

**Anne-Marie di Sciullo**

## **Théorie et description en grammaire générative**

### **Remerciements**

Ma reconnaissance va à Yves-Charles Morin qui a eu la gentillesse de se livrer à l'examen critique d'une version préliminaire de cet ouvrage. Plusieurs de ses commentaires ont été intégrés à la version définitive. Ma reconnaissance va également à Sylvie Robidoux pour ses patientes et minutieuses corrections d'épreuves du texte définitif.

Je dois beaucoup à Noam Chomsky, Jim Higginbotham, Edwin Williams et Judith Mc A'Nulty qui m'ont fourni une aide précieuse sous des formes diverses.

Je suis également redevable à mes étudiants de l'Université du Québec à Montréal, tout particulièrement à José Bonneau et à Pierre Lavoie dont les réactions à mes cours de syntaxe et de sémantique m'ont été très précieuses.

Enfin, je remercie l'Office de la langue française qui a subventionné en partie la recherche préliminaire à la rédaction de l'ouvrage, ainsi que l'Université du Québec à Montréal qui a assuré une partie des coûts d'édition.

### **Introduction**

Ce livre présente les aspects principaux de la grammaire générative. Il peut être utilisé comme texte de référence aux cours de linguistique, telle qu'elle se définit dans les années 80 dans le cadre de la théorie standard étendue ((TSÉ) Chomsky 1981, 1982).

Il s'est élaboré dans la perspective de systématiser les principes de base, les hypothèses et les théories qui ont été proposés dans la TSÉ. Il ne se veut pas un compte rendu exhaustif des problèmes et des résultats obtenus dans le domaine, il trace plutôt des points de repère à l'intérieur des développements récents de la grammaire générative et cherche à mettre en évidence les points saillants de cette théorie linguistique. Essentiellement, la TSÉ vise à formuler un modèle des langues naturelles et de leur acquisition qui puisse décrire et expliquer plusieurs phénomènes linguistiques observables. Alors que les grammaires particulières décrivent les

propriétés de langues spécifiques, la théorie des grammaires, valable pour toutes les langues, présente les règles et les principes qui rendent possible l'acquisition du langage. Dans ce livre, nous présentons les règles et les principes grammaticaux surtout à partir de faits du français, bien que dans certains cas nous ferons référence à d'autres langues : l'italien, l'anglais, le walpiri, le gallois, l'allemand, le latin, le japonais, le hindi, le navajo, le guarani.

Le premier chapitre situe brièvement la grammaire générative dans l'histoire des sciences du langage et fait ressortir quelques différences entre le structuralisme et la grammaire générative. En particulier le fait que les hypothèses formulées dans le cadre de la TSÉ ont renouvelé la problématique de la linguistique et ont modifié son objet d'étude. Alors que ce sont les propriétés des langues naturelles qui constituent l'unique objet des descriptions structuralistes, ce sont les propriétés des grammaires, et seulement de manière dérivée celles des langues naturelles qui sont au centre des recherches dans le cadre de la TSÉ. De plus, les hypothèses sur l'organisation modulaire des grammaires et la dépendance de la structure des règles grammaticales ont renouvelé la conception des grammaires et des règles. En outre, les hypothèses sur la faculté du langage et la théorie des grammaires ont permis de formuler une théorie de l'acquisition du langage qui explique comment les langues sont acquises si rapidement et à partir de très peu de données. Ces hypothèses ont permis la découverte de faits nouveaux. En particulier, l'existence de catégories abstraites (sans réalisation phonétique) constitue l'une des découvertes les plus intéressantes à l'intérieur de la TSÉ. De tels éléments abstraits sont inaccessibles à la linguistique structurale, limitée à l'empirisme et à la taxinomie.

Les chapitres 2 à 5 développent les aspects fondamentaux des systèmes de règles et de principes qui interagissent dans la description et l'explication des phénomènes linguistiques.

Le chapitre 2 contient plusieurs règles lexicales de la grammaire du français, une analyse de la sous-catégorisation lexicale et une première formulation du principe de projection, un des principes de la théorie des grammaires, qui assure que l'information lexicale figure à tous les niveaux de représentation syntaxique.

Au chapitre 3, les propriétés des règles de base sont présentées, ainsi que les principes de la théorie X-barre qui limite leur forme. Une grammaire partielle du français est formulée d'après ces principes. En outre, la théorie des Cas et de la théorie de la rection gouvernent la bonne formation des structures syntaxiques engendrées et permettent d'éliminer plusieurs règles de base des grammaires particulières, étant donné le principe de projection et la théorie X-barre.

Les transformations sont présentées au chapitre 4. Des arguments empiriques et théoriques viennent justifier ce type de règle. Dans la TSÉ, il n'y a qu'une seule transformation générale : déplacer alpha, où alpha est une catégorie. Ceci permet de simplifier le composant transformationnel des grammaires, qui dans des modèles antérieurs (p. ex. la théorie standard (TS) Chomsky 1965) se composait d'un grand nombre de transformations effectuant différentes opérations à la fois. Cette simplification rend la TSÉ plus adéquate au niveau explicatif, étant donné le principe méthodologique propre à ce modèle de réduire la classe des grammaires possibles et accessibles à l'enfant. La rapidité d'acquisition du langage se trouve expliquée par le fait que cette classe est très limitée et que les règles grammaticales sont très simples. Deux transformations de la grammaire du français sont présentées : Mouvement qu et Mouvement N, des cas particuliers de la règle générale déplacer alpha. Les structures engendrées par cette règle sont filtrées par des conditions de bonne formation, notamment la théorie des limites et la théorie du liage. Ces dernières simplifient et généralisent plusieurs contraintes proposées par Ross (1967) et par Chomsky (1973, 1980) pour bloquer l'application des transformations ou pour assurer la bonne formation des relations anaphoriques. Il sera également question des principes de la théorie du contrôle, qui gouvernent la référence de l'élément pronominal abstrait PRO.

Le dernier chapitre porte sur les règles du composant de forme logique (FL) qui dérivent la FL : le seul niveau accessible à l'interprétation sémantique dans la TSÉ. Les concepts de prédicat, d'argument, d'opérateur et de variable sont définis en termes configurationnels, de sorte que plusieurs aspects de l'interprétation sémantique des phrases peuvent s'établir à partir de la FL. Des conditions de bonne formation, dont le critère thématique, viennent filtrer les FL, acceptant celles qui sont bien formées et rejetant les structures agrammaticales.

L'exposé met en relief le développement rapide de la théorie linguistique. Il sera ainsi possible de voir la grammaire générative comme un espace en devenir, où l'organisation générale des grammaires, les règles et les principes qui les composent se modifient face à des impératifs d'adéquation explicative.

## **Chapitre 1 : L'investigation linguistique**

### **1.1 Préliminaires**

L'investigation linguistique, comme toute autre investigation scientifique, vise l'explication des phénomènes naturels. À la base, on y trouve une curiosité face à des faits qui semblent aller de soi. L'histoire des sciences montre comment le développement de la

connaissance a été alimenté par des questions apparemment triviales. Pourquoi tout corps solide est-il attiré vers le sol sous l'influence de son poids ? Galilée, Kepler, Newton, entre autres, ont présenté des réponses à cette question, et des lois générales expliquent le phénomène. Chacune de ces théories s'appuie sur la critique des théories précédentes, et dans la plupart des cas présentent une problématique nouvelle. Les révolutions scientifiques opèrent, selon l'expression de Kuhn, un changement de paradigme, c'est-à-dire théorie dominante. Elles modifient radicalement la définition du domaine d'investigation, des techniques d'observation et des lois qui caractérisent les phénomènes.

Face aux questions suivantes : qu'est-ce que le langage ? Quelles sont les propriétés des langues naturelles ? Sont-elles toutes différentes ou ont-elles des points communs ? Et comment sont-elles acquises ? Plusieurs hypothèses ont été formulées, et certaines se sont constituées en paradigme.

La grammaire générative constitue une révolution en matière de science du langage et modifie profondément la conception des langues et des grammaires qui s'est développée dans la première partie du vingtième siècle.

Il est intéressant de considérer la révolution scientifique effectuée par la grammaire générative d'un point de vue épistémologique. À ce chapitre, les théories de Popper, de Kuhn et de Lakatos ont leur intérêt : elles ont influencé la vision contemporaine du développement des sciences et de l'émergence des révolutions scientifiques.

Selon Popper et Kuhn, l'histoire des sciences progresse de façon linéaire : les théories scientifiques sont suivies de réfutations qui les éliminent et qui donnent lieu à de nouvelles théories. À l'intérieur de ce continuum, Popper met de l'avant l'idée que la science se développe par la discussion de points de vue différents et que les révolutions scientifiques s'opèrent par la critique. Une théorie nouvelle s'impose si elle peut expliquer un plus grand nombre de phénomènes que les théories précédentes.

Alors que Popper souligne l'aspect révolutionnaire-critique de l'activité scientifique, Kuhn met en relief la fonction de ténacité liée aux théories scientifiques dominantes. L'activité scientifique normale, ou « science normale », est essentiellement conformiste, elle se développe en accord avec les principes généraux des théories dominantes acceptés par les communautés scientifiques. Les révolutions scientifiques sont par contre des états de crise qui viennent ébranler

les paradigmes, des moments extraordinaires qui se situent en dehors du domaine de la science normale.

Lakatos souligne que les théories scientifiques ne sont pas des hypothèses isolées, mais plutôt des programmes de recherche complexes, composés de plusieurs théories en développement. Son point de vue éclaire particulièrement la nature des théories et l'émergence des révolutions scientifiques. Selon Lakatos les théories se composent d'une partie centrale, d'une périphérie et d'un ensemble de procédures de découvertes. Ces dernières incluent des techniques pour effectuer les analyses et solutionner des problèmes. La périphérie se compose de théories auxiliaires qui permettent de traiter les anomalies ou les contre-exemples sans que le centre ne soit affecté.

