Lexical Rule », dans D. Flickinger and A. Kathol (éds.) *Proceedings of the 7<sup>th</sup> International HPSG Conference*, UC Berkeley.
- Müller, St. & W. Kasper 2000. « HPSG analysis of German » W. Wahlster (Ed) *VerbMobil: Foundations of Speech-to-Speech Translation*, Artificial Intelligence, Springer, Berlin, Heidelberg, New York, pp. 238-253.
- Nasr, A. 1996. *Un modèle de reformulation automatique fondé sur la Théorie Sens-Texte*, Thèse de doctorat, Université Paris 7, Paris.
- Netter, K. 1996. *Functional Categories in an HPSG for German*, thèse de doctorat, Université de la Sarre, Sarrebruck.
- Office de la langue française, 2000 <http://www.olf.gouv.qc.ca>
- Pollard, C. & I. Sag. 1994. *Head driven phrase structure grammar*, Studies in Contemporary Linguistics, Chicago, London: University of Chicago Press.
- Pollard, C. 1990. « On Head Non-Movement ». H. Bunt & A. van Horck (eds.) *Discontinuous Constituency*, Natural Language Processing, N°. 6, Mouton de Gruyter, Berlin, New York, pp. 279-305.

- Pollard, C. 1994. « Toward a Unified Account of Passive in German ». J. Nerbonne, K. Netter, and C. Pollard (eds.) *HPSG for German*. Stanford: CSLI Publications.
- Poller, P. 1994. « Incremental Parsing with LD/LP-TAGs » *Computational Intelligence*, 10(4):549-562
- Poller, P., & T. Becker 1998. « Two-step TAG Parsing Revisited » *Proceedings of the Fourth International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+4)* IRCS, Université de Pennsylvanie, pp. 143-146.
- Rambow, O. 1994. *Formal and Computational Aspects of Natural Language Syntax*, Institute For Research in Cognitive Science, Thèse de Doctorat, University of Pennsylvania, Philadelphia.
- Rambow, O. & Y-S Lee, 1994. « Word Order Variation and Tree Adjoining Grammar ». *Computational Intelligence* 10(4).
- Rambow, O. and A. Joshi, 1997. « A Formal Look at Dependency Grammars and Phrase-Structure Grammars, with Special Consideration of Word-Order Phenomena ». Wanner L. (éd), *Current Issues in Meaning-Text Theory*. Pinter, London.
- Rambow, O., K. Vijay-Shanker, D. Weir. 1995. « D-Tree Grammars », *Actes d' ACL'95*.
- Reape, M., 1994. « Domain Union and Word Order Variation in German », J. Nerbonne et al. (éds.), *German in Head-Driven Phrase Structure Grammar*, CSLI Lecture Notes, N° 46, Stanford.
- Reis, M. 2000 « On the Parenthetical Features of German Was...W-Constructions and How to Account for Them». In Lutz et. al (éds.) *Wh-Scope Marking*. John Benjamins.
- Rogers, J. & K. Vijay-Shanker. 1994. « Obtaining Trees from their Descriptions: An Application to Tree-Adjoining Grammars » *Computational Intelligence*, 10:401-421
- Ross J.R.. 1967. *Constraints on Variables in Syntax*, Thèse de Doctorat, MIT, Cambridge, Mass.
- Rossi M., A. Di Cristo, D. Hirst, P. Martin, Y. Nishinuma 1981. *L'intonation, de l'acoustique à la sémantique*, Klincksieck, Paris.
- Rossi M., 1999. *L'intonation, le système du français*, Paris, Ophrys.
- Schabes & Joshi 1988. *An Early-Type Parsing Algorithm for Tree Adjoining Grammars*, Rapport technique CIS 88-36, Université de Pennsylvanie, Philadelphie.
- Schabes, Y & S. Shieber 1990. « Synchronous Tree Adjoining Grammars », *Actes de Coling 1990*, Helsinki.
- Schabes, Y & S. Shieber 1994. « An Alternative Conception of Tree-adjoining Derivation », *Computational Linguistics*, vol. 20, n° 1, pp. 91-124.
- Selkirk, E. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, MA: MIT Press.
- Stechow, A. von & W. Sternefeld. 1988. *Bausteine syntaktischen Wissens. Ein Lehrbuch der generativen Grammatik*. Westdeutscher Verlag, Opladen.