Ainsi définies, les théories ne peuvent être falsifiées par des contre-exemples, mais uniquement par d'autres théories, et il n'y a pas de falsification sans l'émergence d'une meilleure théorie. Pour Lakatos, une théorie est révolutionnaire si elle permet la découverte de phénomènes nouveaux, et à partir d'eux, de généralisations importantes, masquées par les paradigmes précédents. Lorsque de nouvelles théories émergent, c'est que des programmes de recherche en progrès remplacent des programmes de recherche en déclin. Les premiers se caractérisent par le fait qu'ils peuvent prédire plusieurs phénomènes dont certains sont corroborés par des faits nouveaux, jamais observés auparavant, les seconds n'ont pas cette propriété.

L'aspect révolutionnaire-critique de l'activité scientifique, la crise de la science normale, la découverte de phénomènes nouveaux éclairent l'évolution des théories scientifiques et les changements de paradigmes. Ils permettent d'interpréter le passage de la linguistique structurale à la grammaire générative.

## **1.2 Le changement de paradigme**

Le paradigme structuraliste se développe en linguistique, en Europe et en Amérique dans la première moitié du vingtième siècle et s'étend aux autres domaines de la connaissance, tels que la sociologie et l'anthropologie.

La linguistique structurale s'oppose à la tradition philologique du dix-neuvième siècle, et présente des aspects novateurs dans le domaine des sciences du langage. Les structuralistes veulent faire de la linguistique une discipline scientifique en appliquant rigoureusement le principe de l'observation à l'analyse des langues. L'objet d'étude de la linguistique se limite aux faits qui sont directement observables à l'intérieur d'un corpus, c'est-à-dire d'un ensemble fini

d'énoncés effectivement produits, et la théorie linguistique met de l'avant l'idée qu'une langue est un système d'unités (p. ex. phonèmes, morphèmes) identifiables par des procédures basées sur des opérations de découpage et de classement (1).

L'analyse linguistique conduit à les identifier et à les classer à l'intérieur d'un système hiérarchique. Toutefois, aucune règle ne vient prédire leur occurrence ni leur relation à l'intérieur du système. En outre, aucun élément implicite, absent en surface, n'est jamais admis dans l'analyse. Ainsi par exemple, le sujet de l'infinitive enchâssée de la phrase [1] est absent en surface, bien qu'il réfère à l'objet de la phrase matrice, soit Marie. Ce phénomène, comme tout ce qui ne relève pas de l'observable est ignoré.

(1) Julie permet à Marie \_\_\_ de partir

À cette conception empirique et taxinomique de la linguistique, s'ajoute une conception behavioriste du comportement verbal et de l'acquisition du langage.

Pour Skinner, Tadwell et les autres behavioristes, le langage est un ensemble d'habitudes, les échanges linguistiques sont des réponses à des stimuli du monde extérieur, et n'impliquent pas en premier lieu l'intervention de facultés cognitives. Les structuralistes adoptent cette conception du langage et supposent que la production des phrases relève de l'analogie et du conditionnement. Bloomfield (1933) soutient que « c'est la dernière phrase qui conditionne la suivante », et Hockett (1958) affirme que « lorsqu'une personne parle, c'est soit de la répétition, soit de l'analogie ».

---

<sup>1</sup> L'opération de commutation permet de découvrir les unités linguistiques. Elle consiste à substituer une unité à une autre dans un même contexte, et à voir si le résultat produit une différence significative dans le système de la langue. Si c'est le cas cette unité appartient au système. Ainsi, à partir des exemples [1] a.b. il est possible de constater que /t/ et /d/ sont des phonèmes du français, et que dé et té sont des morphèmes. L'assignation des morphèmes à des catégories grammaticales est déterminée par la distribution des morphèmes, c'est-à-dire par leur occurrence dans des contextes typiques. En (1)c., les morphèmes grammaire et règle figurent dans les mêmes contextes, et sont assignés à la même classe distributionnelle, celle des N.

(1)

a. /to/ : tôt ; /do/ : dos  
/te/ : thé ; /de/ : dé

b. re-faire / dé-faire ; re-mettre / dé-mettre

c. la\_\_ : la\_\_formelle ; écrivez une\_\_

En outre, la structure des phrases est conçue comme une suite finie de positions, chacune pouvant être occupée par une classe de morphèmes. La relation de co-occurrence entre les unités est un aspect du système linguistique désigné par l'expression « relation syntagmatique ». En outre, pour chaque position, il y a un choix possible à l'intérieur de classes d'unités en « relation paradigmatic ». Ces deux relations syntagmatique paradigmatic caractérisent l'organisation générale des systèmes linguistiques dans la linguistique structurale.

Chomsky critique la linguistique structurale et montre que ce paradigme ne tient pas compte, entre autres, de la capacité des locuteurs de comprendre et de produire une infinité de phrases nouvelles. En outre, il souligne que le corpus n'est qu'une manifestation de la connaissance implicite que tout locuteur possède de sa langue, et c'est plutôt cette connaissance [2] qui est à investiguer lorsqu'il s'agit de caractériser les propriétés fondamentales du langage. Cette connaissance met en jeu différents types d'habilités qui nous font notamment différencier les phrases grammaticales des phrases agrammaticales, précédées d'un \* dans les exemples comme (2), ou encore les phrases interprétables de celles qui ne le sont pas, notées ! comme en (3), et juger la prononciation des phrases, comme en (4) où les syllabes soulignées sont accentuées.

- (2) a. Les linguistes aiment la musique.  
 b. \*Les la musique linguistes aiment.
- (3) c. Voici un homme intelligent.  
 d. !Voici un homme rance.
- (4) e. Pierre aime Paule.  
 f. \*Pierre aime Paule.

En outre, les langues ont des propriétés structurelles qui sont facilement accessibles, et qui nous permettent d'affirmer, entre autres, que les deux premiers termes de la phrase (5a) forment une unité syntaxique (un constituant) et non les deuxième et troisième termes, ou encore de constater que les phrases (6) ont des interprétations sémantiques distinctes selon la position du pronom il, et que les suites (7) sont interprétées comme un seul mot ou trois mots différents selon la position de l'accent (').

- (5) a. [Les linguistes] inventent des théories.  
 b. \*Les [linguistes inventent] des théories.
- (6) a. Il pense que Pierre est fou.  
 b. Pierre pense qu'il est fou.
- (7) a. pied de poule.  
 b. pièd dè poule.

La linguistique structurale ne considère pas le langage dans cette perspective et ne peut conséquemment offrir d'explication à ces phénomènes. Ses limites ont conduit à la formulation

---

<sup>2</sup> Le terme compétence linguistique est souvent utilisé pour désigner la connaissance implicite que les locuteurs ont de leur langue. La compétence se distingue de la performance qui est l'emploi effectif de la langue dans des situations concrètes.

d'un nouveau paradigme, la grammaire générative, qui s'est développée surtout grâce aux travaux de Noam Chomsky et qui s'est imposée en Amérique et en Europe dans la seconde partie du vingtième siècle. Chomsky renouvelle la problématique générale de la linguistique. Il propose un programme de recherche qui vise l'élaboration d'une théorie capable de caractériser les propriétés des langues et qui soit compatible avec une théorie explicative de l'acquisition du langage. Cette caractérisation se fait par l'interaction de règles et de principes. Alors que les grammaires particulières définissent les règles spécifiques aux langues, les principes de la théorie des grammaires viennent limiter la classe des grammaires propres aux langues naturelles. Cette interaction permet de relier la théorie linguistique à la théorie de l'acquisition, puisqu'il est naturel de penser que la classe des grammaires accessibles à l'enfant est très petite, étant donné la rapidité avec laquelle il acquiert le langage, et ce à partir de très peu de données.

La révolution chomskyenne modifie l'objet d'étude de la linguistique. Alors que ce sont les langues naturelles qui constituent l'unique objet de la linguistique structurale, ce sont les propriétés des grammaires, et seulement de manière dérivée celle des langues naturelles, qui sont au centre des recherches dans le cadre de la grammaire générative.

Lorsque l'objet d'étude est confiné aux propriétés observables des langues, l'investigation linguistique ne peut dépasser l'empirisme, et les analyses ne traitent que les propriétés de surface des langues. De plus, il est difficile de voir, à l'intérieur de la linguistique structurale, ce qui fait que les langues, bien que différentes en surface, ont des propriétés communes. En outre, il est difficile d'expliquer comment elles peuvent être apprises rapidement à partir d'un ensemble réduit d'énoncés.

Quand l'objet d'étude devient plus abstrait, et passe des langues aux grammaires, il est possible de dépasser l'empirisme et d'accéder à la théorie en matière de science du langage.

Ces modifications de l'objet d'étude et de la problématique générale peuvent être vues, à la lumière des idées de Lakatos, comme la manifestation d'un changement qualitatif dans l'histoire des sciences du langage par lequel l'investigation linguistique accède à un stade théorique.

Le paradigme générativiste est un paradigme en progrès, qui a un rôle révolutionnaire dans l'histoire des sciences du langage, et qui a favorisé le développement d'hypothèses prometteuses sur la théorie des grammaires et sur l'acquisition du langage.



## 1.3 Des hypothèses

### 1.3.1 La faculté du langage

La conception mentaliste du langage, qui émerge à travers l'histoire dans les travaux de grammairiens et de philosophes tels que Descartes, Du Marsais, Humbolt [<sup>3</sup>], est reprise par Chomsky. Il développe l'idée que le langage est un processus cognitif, rendu possible par la faculté du langage, qu'il définit comme une représentation abstraite des capacités innées, spécifiques à l'espèce humaine, qui rendent possible l'acquisition.

Selon cette hypothèse, certains mécanismes grammaticaux communs à toutes les langues sont présents à la naissance dans la structure même du cerveau. C'est précisément parce que les humains sont biologiquement prédisposés à acquérir les grammaires des langues naturelles que cet apprentissage se fait en très peu de temps (de la naissance à 3 ans) et à partir de très peu de données. La faculté du langage offre ainsi une explication à la rapidité de l'apprentissage linguistique, ce qui n'est pas le cas de l'hypothèse behavioriste selon laquelle l'acquisition se fait sur la base de la répétition et de l'analogie.

Chomsky suppose que la faculté du langage contient une procédure d'évaluation des grammaires qui joue un rôle important dans l'acquisition.

Pour acquérir une langue, un enfant doit élaborer des hypothèses compatibles avec les données auxquelles il est exposé. Il doit choisir dans l'ensemble des grammaires possibles et compatibles avec les données de l'expérience, une grammaire spécifique.

La théorie linguistique qui veut rendre compte de l'acquisition contient une procédure d'évaluation qui limite le choix des grammaires. Soit les trois critères d'évaluation suivants. Une grammaire atteint l'adéquation observationnelle si elle prédit correctement quelles sont les phrases grammaticales et celles qui ne le sont pas. Elle est adéquate descriptivement si en outre, elle décrit correctement les structures des phrases et rend compte par là de l'intuition des locuteurs. Enfin, elle atteint l'adéquation explicative si elle décrit la structure des phrases par un ensemble de règles très contraintes et très simples [<sup>4</sup>]. Ce dernier critère est crucial lors de l'acquisition, puisqu'il est naturel de penser que des règles très contraintes et très simples peuvent être acquises par l'enfant très rapidement, et à partir du peu d'information dont il dispose.

---

<sup>3</sup> Voir Chomsky (1966)

<sup>4</sup> La simplicité des grammaires est calculée en termes du nombre de symboles et du nombre de règles qui les constituent. D'après cette métrique d'évaluation entre deux grammaires équivalentes du point de vue descriptif, c'est celle qui contient le moins de symboles ou de règles qui sera choisie.

C'est dans cette perspective que plusieurs contraintes ont été proposées pour limiter les classes des règles possibles (p. ex. la théorie X-barre, la métarègle déplacer alpha), ainsi que la classe des structures engendrées par ces règles (p. ex. le principe de sous-jacence, la théorie des rôles thêta, la théorie du liage). Ces principes appartiennent à la théorie des grammaires, et ils ont une fonction explicative. Ils expliquent pourquoi certaines structures linguistiques sont impossibles et que certaines règles ne peuvent appartenir aux grammaires des langues naturelles. En plus d'avoir une fonction filtrante, ils limitent la classe des règles, et à fortiori la classe des grammaires compatibles avec les données de l'expérience.

Les principes et les règles de la théorie des grammaires, valables pour toutes les langues, sont associés à des paramètres ouverts qui déterminent la variation possible entre les langues.

Un paramètre est un aspect de la grammaire qui varie d'une langue à l'autre, tel l'ordre des mots, la configurationnalité, la présence obligatoire ou facultative d'un sujet lexical.

On sait que les langues diffèrent du point de vue de l'ordre des mots dans les phrases affirmatives simples. En japonais par exemple, le verbe est en position finale de phrase, en français, il suit le sujet et précède l'objet, et en gallois il est en position initiale de phrase. L'ordre des mots est un paramètre de la théorie des grammaires, tout comme le paramètre de la configurationnalité (cf. Hale (1979), Higginbotham (1984)) qui permet de distinguer les langues où l'ordre de mots est libre, telle le walpiri (une langue australienne), des langues où il est fixe, telles le français ou l'italien. En outre, le paramètre Pro-drop (cf. Rizzi (1982)) permet de distinguer le français et l'italien du point de vue de la présence obligatoire ou facultative du sujet. Cette différence peut être observée dans les exemples suivants ;

- (8) a. E venuto.  
(\*Est venu.)
- b. Nevica.  
(\*Neige.)
- c. Sembrava intelligente.  
(\*Semblait intelligent.)

Une des tâches de la linguistique est de définir l'ensemble des paramètres de variation entre les langues. Dans cette perspective, on suppose que la variation linguistique n'est pas due au hasard, mais que les langues varient à l'intérieur de limites strictes. En outre, dans « Principles and Parameters », Chomsky (1979b) fait l'hypothèse que l'acquisition du langage est un processus qui consiste à fixer les paramètres de la grammaire universelle. À la partie centrale de la théorie est associée une théorie auxiliaire, la théorie de la marque, qui préserve le centre de

contre-exemples éventuels. La théorie de la marque permet de distinguer les aspects marqués des aspects non-marqués des langues naturelles.

Chaque grammaire particulière est formée d'une partie centrale (core grammar) qui contient l'ensemble des règles non-marquées, qui se conforment aux principes de la théorie des grammaires. Elle contient également une partie périphérique, composée d'un ensemble de règles marquées [<sup>5</sup>] qui échappent aux principes généraux. La différence entre la partie centrale et la partie périphérique des grammaires permet de traiter les phénomènes marqués sans que le centre ne soit affecté [<sup>6</sup>].

La théorie de la grammaire centrale et la théorie de la marque sont reliées à la théorie de l'acquisition. Dans « *Markedness and core grammar* », Chomsky (1979a) fait l'hypothèse que les humains sont génétiquement munis d'une théorie de la marque et d'une grammaire centrale. Les règles de la grammaire centrale sont acquises facilement et en très peu de temps par l'enfant puisqu'elles se conforment aux principes de la théorie des grammaires, la faculté du langage.

En résumé, la faculté du langage rend possible l'acquisition et la théorie des grammaires en est une représentation abstraite. Elle contient une procédure d'évaluation qui limite le choix des grammaires potentielles, un ensemble de règles et de principes valables pour toutes les langues, associé à des paramètres de variation, et une théorie de la marque qui permet de distinguer les aspects marqués des aspects non-marqués des langues naturelles.

Ces hypothèses sur la faculté du langage et la théorie des grammaires renouvellent la perspective en matière de science du langage au vingtième siècle. Elles font de la linguistique une branche de la psychologie théorique qui étudie les mécanismes cognitifs en jeu dans l'acquisition du langage.

D'autres hypothèses ont été formulées sur l'organisation générale des grammaires et sur les propriétés des règles qui les constituent. La modularité et l'hypothèse de la dépendance de la structure figurent parmi les plus intéressantes.

---

<sup>5</sup> Si on suppose qu'au niveau de la théorie des grammaires toutes les règles ont une propriété P, et que dans une grammaire particulière, une règle R a la propriété P, la règle R est non-marquée ; si elle n'a pas la propriété P, ce sera une règle marquée.

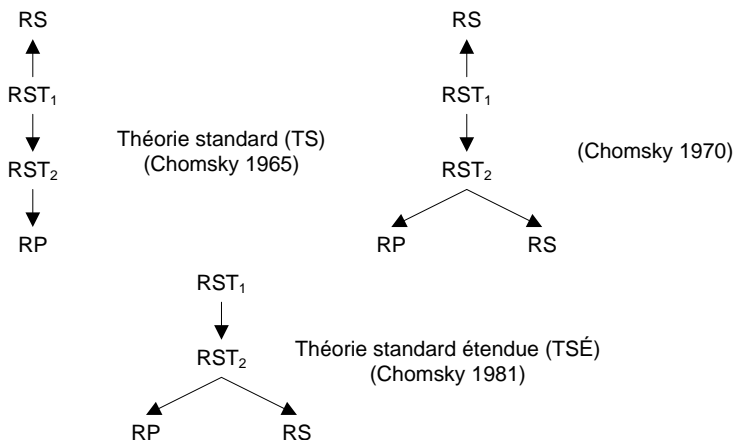
<sup>6</sup>Voir Belletti (à paraître), Chomsky (1979b), di Sciullo (1981a), Hendrick (1980) et Kean (1975) pour des études qui impliquent la théorie de la marque.

### 1.3.2 L'organisation modulaire

Selon l'hypothèse de la modularité, une grammaire est organisée en composants autonomes : les composants syntaxique, sémantique et phonologique. Chaque composant inclut des règles qui décrivent les propriétés des phrases de manière indépendante, de sorte que l'analyse des phrases est le résultat de l'interaction de différents composants de règles. Ces dernières engendrent (dérivent explicitement) des représentations grammaticales : les représentations phonétiques (RP) rendent explicite l'interprétation phonétique des phrases, les représentations sémantiques (RS) caractérisent leur signification et les représentations syntaxiques (RSY) décrivent leurs propriétés formelles. En outre, les principes de la théorie des grammaires interagissent avec les composants de règles et déterminent la bonne formation des représentations grammaticales.

À l'intérieur d'une même théorie, on assiste à des refontes successives du modèle de la grammaire. Ainsi en (9), les modèles proposés dans le cadre de la théorie standard (étendue) (TS(É)) associent des représentations grammaticales par des chemins différents. Néanmoins tous ces modèles actualisent une conception modulaire de la grammaire formée de plusieurs composants autonomes.

(9)



En outre, Chomsky fait l'hypothèse que les règles qui appartiennent au composant syntaxique doivent être étudiées indépendamment des règles qui appartiennent au composant sémantique ou au composant phonologique. Les règles d'un composant donné ne peuvent faire référence à des propriétés qui sont caractérisées par d'autres composants.

Cette idée est exprimée notamment par la thèse de l'autonomie de la syntaxe, formulée en (10). Elle constitue une hypothèse de travail intéressante qui a permis de découvrir plusieurs propriétés formelles des langues. Par exemple, le fait que certains aspects de l'interprétation sémantique des phrases sont établis à partir de propriétés configurationnelles des éléments linguistiques. C'est le cas des relations de coréférence, de l'identification du champ des opérateurs logiques, ou de l'identification de la structure en arguments. La théorie du liage et la théorie des rôles thématiques déterminent en partie ces propriétés, à partir des structures syntaxiques.