- Sternefeld W. 2001. « Semantic vs. Syntactic Reconstruction » C. Rohrer, A. Rossdeutscher et Kamp H (éds) *Linguistic Form and its Computation*. CSLI Publications, Stanford, CA, S. 145-182
- Tesnière, L. 1959. *Eléments de syntaxe structurale*, Klincksieck, Paris.
- The XTAG Research group. 1995. *A lexicalized Tree Adjoining Grammar for English*. IRCS-95-03, Philadelphia, Institute of Research in Cognitive Science.
- Thiersch, C. L. 1978. *Topics in German Syntax*, thèse de doctorat, M.I.T., Cambridge, Mass.
- Truckenbrodt H. 1999. « On the Relation between Syntactic Phrases and Phonological Phrases », *Linguistic Inquiry*, 30:2. pp.219-255.
- Truckenbrodt, H. 1995. *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. Thèse de doctorat, MIT, Cambridge, Mass.
- Tseng J. 2000. « Linearizing Parentheticals in German: The Limitations of Phrase Structure. » *Actes de ESSLLI 2000 Workshop on Linguistic Theory and Grammar Implementation*. Birmingham.
- Uszkoreit H. 1987. *Word Order and Constituent Structure in German*, CSLI Lecture Notes 8, Stanford. CA.
- Vallduví, E. 1992. *The Informational Component*. New York, Garland.
- Vijay-Shanker, K. 1987. *A Study of Tree Adjoining Grammars*, Institute For Research in Cognitive Science, Thèse de doctorat, Université de Pennsylvanie, Philadelphie.
- Wahlster, W. 2000 (ed.) *VERBMOBIL – Foundations of Speech-to-Speech Translation*, Springer.
- Webelhuth, G. 1985. « German is Configurational », *Linguistic Review*, 4, 203-246.
- Weir, D.J. 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms*. Thèse de doctorat, Université de Pennsylvanie.
- Yoo H. 2000. *Etude des variations prosodiques dues à l'organisation pragmatique des énoncés en grec moderne : topique, focus et ordre des mots*, mémoire de DEA, Université Paris 7.



## Table des figures

Figure 1 : les différentes structures de la théorie Sens-Texte .....	3
Figure 2 : les différents structures de la théorie Sens-Texte .....	3
Figure 3 : Un graphe sémantique peut correspondre à deux structures syntaxiques différentes ..	14
Figure 4 : Graphe sémantique et arbre syntaxique de la phrase (18).....	15
Figure 5 : Graphe sémantique et arbre syntaxique de la phrase (19).....	16
Figure 6 : Les trois niveaux sémantique, syntaxique profond et syntaxique de surface.....	18
Figure 7 : Structure de dépendance syntaxique simplifiée de (74).....	33
Figure 8 : Structure de dépendance syntaxique simplifiée de (76).....	34
Figure 9 : Structure de dépendance syntaxique simplifiée de (78).....	35
Figure 10 : Structure de dépendance syntaxique simplifiée de (79).....	36
Figure 11 : Représentation schématique du 59 <sup>th</sup> <i>street bridge accent</i> comme configuration intonative de la phrase allemande. ....	41
Figure 12: Dépendance syntaxique de surface servant de point de départ : .....	44
Figure 13: Les deux représentations syntagmatiques non croisée possibles pour l'arbre de dépendance,.....	45
Figure 14: Représentation syntagmatique illustrant la topicalisation de <i>diesem Mann</i> ‘à ces hommes’ .....	45
Figure 15: Représentation syntagmatique illustrant le scrambling.....	46
Figure 16: Les représentations syntagmatiques de l'extraction lointaine : .....	46
Figure 17: Représentation syntagmatique illustrant un <i>VP-fronting</i> complet.....	47
Figure 18: Représentation syntagmatique illustrant un <i>VP-fronting</i> partiel. ....	47
Figure 19: Représentation syntagmatique illustrant une intraposition complète.....	47
Figure 20: Représentation syntagmatique illustrant une intraposition partielle. ....	47
Figure 21: Représentation syntagmatique illustrant une extraposition du verbe fini <i>versprochen</i> et une intraposition totale de <i>den Roman zu lesen</i> . ....	48