(10) Thèse de l'autonomie de la syntaxe

Aucune règle syntaxique ne peut faire référence à de l'information sémantique.

### 1.3.3 La dépendance de la structure

D'autres hypothèses sur la nature des règles grammaticales ont été proposées dans le cadre de la grammaire générative, dont l'hypothèse de la dépendance de la structure, formulée en (11). D'après cette hypothèse, les règles grammaticales mettent en jeu les configurations sous-jacentes aux phrases, et non la fonction des éléments qui les composent, ou leur ordre d'occurrence.

(11) La dépendance de la structure

Les règles des grammaires s'appliquent à des structures

Cette hypothèse est confirmée par plusieurs phénomènes, Chomsky (1975) montre que la règle de formation des questions fermées [<sup>7</sup>] de la grammaire de l'anglais met en jeu la structure des phrases et non l'ordre des mots qui les composent. Si c'était le cas, et que cette règle consistait à déplacer la première occurrence de la forme *is* au début de la phrase, comme dans l'exemple (12), il serait impossible d'expliquer le fait que cette opération produit de mauvais résultats en (13b).

- (12) a. The man is tall.  
       (L'homme est grand).  
       b. Is the man tall ?  
           └──────────┘  
           (est-ce que l'homme est grand ?)

- (13) a. The man who is tall is here.  
       (L'homme qui est grand est ici.)  
       b. \*Is the man who is tall is here ?

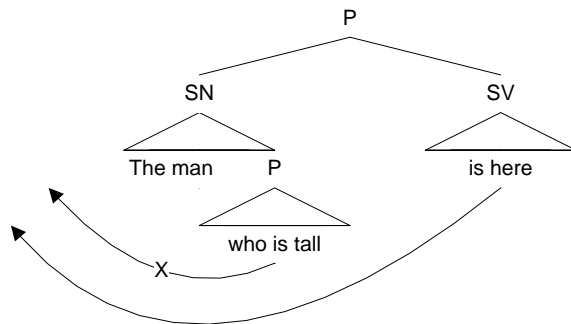
---

<sup>7</sup>Les questions fermées s'opposent aux questions ouvertes. Leur seule réponse possible est oui ou non, ce qui n'est pas le cas pour les questions ouvertes.

- ┌──────────┐
- c. Is the man who is tall here ?  
(Est-ce que l'homme qui est grand est ici ?)

L'explication de ce phénomène est une conséquence de l'hypothèse de la dépendance de la structure. La règle de formation des questions fermées s'applique aux phrases matrices et non aux phrases enchâssées [<sup>8</sup>], comme dans la représentation suivante :

(14)



Ce phénomène s'observe également dans d'autres langues. En français par exemple, la règle d'inversion du pronom sujet ne s'applique pas aux phrases enchâssées, comme le montrent les exemples (15). Ces faits viennent corroborer l'hypothèse de la dépendance de la structure.

- (15) a. Elle partira.  
 b. partira-t-elle.  
 c. Je me demande si elle partira.  
 d. \*Je me demande si partira-t-elle.

Cette hypothèse conduit à l'élaboration d'un ensemble de règles et de principes qui s'appliquent à des structures. Les principes de la théorie des Cas (abstrait) s'appuient sur des propriétés configurationnelles et contribuent à déterminer la bonne formation des structures sous-jacentes des phrases. Les principes de la théorie de la rection définissent une relation structurale de base à l'œuvre dans plusieurs autres théories.

Les programmes de recherche qui se développent dans le cadre de la grammaire générative ont favorisé le développement d'hypothèses intéressantes sur la nature des langues naturelles et des grammaires. L'idée que les phénomènes linguistiques sont décrits par des règles

<sup>8</sup>Une phrase matrice est dominée par un seul nœud P dans la structure syntaxique alors qu'une phrase enchâssée est dominée par plus d'un nœud P.

formelles, que ces règles appartiennent à des composantes autonomes, et qu'elles sont dépendantes de la structure constituent une série d'hypothèses nouvelles qui permettent à la linguistique d'accéder à une étape théorique où il est possible d'expliquer plusieurs phénomènes et de découvrir des faits nouveaux.

#### 1.4 Des faits nouveaux

L'existence de catégories grammaticales abstraites (sans réalisation phonétique) et leur liage à d'autres éléments des structures syntaxiques figurent parmi les découvertes les plus intéressantes de la grammaire générative. Dans la TSÉ, on suppose qu'il existe différents types de catégories abstraites, les unes dérivées par des règles de mouvements, comme dans les structures (16), les autres engendrées directement dans la position qu'elles occupent, comme dans les exemples (17). Les principes de la théorie des grammaires déterminent leur occurrence et leur liage à d'autres catégories à l'intérieur des structures.

- (16) a. Les documents<sub>i</sub> ont été retrouvés t<sub>i</sub>.  
b. Qui<sub>i</sub> Pierre voit-il t<sub>i</sub>.
- (17) a. Pierre permet à Lucie<sub>i</sub> [PRO<sub>i</sub> de partir]  
b. Pierre<sub>i</sub> promet à Lucie [PRO<sub>i</sub> de partir]  
c. Il est nécessaire [PRO de partir]

Les catégories abstraites permettent de décrire les dépendances structurales entre éléments. En outre, elles déterminent, en partie, l'interprétation sémantique des phrases. La présence de traces t<sub>i</sub> en position objet dans la structure (16) permet d'identifier la fonction grammaticale des constituants auxquels elles sont co-indexées, soit la fonction objet. La présence de PRO en (17) permet d'identifier la référence du sujet de l'infinitive enchâssée. Celui-ci peut référer au sujet, comme en (17a), à l'objet, comme en (17b), ou encore, s'il n'y a pas d'antécédent disponible dans la structure, comme en (17c), sa référence est arbitraire.

L'étude des principes qui régissent la distribution des catégories abstraites et leur liage à d'autres éléments de la structure se situe au centre des recherches élaborées dans le cadre de la TSÉ. On peut supposer que ces principes appartiennent à la théorie des grammaires puisque les grammaires particulières ne les manifestent pas directement. Ces phénomènes ne pouvaient être découverts par la linguistique structurale qui a écarté de son domaine d'étude les phénomènes qui se situent au-delà de l'observable.

## 1.5 Les règles et les principes

Considérons de plus près les systèmes de règles et de principes de la TSÉ qui interagissent dans la description et l'explication des phénomènes linguistiques.

Soit le système de règles suivant, proposé dans Chomsky (1981, 1982) :

(18) .le lexique

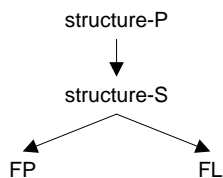
- . la syntaxe
  - i) les règles de base
  - ii) les règles transformationnelles
  
- .l'interprétation
  - i) les règles de forme phonétique
  - ii) les règles de forme logique

Le lexique spécifie les propriétés des éléments lexicaux (mots) qui ne peuvent être dérivées de règles ou de principes généraux. Il contient en outre, des règles qui expriment les régularités des structures lexicales. Ces règles sont soumises aux principes de la théorie des grammaires.

Les règles de base engendrent des structures profondes (structures-P) qui sont des représentations abstraites de la structure des fonctions grammaticales des phrases.

Les règles transformationnelles de la syntaxe déplacent des constituants et laissent des traces qui permettent de récupérer la position de ces constituants en structure-P. Les transformations syntaxiques associent les structures-P aux structures-S. Ces dernières contiennent des éléments abstraits *t*, PRO, qui sont interprétés sémantiquement. Appliquées aux structures-S, les règles de forme logique (FL) dérivent des FL, qui sont les seules structures accessibles à l'interprétation sémantique des phrases. Les structures-S sont également l'entrée des formes phonétiques (FP), qui interprètent les structures-S phonétiquement. La forme générale du modèle de grammaire qui a été décrit sommairement est présentée ci-dessous.

(19)





Les grammaires des langues naturelles, modelées selon le schéma (19) engendrent les structures sous-jacentes des phrases, en fournissant pour chacune d'elles une ou plusieurs dérivations. La capacité générative forte des grammaires leur permet d'engendrer des structures, alors que la capacité générative faible leur permet d'engendrer l'ensemble infini des phrases d'une langue.

Au système des règles (18) s'ajoute le système de principes (20) qui limite la forme et l'application des règles et qui assure la bonne formation des structures syntaxiques.

- (20) théorie X-barre
- théorie des Cas
- théorie de la rection
- théorie des limites
- théorie des rôles thématiques
- théorie du liage
- principe de projection

La théorie X-barre limite la classe des catégories grammaticales et la forme des règles de base. Les théories des Cas et de la rection imposent des contraintes sur la co-occurrence des éléments grammaticaux. La théorie des limites impose des contraintes sur l'application des règles de mouvements. La théorie des rôles thématiques et la théorie du liage imposent des conditions de bonne formation sur les structures syntaxiques, la première s'applique au niveau FL et la seconde s'applique au niveau des structures-S. Enfin, le principe de projection assure que les structures syntaxiques à tous les niveaux de représentation, c.-à-d. structure-P, structure-S et FL, soient des projections des propriétés des éléments lexicaux.

Ces principes appartiennent à la grammaire universelle et interagissent dans la description et l'explication des phénomènes linguistiques.

Voyons maintenant la capacité descriptive et explicative des systèmes de règles et de principes, tels qu'ils se manifestent dans une grammaire générative du français.

Nous considérons d'abord le lexique, puis les règles de base, les transformations et les règles de forme logique. Les principes de la théorie des grammaires interviendront au cours de l'exposé.