Figure 22: Les 99 possibilités de linéariser l'arbre de dépendance syntaxique de la Figure 12...	49
Figure 23: Illustration du domaine principal de la structure topologique.....	55
Figure 24: Illustration du domaine enchâssé de la structure topologique.....	55
Figure 25 : Dépendance syntaxique de surface servant de point de départ .....	57
Figure 26 : Structure de constituants sans domaines enchâssés.....	58
Figure 27 : Structures topologiques sans domaines enchâssés (respectivement phrases (106) et (107)).....	59
Figure 28 : Structure topologique de la phrase 0 illustrant le brouillage d'argument. ....	60
Figure 29: Structure de constituants avec domaine enchâssé ouvert par <i>zu lesen</i> . ....	60
Figure 30: Structure topologique des phrases (109) illustrant le brouillage d'argument.....	61
Figure 31: Structure topologique complète de la phrase (109)b.....	62
Figure 32: Structure de constituants avec émancipation du dépendant <i>den Roman</i> de <i>zu lesen</i> ...	63
Figure 33: Structures topologiques avec émancipation (la phrase (111) correspond à un cas de VP fronting partiel et la phrase (112) à un cas d'émancipation de l'objet). ....	64
Figure 34 : Arbre de dépendance syntaxique de surface avec un verbe le verbe <i>anfangen</i> ‘crier sur’. ....	64
Figure 35 : Structures topologiques indiquant le placement du préfixe <i>an</i> . ....	65
Figure 36: Structure topologique de la « troisième construction ». ....	69
Figure 37: Les champs du domaine de la complétive ouvert par le complémenteur.....	71
Figure 38: Arbre de dépendance syntaxique de surface avec un verbe le verbe <i>anfangen</i> ‘crier sur’. ....	71
Figure 39: Domaine du complémenteur ouvert dans le Nachfeld : .....	72

Figure 40: Les dépendants nominaux dont le gouverneur se trouve dans un champ complémentateur ne peuvent pas s'émanciper. Cette linéarisation est donc impossible. ....	72
Figure 41: Les dépendants nominaux dont le gouverneur se trouve dans un champ complémentateur ne peuvent pas s'émanciper. Cette linéarisation est donc impossible. ....	72
Figure 42: Domaine spécifique ouvert par la conjonction <i>als</i> 'comme' .....	74
Figure 43: Domaine spécifique ouvert par la conjonction <i>als</i> 'comme' .....	74
Figure 44: Domaine spécifique de la phrase parenthétique. ....	76
Figure 45: Graphe sémantique pour les phrases (141) et (142) .....	77
Figure 46: arbres de dépendance syntaxique de surface des phrases (141) et (142) .....	78
Figure 47: Arbres de dépendance syntaxique de surface des phrases (141) et (142) .....	79
Figure 48 : graphe sémantique de la phrase (163). ....	83
Figure 49: Arbre de dépendance syntaxique de surface de la phrase (163) .....	84
Figure 50: graphe sémantique de la phrase (166) .....	84
Figure 51: arbre syntaxique de la phrase (166) .....	85
Figure 52 : Arbres de dépendance syntaxique de surface de la phrase relative sous (167) .....	86
Figure 53 : Représentation de la structure topologique d'une relative : .....	86
Figure 54 : Représentation de la structure topologique d'une relative : .....	86
Figure 55: Arbre de dépendance syntaxique de surface de la phrase (168) .....	87
Figure 56: Représentation de la structure topologique de la relative dans le cas du pied piping. ....	87
Figure 57: Première étape dans la linéarisation topologique de la partie relative de la phrase (169) .....	88
Figure 58: Placement de la préposition dans le champ complétif pour la linéarisation topologique de la partie relative de la phrase (169) .....	88
Figure 59: Arbre de dépendance syntaxique de surface des phrases (175)et (176) .....	89
Figure 60: Linéarisation topologique sans pied piping (phrase (175)) .....	89
Figure 61: Linéarisation topologique avec intraposition phrase (175) .....	90
Figure 62: Linéarisation topologique avec pied piping (phrase (176)) .....	90
Figure 63: Linéarisation topologique avec pied piping (phrase (176)) .....	90
Figure 64: Les trois structures topologiques différentes pour la phrase (192). ....	99
Figure 65: Analyse prosodique de trois structures topologiques (dans chaque figure, nous voyons la courbe de la fréquence fondamentale ( en Hertz), le texte aligné, la courbe d'intensité (en dB) et le signa acoustique. Première figure : complexe verbal ; deuxième figure : enchâssement obligatoire ; troisième figure : enchâssement supposé. ....	101
Figure 66 : Architecture de base des modèles chomskyens .....	104
Figure 67 : Une manière plus expressive de représenter l'architecture des modèles chomskyens. ....	106
Figure 68 : Diagramme de l'architecture d'un modèle Sens-Texte .....	106
Figure 72: arbres élémentaires <i>A, B, C, D</i> et <i>E</i> .....	118
Figure 73: arbre de dérivation de <i>Jean dort</i> avec la grammaire $\{A, C, D\}$ .....	119
Figure 74: un arbre identique .....	119
Figure 75: arbre de dérivation de <i>Jean dort</i> avec la grammaire $\{B, C\}$ .....	119
Figure 76: .....	123
Figure 77: Cas des dépendances sérielles croisées : l'arbre de $V_1$ doit croiser l'arbre de $V_2$ et de $V_3$ .....	123
Figure 78: structure sémantique de l'exemple (201) .....	125
Figure 79: arbres élémentaires et arbre de dérivation pour la phrase (201). ....	125
Figure 80 : arbres élémentaires pour la phrase (202) .....	126
Figure 81 : arbre dérivé et arbre de dérivation de la phrase (202) .....	126
Figure 82 : arbre dérivé intermédiaire de la phrase (203)avec deux nœuds pied .....	127
Figure 83: un arbre dérivé intermédiaire avec deux nœuds pied .....	128
Figure 84: les arbres de dérivation de la phrase (204) .....	129
Figure 85: structure dérivée intermédiaire .....	130

Figure 86: les arbres de dérivation de de <i>warmes schales Bier</i> et de <i>schales warmes Bier</i> .....	130
Figure 87: les deux façons d'adjoindre .....	131
Figure 88: arbre dérivé.....	132
Figure 89: arbres élémentaire pour l'analyse de (205) .....	133
Figure 90: arbres de dérivation de (205).....	133
Figure 91: arbres dérivés après adjonction de ② dans ③ .....	134
Figure 92: structure argumentale (sémantique) de la phrase (209).....	140
Figure 93: arbre de dérivation de la phrase (209) .....	140
Figure 94: combinaison des arbres élémentaires pour la phrase (209) .....	141
Figure 95: combinaison des arbres élémentaires pour la phrase (209) .....	144
Figure 96: graphe sémantique et arbres syntaxiques de (212) et (213) .....	144
Figure 97: arbre élémentaire de <i>schreibt</i> avec le sujet relativisé et arbre de dérivation de (212) .....	145
Figure 98: graphe sémantique et arbre syntaxique de (214) .....	146
Figure 99: arbre élémentaire possible pour une préposition introduisant un relative .....	147
Figure 100: arbre graphe sémantique et arbre syntaxique de (215).....	148
Figure 101: Modèles satisfaisants D2 – arbres <i>A, B, C, D</i> .....	154
Figure 102: Un des modèles de D3 – arbre <i>D</i> .....	155
Figure 103: Modèles référents de D4 – arbres <i>E, F</i> .....	156
Figure 104 : Arbre <i>G</i> .....	156
Figure 105 : Les cinq modèles référents de D3 – arbres <i>H, I, J, K, L</i> .....	157
Figure 106 : Les modèles référents de D5 .....	158
Figure 107 : Les modèles référents de D3 <sup>c</sup> .....	160
Figure 108 : Les modèles référents de D5 .....	160
Figure 109 : Les modèles référents de D5 <sup>c</sup> .....	161
Figure 110 : Les résultat du générateur de Bertrand Gaiffe pour D5 <sup>c</sup> .....	161
Figure 111 : Le modèle référent de <i>V</i> .....	163
Figure 112 : Le modèle référent de $V \times D \times I$ .....	163