## Chapitre 2 : Le lexique

Le lexique est un composant de la grammaire qui présente les propriétés phonétiques, syntaxiques et sémantiques des éléments lexicaux (mots). Il contient une liste d'articles lexicaux qui spécifient les propriétés non-prédictibles des mots et un ensemble de règles qui énoncent des généralisations sur leur forme.

### 2.1 Les règles lexicales

Les recherches développées dans le cadre de la grammaire générative sur le lexique et la morphologie (p. ex. Halle (1973), Aronoff (1976), Williams (1981), Selkirk (1982), Lieber (1983)) ont exploré l'idée que les éléments lexicaux ont une structure sous-jacente dérivée par des règles, et soumise à des conditions de bonne formation. Cette approche éclaire, entre autres, les propriétés lexicales des mots du français.

Il y a des éléments lexicaux qui ne sont pas morphologiquement complexes, c'est le cas de mots tels que livre, de ou rouge. Ces éléments appartiennent à une catégorie grammaticale : nom (N), préposition (Prép), adjectif (A), et leur structure se limite au schéma suivant :

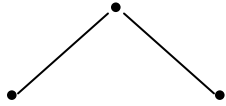
(1) N	Prép	A
Livre	de	rouge

Par contre, d'autres éléments lexicaux, tels que (2) par exemple, ont une structure sous-jacente plus complexe, qui peut être représentée par le graphe arborescent (3).

- (2)
- a. transporter  
transmettre  
transplanter  
etc.
  - b. pensable  
imaginable  
discutable  
etc.
  - c. indélicat  
inconséquent  
instable  
etc.
  - d. lentement  
doucement  
rapidement

etc.

(3)



En outre, on remarquera en (2) que certains morphèmes se retrouvent dans la même position dans plusieurs éléments lexicaux ; c'est le cas de -able, -ment, in-, trans-. Les exemples (4) montrent que les morphèmes ne se combinent pas au hasard, puisque seulement certaines suites de morphèmes sont possibles (\*indique les suites impossibles).

(4)

a. intransportable

b. \*transportablein  
\*transportinable  
\*abletransportin

c. considérablement

d. \*considérmentable  
\*mentconsidérable  
\*ableconsidérment

De plus, on observe en (5) que c'est la partie droite d'un élément lexical morphologiquement complexe qui détermine sa catégorie. Les préfixes trans- et in- ne déterminent pas la catégorie des mots auxquels ils appartiennent, ce sont les suffixes, tels que -isme qui ont cette propriété.

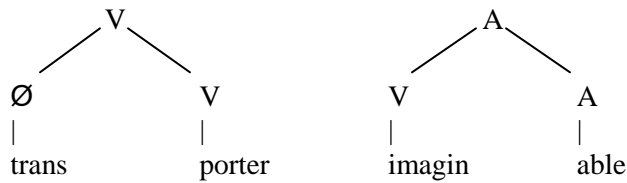
(5)

a. [<sub>V</sub> trans [<sub>V</sub> porter]]  
[<sub>A</sub> trans [<sub>A</sub> portable]]  
[<sub>N</sub> trans [<sub>N</sub> porteur]]

b. [<sub>N</sub> [<sub>N</sub> symbol] isme]  
[<sub>N</sub> [<sub>A</sub> vital] isme]  
[<sub>N</sub> [<sub>V</sub> arriv] isme]

Williams (1981) propose que la partie droite d'un élément lexical complexe soit analysée comme la « tête » de la structure lexicale. La tête lexicale, comme la tête syntaxique, que nous définirons plus loin, détermine la catégorie de l'ensemble de la structure, comme dans les représentations suivantes :

(6)



On observe en (6) que le suffixe -able appartient à une catégorie grammaticale. En fait, tous les morphèmes sont assignés à une catégorie, et ne peuvent être associés qu'à des éléments qui appartiennent à des catégories spécifiques : -able est un adjectif (A) qui est associé à un verbe (V), -ment est un adverbe (ADV) qui est associé à un A, trans- est une catégorie zéro (Ø) qui est associée à un V, à un A, ou à un N et ainsi de suite.

Les structures sous-jacentes telles que (6) sont dérivées par des règles dont la forme générale est présentée en (7). Plusieurs contraintes ont été proposées pour limiter la forme et l'application de ces règles (voir Aronoff (1976), Selkirk (1982), Lieber (1983)).

(7) X -> Y

où X et Y sont des catégories lexicales.

Les règles lexicales prédisent l'existence des mots réels et des mots possibles (qui peuvent éventuellement faire partie du lexique d'une grammaire donnée). En outre, elles permettent d'extraire la redondance du lexique, comme nous allons le voir tout de suite.

Considérons à titre d'exemple, la règle de formation des ADV en -ment à partir d'A.

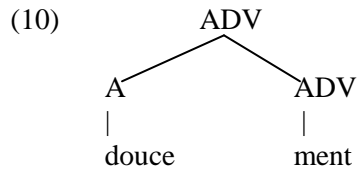
En français, la relation entre les A et les ADV est très productive, c'est-à-dire qu'à chaque A correspond généralement un ADV en -ment, et inversement, comme dans les exemples suivants :

(8)	doux	doucement
	lent	lentement
	fort	fortement
	sensible	sensiblement
	léger	légalement
	tendre	tendrement
	poli	poliment
	etc.	etc.

La relation entre les A et les ADV en -ment est exprimée par la règle (9). Cette règle a pour effet d'associer le morphème -ment à la forme féminine d'un A quelconque représenté par

la variable X<sup>9</sup>]. Elle effectue également un changement de catégorie, elle transforme un A en ADV. Cette règle décrit la structure des ADV en *-ment* qui existent déjà, et prédit la formation de nouveaux ADV. Cette structure peut être représentée par un graphe arborescent comme en (10) pour l'ADV doucement. On remarquera que c'est la partie droite de la structure qui détermine la catégorie de l'élément lexical.

(9) [A X] -> [ADV [A X] -ment]  
X = [+feminine]



Interprétée comme une règle de redondance [10], la règle (9) permet de simplifier le lexique : les ADV en *-ment* prédictibles à partir de cette règle n'ont pas à être spécifiés dans le lexique.

Les règles lexicales permettent ainsi d'extraire l'information redondante à l'intérieur de la grammaire. Les simplifications qu'elles entraînent sont souhaitables, étant donnés les critères d'évaluation des grammaires qui, parmi l'ensemble des grammaires possibles et compatibles avec les données de l'expérience, privilégient les grammaires les plus simples.

Considérons une autre règle lexicale, la règle qui gouverne la formation des A en *-eux* à partir de N. On observe en (11) que la relation entre les N et les A est assez productive, bien qu'il existe des lacunes dans la liste des A : il n'y a pas d'A en *-eux* qui correspond aux N *ramage* et *violence*.

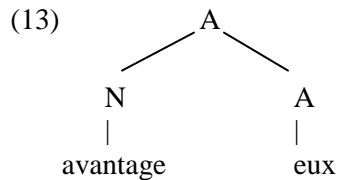
(11)	avantage	avantageux
	courage	courageux
	fougue	fougueux
	ombre	ombrageux
	ramage	*ramageux
	violence	*violenceux

<sup>9</sup> Il est à noter que dans certains cas c'est la forme masculine de l'adjectif qui sert de base à la formation d'adverbes en *-ment* : *savant/savamment*.

<sup>10</sup> Une règle de redondance R permet d'exprimer des généralisations sur des classes d'éléments à l'intérieur d'une grammaire G. Soit X et Y deux classes d'éléments de G, et une règle de redondance R qui relie X et Y : X -> Y. La règle prédit que si G contient un élément de X, alors elle contient également un élément de Y. Conséquemment, les éléments de Y n'ont pas à être spécifiés dans G, ils sont redondants.

La relation entre les N et les A est exprimée par la règle (12), qui élimine la spécification des A en -eux dans le lexique. La structure de ces A est présentée en (13). Encore ici, la partie droite de la structure détermine la catégorie de l'élément lexical.

(12)  $[_N X] \rightarrow [_A [_N X]].-eux]$

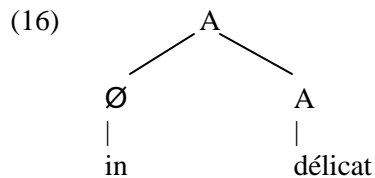


Certaines règles lexicales ne changent pas la catégorie des éléments auxquels elles s'appliquent. C'est le cas de la règle (15) qui exprime la relation entre les éléments lexicaux suivants :

(14)

délicat	indélicat
conséquent	inconséquent
tolérable	intolérable
vraisemblable	invraisemblable
avoué	inavoué
etc.	etc.

(15)  $[_A X] \rightarrow [_A in- [_A X]]$



La règle (15) associe le préfixe in- à des A. Elle permet de décrire la structure des éléments lexicaux tels que (16), et prédit la formation d'A du même type. En outre, elle élimine la redondance dans le lexique, puisque les A qui résultent de la règle (15) n'ont pas à être spécifiés.

Toutefois, les règles lexicales ne sont pas entièrement productives : il y a des éléments lexicaux qui échappent à leurs effets. Les exemples suivants sont des exceptions aux règles que nous avons considérées jusqu'ici.

(17)

- menu/\*menuement
- gros /\*grossement
- ramage/\*ramageux
- violence/\*violenceux

- e. joli/\*injoli
- f. tranquille/\*intranquille

Il y a différentes manières de traiter les exceptions dans le cadre de la grammaire générative. On peut supposer qu'une règle donnée, la règle (15) par exemple, qui décrit la structure des A en in-, est bloquée s'il existe déjà à l'intérieur du lexique, un A qui correspond au résultat de cette règle. Ainsi, l'existence des A laid et turbulent bloquerait l'application de cette règle aux A joli et tranquille. On peut également supposer que les règles qui ont des exceptions sont lexicalement gouvernées, c'est-à-dire que les éléments qui subissent ces règles sont spécifiés dans le lexique. Dans ce cas, chaque élément lexical est associé à des traits de règle (p. ex. +[R9], +[R12], +[15]), et les règles gouvernées lexicalement ne s'appliquent qu'aux éléments qui sont marqués de ces traits. Ainsi, la règle (9) s'applique à l'A doux qui a le trait +[R9] dans son entrée lexicale, mais non à l'A laid qui n'est pas marqué pour cette règle.