Figure 113 : Les modèles référents de $V \times I$ .....	164
Figure 114 : Le modèle référent.....	165
Figure 115 : le croisement des classes ne donne pas un résultat neutre.....	167
Figure 116 : le croisement des classes donne un résultat neutre.....	167
Figure 117 : après débarras des valeurs doubles, le croisement des classes donne un résultat neutre.....	168
Figure 118 : après débarras des valeurs doubles, le croisement des classes donne un résultat non neutre.....	168
Figure 119 : le croisement des classes F1 à F4 donne un résultat neutre grâce aux multiset....	169
Figure 120 : le croisement des classes donne un résultat neutre grâce aux multiset.....	169
Figure 122 : structure de traits .....	170
Figure 121 : structure hiérarchique .....	170
Figure 123 : Arbre élémentaire du verbe <i>aufschreiben</i> en position <i>V2</i> .....	174
Figure 124 : Arbre élémentaire du verbe <i>aufschreiben</i> en position <i>V2</i> .....	175
Figure 125: Arbre élémentaire auxiliaire .....	178
Figure 126: Expression à l'aide des traits, des opérateurs réguliers .....	178
Figure 127 : Adjonction unique (à gauche du verbe en position <i>V2</i> ) .....	179
Figure 128 : Arbre élémentaire du verbe <i>aufschreiben</i> en position <i>V2</i> .....	180
Figure 129: Arbre élémentaire de relative illustrant la modification adjacente.....	183
Figure 130: Arbre élémentaire d'une relative extraposée illustrant l'extraposition.....	183
Figure 131: Combinaison d'arbres élémentaires dans l'analyse de la phrase (235) .....	184
Figure 132: Arbre de dérivation qu'on obtient dans le cas de la modification relative adjacente .....	185
Figure 133: Arbre de dérivation erroné correspondant au cas de l'extraposition de la relative .	186

Figure 134: Combinaison d'arbres élémentaires dans l'analyse de la phrase (236) et passage des traits .....	187
Figure 135: Les deux possibilités d'adjonction d'un verbe matriciel dans la parenthèse droite	190
Figure 136: Constructions d'un verbe matriciel dans la parenthèse droite (phrases (245)et (246)) .....	191
Figure 137: Construction d'un SN avec proposition relative et verbe à montée.....	192
Figure 138: Construction d'un SN avec proposition relative et verbe à contrôle.....	192
Figure 139: Substitution du verbe enchâssé dans le cas de la relativisation d'un argument nominal d'un verbe à montée (arbre de gauche) et d'un verbe à contrôle (arbre de droite), nécessaire pour l'analyses des phrases (249) et (250) .....	193
Figure 140 : Représentation de la description de la classe « verbal-anchor » .....	195
Figure 141 : Classes principales de la hiérarchie de la réalisation verbale .....	195
Figure 142 : Détail de la capture d'écran de l'éditeur de métagrammaire de Gaiffe 2002. ....	196
Figure 143 : Hiérarchie de réalisation verbale pour les verbes non finis utilisés dans des constructions auxiliaire et à montée.....	196
Figure 144 : Hiérarchie de réalisation verbale autour des constructions pour les propositions relatives. La classe <i>VRelTwoVerbs</i> nous servira comme exemple dans la suite.....	196
Figure 145 : Description de la classe <i>VRelTwoVerbs</i> pour la construction verbale du verbe <i>schicken</i> 'offrir' dans le contexte de la proposition relative dans (251) .....	197
Figure 146: Neutralisation des traits polaires pour obtenir la classe croisée $accArgFull \times N1N3N4V \times diaVirtualAcc \times diaDativePassiveVerb \times relSetupArgN \times StandardLocal$ .....	198
Figure 147 : Début de l'entrée de la classe correspondant à la famille des verbes ditransitifs <i>N1N3N4V</i> . Cette classe a besoin d'être croisée avec des classes proposant des diathèses pour chacun des arguments nominaux ainsi que pour le verbe. ....	199