Les exceptions reçoivent ainsi un traitement systématique à l'intérieur de la grammaire, qui ne fait pas que les énumérer, mais qui cherche au contraire à trouver des mécanismes qui expliquent leur existence. En outre, la théorie de la marque, que nous avons décrite brièvement au premier chapitre, prédit l'existence d'exceptions aux règles des grammaires particulières et permet de différencier les propriétés marquées, ou d'exception, des propriétés non-marquées, ou régulières, des langues naturelles.

Les règles lexicales permettent également d'analyser la structure des composés tels que (18) et (20). Ils sont respectivement décrits par les règles (19) et (21). La règle (19) associe un V et un N pour former un N complexe, et la règle (21) associe un V et un constituant syntaxique SX pour former un N complexe.

- (18) attrape-nigaud  
casse-pied  
pisse-vinaigre  
croque-mort

(19)  $[_V X] \rightarrow [_N [_V X] - [_N X]]$

- (20) coupe-la-soif  
trompe-l'œil  
crève-la-faim  
pince-sans-rire  
boit-sans-soif

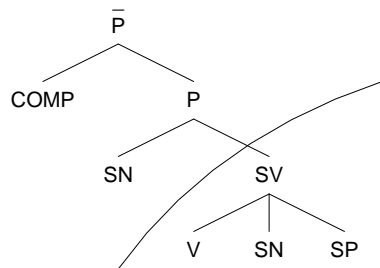
(21)  $[_V X] \rightarrow [_N [_V X] - [_{SX} X]]$

L'existence de composés tels que (20) est particulièrement intéressante pour une théorie grammaticale qui suppose l'existence de principes explicatifs. Nous avons proposé (voir di Sciullo 1982a) que la bonne formation de ces éléments lexicaux est gouvernée par la théorie de la rection, un des principes de la théorie des grammaires. Selon cette théorie, le V régit ses compléments, mais pas le sujet, ni les éléments extérieurs au SV. Les constituants syntaxiques SX des composés (20) sont régis par le V et ils doivent l'être pour que ces composés soient bien formés. En effet, des composés tels que (22) où la catégorie syntaxique n'est pas régie par le V sont impossibles.

- (22) \*le-gars-coupe \*<sub>[N]</sub> SV V]  
 \*sur-qui-saute \*<sub>[N]</sub> COMP V]

Alors que coupe-la-soif est une expression bien formée en français, le-gars-coupe ne l'est pas. Les structures agrammaticales (22) qui correspondent aux composés impossibles sont exclues puisque dans ces structures, le constituant syntaxique n'est pas régi par le V. Un V ne peut régir que les constituants à l'intérieur du SV dans la structure suivante :

(23)



Ainsi les règles lexicales expriment des régularités sur la forme des éléments lexicaux. Interprétées come des règles de redondance, elles simplifient les grammaires. La bonne formation des structures qu'elles engendrent (décrivent explicitement) dépend des principes de la théorie des grammaires. L'interaction des règles et des principes permet d'expliquer l'existence d'éléments lexicaux réels ou possibles, et d'exclure ceux qui sont impossibles.

## 2.2 Les entrées lexicales

Le lexique se compose également d'un ensemble d'articles lexicaux qui spécifient les propriétés non-prédictibles des mots.



La théorie des représentations lexicales s'est développée dans le cadre de la grammaire générative surtout par les travaux de Chomsky (1965, 1970, 1981), Jackendoff (1972, 1976), Bresnan (1978), Brame (1978). Ces travaux présentent des hypothèses différentes, bien qu'ils admettent tous l'idée que le lexique est un composant autonome de la grammaire qui spécifie les propriétés idiosyncratiques.

### 2.3 Les mots

Nous ne discuterons que de l'information qui est spécifiée dans les articles lexicaux, telle qu'elle est définie dans le cadre de Chomsky (1981).

La représentation (24) est un exemple d'article lexical [<sup>11</sup>]. Elle décrit explicitement les propriétés syntaxiques du V clarifier. Le trait [+V] indique la catégorie grammaticale de cet élément, le trait +[SN] spécifie sa sous-catégorisation, et le symbole <sup>θ</sup> TH indique le rôle thématique (θ) associé au complément.

(24) clarifier : +V  
                   + [SN]  
                           |  
                           <sup>θ</sup> TH

Nous discuterons plus loin des rôles θ et de la théorie qui gouverne leur assignation au niveau des structures syntaxiques. Considérons pour l'instant la sous-catégorisation lexicale, tout en gardant à l'esprit qu'à chaque élément de la sous-catégorisation lexicale est associé un rôle θ (c.-à-d. patient, but, source, position, etc.) [<sup>12</sup>] qui lui attribue un rôle sémantique d'argument.

Soulignons également que les articles lexicaux contiennent des informations sémantiques et phonétiques en plus des informations syntaxiques qui sont présentées de manière très générale dans la représentation suivante :

(25) +X  
           + [SX]  
               |  
               θ

Où X est une catégorie lexicale,  
 SX est une catégorie syntaxique,  
 et θ est un rôle thématique.

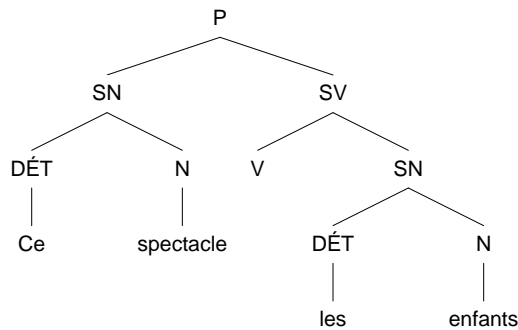
<sup>11</sup>Cet article lexical est incomplet mais l'information qu'il contient est suffisante pour illustrer notre propos.

<sup>12</sup>Voir Gruber (1965) et Jackendoff (1972, 1976) pour des définitions des rôles thématiques.

## 2.4 La sous-catégorisation

La sous-catégorisation est un mécanisme grammatical qui a été proposé dans Chomsky (1965) pour contraindre l'insertion des éléments lexicaux dans les structures syntaxiques (c.-à-d. les structures engendrées par les règles catégorielles). Cette insertion ne peut se faire librement puisque dans la structure (26) par exemple, il existe des restrictions sur le type de V qui peut être inséré sous le nœud SV.

(26)

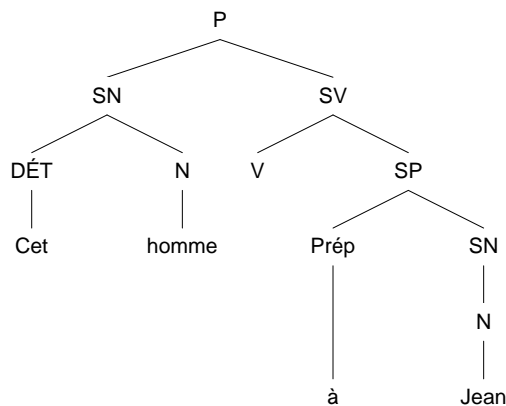


Les exemples (17) montrent que le V amuser peut être inséré dans cette structure, et que les V ressembler et dormir ne peuvent l'être puisque le résultat de cette insertion est agrammatical.

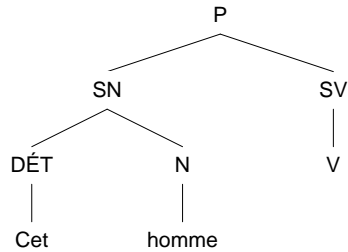
- (27)
- a. Ce spectacle amuse les enfants.
  - b. \*Ce spectacle ressemble les enfants.
  - c. \*Ce spectacle dort les enfants.

Le V ressembler peut être inséré dans la structure (28), ce qui n'est pas le cas du V dormir, qui par contre peut être inséré dans la structure (29).

(28)



(29)



Ces faits indiquent que seulement certains V peuvent être insérés dans certaines structures.

De plus, il est impossible à l'heure actuelle de prédire quels V peuvent figurer dans telle ou telle structure. A témoin, des V très proches sémantiquement tels amuser et divertir diffèrent quand au type de structure où ils peuvent apparaître. Les exemples (30) montrent que le V amuser, contrairement au V divertir ne peut figurer dans une phrase sans être suivi d'un complément.

- (30) a. Ce spectacle amuse les enfants.  
b. Ce spectacle divertit les enfants.  
c. \*Ce spectacle amuse.  
d. Ce spectacle divertit.

On dira que le V amuser est obligatoirement sous-catégorisé par un complément (d'objet) SN. Cette information idiosyncratique est spécifiée dans l'article lexical de ce V, comme en (31). La sous-catégorisation spécifie le type de structure dans laquelle un élément peut être inséré. Le V amuser par exemple, peut être inséré dans la position \_\_ qui précède immédiatement un complément (d'objet) SN, contenu dans le SV.

- (31) amuser : +V  
          +[\_\_SN]

Par ailleurs, on observe en (32) que le V amuser doit être suivi d'un complément SN<sup>[13]</sup> et qu'il ne peut être suivi immédiatement d'un complément SP ou  $\bar{P}$ .

- (32) a. Jean amuse Marie.  
b. \*Jean amuse.

---

<sup>13</sup>On suppose que le V amuser est sous-catégorisé par un SN même si dans des phrases telles que (i) et (iii), ce V n'est suivi d'aucun complément en surface. Dans le cas (i) le pronom clitique l(e) est engendré sous forme de SN en position d'objet en structure-P, de sorte que les traits de sous-catégorisation de ce V sont satisfaits à ce niveau de représentation syntaxique. La règle déplacer alpha permet de lier le clitique l(e) à la position du complément sous-catégorisé, comme en (ii) (voir Kayne (1975)) pour une première formulation de la règle de placement de clitique). Par ailleurs, d'autres mécanismes grammaticaux sont requis pour l'analyse de structure telles que (ii) qui ont une interprétation générique.

- c. \*Jean amuse à Marie.
- d. \*Jean amuse que Marie rencontre Pierre.

Pour rendre compte de ces faits, il serait possible d'ajouter des traits de sous-catégorisation négatifs à l'entrée du V amuser, comme suit :

(33) amuser : +V  
 +[  SN], -[  ], -[  SP], -[   $\bar{P}$ ]

Une solution plus simple est proposée dans Chomsky (1965) : seulement les sous-catégorisations positives sont marquées dans le lexique, et si un V n'est pas sous-catégorisé positivement par un complément, alors il est nécessairement sous-catégorisé négativement par rapport à ce complément.

Cette dernière solution est préférable puisqu'elle permet de traiter les sous-catégorisations positives et négatives tout en simplifiant la grammaire : elle limite les sous-catégorisations qui doivent être spécifiées dans le lexique. Dans le cas du V amuser, c'est l'entrée (31) qui est choisie plutôt que l'entrée plus complexe (33).

Considérons maintenant le cas du V divertir. Ce V peut être suivi immédiatement d'un complément (d'objet) SN, ou d'aucun complément :

- (34) a. Ce spectacle divertit les enfants.
- b. Ce spectacle divertit.

On dira que ce V est facultativement sous-catégorisé par un complément SN. Plutôt que d'assigner deux traits de sous-catégorisation (c.-à-d. +[  ] et +[  SN]) à l'article lexical de ce V, on utilisera des parenthèses, comme en (35). Les parenthèses permettent d'abrégier les grammaires, elles indiquent que l'élément qu'elle contient est facultatif. Selon (35) le V divertir peut être inséré dans une structure SV qui contient ou pas un complément SN.

(35) divertir : +V  
 +[  (SN)]

Telle que présentée dans Chomsky (1965), la sous-catégorisation permet de limiter la règle d'insertion des éléments lexicaux dans les structures syntaxiques. D'après la règle (36),

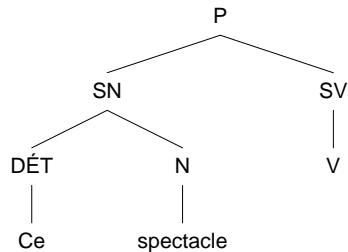
l'insertion d'un élément lexical dans une structure dépend de ses traits de sous-catégorisation et de ses traits catégoriels [<sup>14</sup>].

(36) Règle d'insertion lexicale

- Insérer un élément lexical I sous un nœud X d'une structure syntaxique si
- i) I est de la catégorie X, et que
  - ii) les traits de sous-catégorisation de I sont compatibles avec la structure.

La règle (36) permet, par exemple, d'insérer un V tel que divertir, mais non un V tel que ressembler dans la structure suivante :

(37)



Ainsi, la sous-catégorisation et la règle d'insertion lexicale rendent compte du fait que seulement certains V peuvent figurer dans certaines structures. Considérons d'autres cas de sous-catégorisation verbale.

Il y a des V en français qui doivent être suivis de deux compléments, c'est le cas du V mettre. À partir des exemples (38), on peut supposer que la sous-catégorisation de ce V est telle que spécifiée en (39).

- (38) a. \*Jean met.  
 b. \*Jean est un livre.  
 c. Jean met un livre sur la table.

- (39) mettre : +V  
 +[\_\_SN SP]

Toutefois, il y a des restrictions sur le type de préposition qui introduit le second complément de mettre. On observe en (40) que les prépositions (Prép) sur, près, à côté, devant

<sup>14</sup>La règle d'insertion lexicale proposée dans Chomsky (1965) s'applique aux structures engendrées par les règles catégorielles de la base. Dans Chomsky (1981) toutes les règles, y compris la règle d'insertion lexicale, sont facultatives, c'est-à-dire qu'elles peuvent ou non s'appliquer, à moins que des principes de la théorie des grammaires n'en forcent l'application.

sont possibles, alors que les Prép à, de, pour, sans sont exclues. Pour rendre compte de ce fait, on peut supposer qu'un rôle  $\theta$  POSITION (POSI) est associé au SP de la sous-catégorisation du V mettre, comme en (41). Ce rôle  $\theta$  est également associé au complément des Prép sur, devant, près, à côté, contrairement aux Prép à, de, pour, sans. La grammaire peut alors exclure les structures (40b) et accepter les structures (40a).

- (40) a. Jean met le livre  $\left\{ \begin{array}{l} \text{sur} \\ \text{à côté de} \\ \text{près de} \\ \text{devant} \end{array} \right\}$  la table.
- b. \*Jean met le livre  $\left\{ \begin{array}{l} \text{à} \\ \text{de} \\ \text{pour} \\ \text{sans} \end{array} \right\}$  la table.

- (41) mettre : +V  
 +[  SN SP]  
 |  
 $\theta$ POSI

Par ailleurs, certains V sont sous-catégorisés par des compléments SP avec une Prép spécifique. C'est le cas du V appartenir. Les exemples (42) montrent que ce V doit être suivi immédiatement d'un complément SP qui contient la Prép à.

- (42) a. \*Jean appartient.  
 b. \*Jean appartient une organisation secrète.  
 c. \*Jean appartient à une organisation secrète.  
 d. \*Jean appartient  $\left\{ \begin{array}{l} \text{de} \\ \text{vers} \\ \text{sur} \\ \text{pour} \\ \text{avec} \end{array} \right\}$  une organisation secrète.

Ces faits indiquent que la Prép à doit être spécifiée dans l'article lexical de ce V comme (43). Nous laissons au lecteur le loisir de vérifier la sous-catégorisation du V réagir.

- (43) appartenir : +V  
 +[   à SN]  
 (44) réagir : +V  
 +[   (à SN)]

D'autres V sont sous-catégorisés par des compléments  $\bar{P}$ , c'est le cas du V supposer, comme le montrent les exemples suivants :

- (45) a. \*Jean suppose.  
 b. \*Jean suppose un scénario.  
 c. \*Jean suppose à Marie.  
 d. \*Jean suppose un scénario à Marie.  
 e. Jean suppose que Pierre est un anarchiste.

- (46) supposer : +V  
 +[ $\bar{P}$ ]

Jusqu'ici, nous n'avons considéré que la sous-catégorisation verbale : les autres catégories lexicales, N, Prép, à, doivent également être sous-catégorisées.

Les Prép doivent être sous-catégorisées par rapport au type de complément qui peut ou doit figurer dans la structure qui les contient, soit SP. Ainsi par exemple, la Prép vers doit être suivie immédiatement d'un complément SN, alors que la Prép pour peut être suivie immédiatement par un SN ou par un  $\bar{P}$ . Les exemples (47) et (49) illustrent ces faits, et on trouve en (48) et (50) les sous-catégorisations de ces Prép.

- (47) a. \*Il se dirige vers.  
 b. Il se dirige vers la fenêtre.  
 c. \*Il se dirige vers à la fenêtre.

- (48) vers : +Prép  
 +[ $\bar{SN}$ ]

- (49) a. \*Il est venu pour.  
 b. Il est venu pour le concert.  
 c. Il est venu pour que Marie le rencontre.  
 d. \*Il est venu pour à Marie.

- (50) pour : +Prép  
 +[ $\bar{SN}$ ], +[ $\bar{P}$ ]

Les A doivent également être sous-catégorisés par rapport au type de complément qui peut figurer dans la structure qui les contient, soit SA. L'A fidèle par exemple, peut être suivi d'un SP qui contient exclusivement la Prép à, ce que montrent les exemples (51), alors que l'A pour a une sous-catégorisation différente.

- (51) a. Jean est fidèle.  
 b. Jean est fidèle à ses convictions.  
 c. \*Jean est fidèle  $\left. \begin{array}{l} \text{de} \\ \text{avec} \\ \text{pour} \\ \text{vers} \end{array} \right\}$  ses convictions.

(52) fidèle : +A  
 +[ $\_\_$  à SN]

- (53) a. Jean a peur.  
 b. Jean a peur  $\left. \begin{array}{l} \text{de} \\ \text{pour} \end{array} \right\}$  Pierre.  
 c. \*Jean a peur  $\left. \begin{array}{l} \text{vers} \\ \text{sur} \\ \text{contre} \end{array} \right\}$  Pierre.  
 d. Jean a peur que Pierre rencontre Paul.

(54) peur : +A  
 +[ $\_\_$  ( $\left. \begin{array}{l} \text{de} \\ \text{pour} \end{array} \right\}$  SN)], +[ $\_\_$  ( $\bar{P}$ )]

Les N sont également sujets à la sous-catégorisation. Un N tel que trajet peut être suivi de deux compléments SP, le premier introduit par la Prép de et le second par la Prép à. Alors que le N vote est sous-catégorisé par un SP qui contient les Prép pour ou contre. Les exemples suivants illustrent ces faits.

- (55) a. le trajet.  
 b. \*le trajet de Paris.  
 c. le trajet de Paris à Rome.

(56) trajet : +N  
 +[ $\_\_$  (de SN à SN)]

- (57) a. le vote  
 b. le vote  $\left. \begin{array}{l} \text{pour} \\ \text{contre} \end{array} \right\}$  Reagan.  
 c. \*le vote  $\left. \begin{array}{l} \text{à} \\ \text{vers} \\ \text{sans} \\ \text{devant} \end{array} \right\}$  Reagan.