Figure 148 : Familles avec argument verbal et les familles sans argument au nominatif. ....	199
Figure 149 : Les familles sans argument verbal et avec au moins un argument au nominatif ( <i>NI</i> ). .....	199
Figure 150 : La diathèse de la métagrammaire .....	201
Figure 151 : Classe de la diathèse verbale du passif de l'accusatif <i>diaPassiveVerb</i> .....	201
Figure 152 : Classe de la diathèse du nominatif au passif <i>diaPassiveNom</i> . ....	202
Figure 153 : Classe de la diathèse du passif du datif <i>diaDativePassiveVerb</i> . ....	202
Figure 154 : Classe de la diathèse virtuelle (à l'actif) de l'accusatif qui sert aussi pour l'accusatif dans le contexte du passif du datif. ....	203
Figure 155 : Classe de la réalisation de la structure verbale du passive du datif.....	203
Figure 156 : Réalisations verbales et de différents arguments nominaux .....	204
Figure 157 : Réalisation du <i>Vorfeld</i> .....	205
Figure 158 : Hiérarchie de l'accord.....	205
Figure 159 : Représentation simplifiée de la classe $accArgFull \times N1N3N4V \times diaVirtualAcc \times diaDativePassiveVerb \times relSetupArgN \times StandardLocal Agreement \times 2VerbsRight \times vonArg \times realDatPassiveVerb \times nomRel \times VRelTwoVerbs$ .....	206
Figure 160: Premier modèle minimal de la classe croisée.....	207
Figure 161: Deuxième modèle minimal de la classe croisée.....	207
Figure 162: Analyse complète de la phrase sous (252) : .....	208
Figure 163: Dérivation de la phrase (253). ....	215
Figure 164: Représentations schématiques correspondent à des phrases grammaticales.....	224
Figure 165: partie de gauche : représentation schématique d'une phrase « syntaxiquement correcte » et « sémantiquement déficiente ».....	225
Figure 166: Représentation schématique correspondent à une phrase « syntaxiquement déficiente » : problème d'accord. ....	226
Figure 167 : Représentation schématique correspondent à une phrase « syntaxiquement déficiente » : problème de valence. ....	227

Figure 168: Représentation schématique correspondant à une phrase « topologiquement défectueuse ».....	228
Figure 169: Exemples d'arbres élémentaires TUG.....	229
Figure 170: Exemples de traits fonctionnels.....	230
Figure 171: Les entrées lexicales pour un TUG jouet de l'allemand.....	233
Figure 172: Correspondance arbre de dépendance syntaxique et structure topologique pour la phrase (261) .....	234
Figure 173: Analyse de l'exemple (261) avec une TUG. ....	235
Figure 174: structure f de la phrase (262). ....	237
Figure 175: structure f de la phrase (262) revisitée. ....	238
Figure 176: .....	239
Figure 177: Règles de la structure des constituants pour l'allemand.....	241
Figure 178: structure f de la phrase (263). ....	243
Figure 179 : ① est la représentation graphique de la structure prédicat-arguments, ② représente la structure de dépendance syntaxique et ③ le graphe sémantique des exemples sous (264). .....	244
Figure 180: Première structure c possible pour la phrase (263). ....	244
Figure 180: Première structure c possible pour la phrase (296). ....	245
Figure 182: Deuxième structure c possible pour la phrase (296). ....	245
Figure 183: Schématisations des courbes mélodiques des phrases (265) et (266) .....	247
Figure 184: Schématisations des courbes mélodiques de la phrase déclarative allemande. Le rectangle indique le champ Mittelfeld. Notons que dans toutes les phrases, les topiques sont optionnels. ....	250
Figure 185 : Arbre de dépendance syntaxique de surface augmentée de traits communicatifs..	255
Figure 186: La tête syntaxique <i>hat</i> 'a' ouvre un domaine principal qui comporte ces champs et se place dans la parenthèse gauche.....	255
Figure 187: Placement des dépendants de <i>hat</i> 'a' : le dépendant verbal se place dans la parenthèse droite et son dépendant nominal dans FocF.....	255
Figure 188: Les deux possibilités de placement des dépendants de <i>versprochen</i> 'promis' .....	256
Figure 189: Domaine du syntagme nominal. ....	257
Figure 190 : Tableau de Netter 1994. ....	258
Figure 191: Linéarisation possible de l'arbre de dépendance syntaxique de surface de la Figure 192.....	258
Figure 192 : Arbre de dépendance syntaxique de surface de l'exemple donné par Netter.....	259
Figure 193 : Arbre de dépendance syntaxique de surface d'une complétive.....	259
Figure 194: Linéarisation possible de l'arbre de dépendance syntaxique de surface de la Figure 193.....	259
Figure 195 : Arbre de dépendance syntaxique de surface d'un exemple grec.....	264
Figure 196 : Les trois structures topologiques possibles pour l'arbre de dépendance de la Figure 195.....	265
Figure 197 : Arbre de dépendance syntaxique de surface d'un exemple grec avec un marquage communicatif différent que celui de la Figure 195. ....	266
Figure 198 : Structure topologique pour l'arbre de dépendance de la Figure 197.....	266
Figure 199 : Arbre de dépendance d'une phrase complexe comprenant trois niveaux d'enchâssement. ....	267
Figure 200 : Structure topologique de la première linéarisation possible (282)a. ....	268
Figure 201 : Structure topologique de la deuxième linéarisation possible (282)b.....	268
Figure 202 : Structure topologique de la troisième linéarisation possible (282)c.. ....	269
Figure 203 : Structure topologique de la quatrième linéarisation possible (282) d .....	270

# Table des matières

<b>Introduction</b>	<b>1</b>
Du Sens à la forme phonétique de la phrase	2
La liberté d'ordre en allemand	4
Plan de la thèse	6
<b>I. L'allemand</b>	<b>7</b>
<b>1 Les cas de l'allemand</b>	<b>8</b>
1.1 Définition et conventions	8
1.2 La raison d'être des cas	9
<b>2 Actant sémantique, actant syntaxique et argument syntaxique</b>	<b>10</b>
2.1 Actant et argument	10
2.2 Introduction d'un niveau intermédiaire : la syntaxe profonde	16
<b>3 Le passif</b>	<b>19</b>
3.1 Construction du passif	19
3.2 Passif de processus et passif d'état	21
<b>4 Prépositions et arguments syntaxiques</b>	<b>24</b>
<b>5 Le datif libre</b>	<b>27</b>
5.1 Les phénomènes du passif du datif (construction avec bekommen)	27
5.2 Application du critère de diathèse	29
5.3 Cas structurel et cas lexical	31
<b>6 La notion d'ordre libre des mots</b>	<b>32</b>
6.1 Un ordre libre des mots	32
6.2 Et l'allemand ?	35
6.3 Les structures communicatives	37
6.4 Ordre des mots standard	41
6.5 L'ordre standard : standard par rapport à quoi ?	42
6.6 Les différents phénomènes d'ordre des mots	44
<b>II. Topologie</b>	<b>51</b>
<b>1 Les règles de base de la topologie</b>	<b>53</b>
<b>2 Différentes structures verbales de la topologie</b>	<b>57</b>
2.1 Les structures plates	58
2.2 Analyse du « brouillage d'arguments »	59
2.3 L'enchâssement	60
2.4 L'émancipation	63
2.5 Le traitement des préfixes	64
<b>3 Structure du complexe verbal</b>	<b>65</b>
3.1 Description des possibilités d'ordre dans un complexe verbal	65
3.2 Modélisation du complexe verbal	67
3.3 Complexité du calcul topologique et la « troisième construction »	69
<b>4 Complétives et Relatives</b>	<b>71</b>
4.1 Les complétives	71
4.1.1 Le traitement des complétives	71
4.1.2 Les complétives sans complémenteurs	72
4.1.3 Parenthétique et complémentation	74
4.2 Les relatives et pronoms résomptifs	76
4.2.1 Les phrases relatives dans une grammaire de dépendance	76
4.2.2 Relativisation des modifieurs	82
4.2.3 Topologie et relatives	85
4.2.4 Une place devant le Vorfeld ?	91



<b>5</b>	<b>Sur les traces du modèle topologique</b>	<b>93</b>
<b>6</b>	<b>La notion de syntagme topologique</b>	<b>96</b>
6.1	Critères syntaxiques	96
6.2	Critères communicatifs et prosodiques	98
<b>7</b>	<b>La place de la structure syntagmatique dans un modèle linguistique</b>	<b>102</b>
<b>III. TAG</b>		<b>109</b>
<b>1</b>	<b>Les grammaires d'arbres adjoints lexicalisées</b>	<b>111</b>
1.1	La lexicalisation	111
1.1.1	La délexicalisation partielle	114
1.1.2	La notion de famille	115
1.2	L'idée de métagrammaire	115
1.3	Définition de TAG	117
1.3.1	Substitution et adjonction	118
<b>2</b>	<b>Critiques et extensions</b>	<b>120</b>
2.1	Capacité générative	121
2.2	L'arbre de dérivation	124
2.2.1	Le problème des verbes à contrôle	124
2.2.2	Grammaire d'arbres adjoints – mais adjoints comment ?	128
2.3	Interprétation de la structure dérivée	135
2.3.1	La structure dérivée de FTAG	136
2.3.2	La structure profonde sous-jacente	136
2.4	TAG et l'allemand	138
2.4.1	Brouillage d'arguments d'un verbe	140
2.4.2	Débrouillages des arguments	142
2.4.3	Brouillage de modificateurs	143
2.4.4	Relativement difficile	144
2.4.5	Modificateurs relatifs en TAG	145
2.4.6	La position de la relative dans la phrase	148
2.4.7	XTAG – FTAG – DTAG ?	149
<b>3</b>	<b>Les métagrammaires pour les TAG</b>	<b>150</b>
3.1	Les méta-règles	150
3.1.1	Organisation hiérarchique de l'information contenu dans les arbres élémentaires	151
3.1.2	La métagrammaire de FTAG	151
3.1.3	Vers une métagrammaire de l'allemand	152
3.2	Les descriptions et leurs arbres	153
3.2.1	Notions de base et la fonction d'assignation de catégories	153
3.2.2	Les traits	159
3.2.3	L'implémentation du trait cat	161
3.3	Le modèle minimal dans la métagrammaire	162
3.4	Les traits polaires	165
<b>4</b>	<b>DTAG</b>	<b>170</b>
4.1	Objectifs de DTAG	170
4.2	Les critères retenus pour DTAG	171
4.3	Des solutions compromettantes : les premiers arbres	173
4.3.1	Arbres élémentaires	173
4.3.2	V2 en TAG	176
4.3.3	Les relatives en TAG	181
4.3.4	Les compléments verbaux en TAG	186
4.4	L'organisation de la DTAG dans une métagrammaire	194
4.4.1	La construction d'un schème verbale	194
4.4.2	Un exemple de neutralisation de traits polaires	197
4.4.3	Familles	198
4.4.4	Diathèses	201
4.4.5	Réalisations	203
<b>5</b>	<b>Remarques conclusives sur DTAG</b>	<b>208</b>

<b>IV. Formalisations et extensions de la topologie</b>	<b>211</b>
<b>1 Formalisation d'une analyse topologique</b>	<b>212</b>
1.1 Formalisation de base	212
1.2 Augmentation des règles des traits communicatifs	215
1.3 L'implémentation DepLin	216
<b>2 Le problème du sujet dans un domaine enchâssé</b>	<b>219</b>
2.1 Les données sur le VP fronting	220
2.2 Interprétation prosodique et communicative des données	221
2.3 Le rôle de la structure communicative dans le groupement des mots	222
2.4 Que reste-t-il pour le module syntaxique ?	223
2.5 Exemples de correspondances	223
<b>3 Grammaires d'unification d'arbres (TUG)</b>	<b>228</b>
3.1 Définition des TUG	229
3.2 Où se trouvent les conditions de bonne formation ?	231
3.3 Une grammaire jouet pour l'allemand	232
3.3.1 Description de la grammaire	232
3.3.2 L'analyse avec TUG	234
<b>4 LFG</b>	<b>236</b>
4.1 Quelques ajustements de notation	236
4.2 Formalisation	239
<b>5 L'ordre dans le Mittelfeld – un phénomène topologique ?</b>	<b>246</b>
5.1 L'organisation communicative et prosodique du Mittelfeld	246
5.2 Analyse du Mittelfeld	251
5.3 Réorganisation du champ Mittelfeld	252
5.4 Illustration des nouveaux champs introduits	254
<b>6 La topologie du syntagme nominal allemand</b>	<b>256</b>
<b>7 Topologie comparative : coréen, grec, français</b>	<b>261</b>
7.1 La topologie du grec moderne	261
7.1.1 Les règles topologique du grec	262
7.1.2 Illustration	263
<b>Conclusion</b>	<b>271</b>
<b>Références bibliographiques</b>	<b>275</b>
<b>Table des figures</b>	<b>285</b>
<b>Table des matières</b>	<b>290</b>