(58) vote : +N  
 +[ $\_\_$  ( $\left. \begin{array}{l} \text{pour} \\ \text{contre} \end{array} \right\}$  SN)]

Nous avons mentionné que les compléments qui sous-catégorisent les éléments lexicaux sont à l'intérieur de certains types de structure : les compléments qui sous-catégorisent les V sont à l'intérieur du SV, les compléments qui sous-catégorisent les Prép sont à l'intérieur du SP, et



ainsi de suite pour les autres catégories. Nous n'avons pas spécifié toutefois leur position exacte. Pour ce faire, il convient de considérer également les compléments qui ne sous-catégorisent pas. Considérons le cas des V.

Il y a des différences empiriques entre les compléments qui sous-catégorisent les V (c.-à-d. les objets) et les compléments adverbiaux de temps, de lieu, de manière qui n'entrent pas dans la sous-catégorisation verbale.

Les compléments qui sous-catégorisent sont soumis à plusieurs restrictions. Ils se trouvent généralement en position adjacente au V, comme en (59a)<sup>[15]</sup>, ils ne peuvent être antéposés, comme en (60c), et dans certains cas, leur présence est obligatoire pour la bonne formation des structures, comme en (61). Les compléments qui ne sous-catégorisent pas ne sont pas soumis à ces restrictions.

- (59) a. Paul traversera ce pont vers deux heures.  
b. \*Paul traversera vers deux heures ce pont.

- (60) a. Paul écrit un roman la nuit.  
b. La nuit Paul écrit un roman.  
c. \*Un roman Paul écrit la nuit.

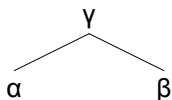
- (61) a. Paul exige des explications.  
b. \*Paul exige.  
c. Paul infère le résultat.  
d. \*Paul infère.

Ces différences entre les compléments sont exprimées au niveau des représentations syntaxiques. Les compléments qui sous-catégorisent sont en position « sœur »<sup>[16]</sup> de V, c'est-à-dire que le V et le complément qui le sous-catégorise sont dominés immédiatement par le même nœud. Dans la structure (62), le complément qui sous-catégorise est dominé immédiatement par

---

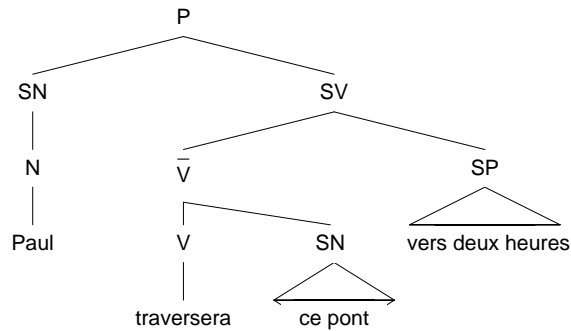
<sup>15</sup>Ces compléments n'occupent pas toujours cette position en surface. L'ordre des compléments à droite du V dépend également de facteurs stylistiques, tels que la longueur, comparer (59b) à l'exemple suivant qui est acceptable et où les deux compléments sont d'égale longueur : Paul traversera à l'aube ce pont. Une analyse possible de ce phénomène est de supposer que les compléments ne sont pas ordonnés dans la syntaxe, mais seulement dans le composant de forme phonétique. Voir di Sciullo (1981b).

<sup>16</sup>Plus généralement, dans une structure arborescente, un nœud alpha est en position sœur par rapport à un autre nœud beta, si alpha et beta sont immédiatement dominés par le même nœud gamma, comme dans la représentation suivante :



le nœud V, et se trouve en position sœur de  $\bar{V}$ . Le complément non sous-catégorisant n'est pas en position sœur de V, il est dominé par SV<sup>[17]</sup>.

(62)



Les compléments qui sont associés aux autres catégories lexicales, N, A, Prép, ont un traitement similaire, de sorte que :

(63) tout complément qui sous-catégorise un élément lexical X est en position sœur de X en structure-P.

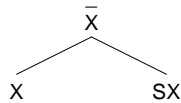
En résumé, la sous-catégorisation est un mécanisme grammatical qui exprime le fait que les éléments lexicaux ont des propriétés idiosyncratiques par rapport aux types de compléments qui peuvent ou qui doivent les suivre immédiatement en structure-P. Elle prend la forme de traits contextuels positifs : +[  ], +[  SN], etc. Ces traits sont spécifiés dans les articles lexicaux et limitent l'application de la règle d'insertion lexicale.

La forme générale des traits de sous-catégorisation est présentée en (64). Le trait vertical indique la position de l'élément lexical X dans la structure-P. Le symbole SX peut être remplacé par n'importe quelle suite de compléments qui peut ou doit figurer dans la structure (65) : une représentation générale des structures (66).

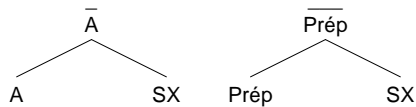
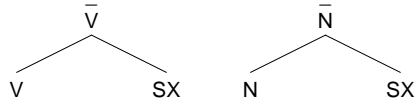
(64) X : +[  SX]

(65)

<sup>17</sup>Cette analyse des compléments est généralement adoptée dans la TSÉ (voir Chomsky (1981)) et a été proposée à partir des faits de l'anglais. En français, comme nous l'avons déjà noté, certains arguments sur lesquels elle s'appuie ne s'appliquent pas. Une analyse différente est proposée dans di Sciullo (1981a) à partir des faits du français.



(66)



## 2.5 Le principe de projection

Le lexique est un composant indépendant des grammaires qui présente les propriétés des éléments lexicaux.

La théorie des grammaires inclut un principe universel, le principe de projection (67) [<sup>18</sup>] qui assure que toute l'information lexicale soit présente au niveau des représentations syntaxiques.

(67) Principe de projection

Les représentations syntaxiques à tout  
niveau de description sont des projections du lexique.

En particulier, l'information qui a trait à la sous-catégorisation des éléments lexicaux, ainsi que leurs propriétés thématiques doivent être présentes à chaque niveau de représentation syntaxique : au niveau des structures-P, des structures-S et des formes logiques. Le principe de projection permet de préserver l'information lexicale à tous les stades des dérivations.

Considérons maintenant les propriétés des composantes de règles qui engendrent les représentations syntaxiques.

---

<sup>18</sup>Le principe (67) est une première formulation du principe de projection. Voir les chapitres 3 et 5 pour d'autres formulations.

## Chapitre 3 : La base

Le composant de base [<sup>19</sup>] d'une grammaire générative du français décrit plusieurs aspects de la forme syntaxique d'un sous-ensemble de phrases du français. Il est défini par les symboles suivants :

$$(1) \langle \Sigma, V_n, V_t, R \rangle$$

Le symbole  $\Sigma$  désigne l'élément initial de la grammaire. Cet élément est une catégorie P (P, P-barre ou P-deux-barres) et il commence les dérivations. Le vocabulaire non-terminal ( $V_n$ ) contient l'ensemble des catégories grammaticales N, V, A-barre, N-deux-barres, etc. Le vocabulaire terminal ( $V_t$ ) contient les éléments lexicaux et les catégories abstraites : alpha, PRO, etc. Les propriétés générales du  $V_n$  et du  $V_t$  sont déterminées par les principes de la théorie des grammaires, et chaque grammaire fixe les valeurs des paramètres associés à leurs vocabulaires.

La grammaire contient, en outre, un ensemble de règles ( $R$ ) qui engendrent (décrivent explicitement) des dérivations dont l'essentiel peut être représenté sous forme de graphes arborescents. Les règles s'appliquent à des éléments qui constituent ce qu'il est convenu d'appeler l'entrée de la règle, les éléments qui résultent de l'opération constituent la sortie de la règle :

$$(2) R : \text{Entrée} \rightarrow \text{Sortie}$$

Les règles catégorielles sont soumises aux principes de la théorie des grammaires, et chaque grammaire en fixe les valeurs paramétriques. Elles engendrent des structures-P qui représentent la structure des phrases avant l'application des transformations.

### 3.1 Les règles de réécriture

La forme générale des règles catégorielles est présentée par le schéma (3). Ce sont des règles formelles : les termes qui les composent, et l'opération qu'elles effectuent sont définis explicitement. Les symboles X, Y appartiennent au vocabulaire de la grammaire, défini indépendamment, et l'opération qu'elles effectuent est la substitution : le symbole à gauche de la flèche est remplacé par la suite non nulle de symboles à droite. Le signe de réécriture «  $\rightarrow$  » déclenche cette opération. Les règles formulées selon le schéma (3) sont désignées par l'expression règles de réécriture indépendantes du contexte ». Elles s'appliquent aveuglément, là

---

<sup>19</sup>Le composant de base est également désigné par les expressions : le composant catégoriel, la grammaire syntagmatique (phrase structure grammar, constituent structure grammar) grammaire indépendante du contexte (context tree grammar).

où elles peuvent s'appliquer, et ne sont pas ordonnées entre elles, sauf pour la règle qui développe le symbole initial de la grammaire, qui doit s'appliquer avant toutes les autres.

$$(3) X \rightarrow \dots Y \dots$$

En outre, les règles de réécriture sont récursives, elles peuvent se réappliquer à leur sortie, et intervenir plusieurs fois dans une même dérivation. La présence d'un même symbole à gauche et à droite de la flèche déclenche la réapplication d'une règle, comme c'est le cas pour les symboles A et C des règles (4). Le résultat de l'application de ces règles est présenté en (5) : les suites de symboles engendrées par les règles (4) sont potentiellement infinies.

$$(4) \begin{array}{l} \text{a. } A \rightarrow B C \\ \text{b. } C \rightarrow A \end{array}$$

$$(5) \begin{array}{ll} A & \text{(symbole initial)} \\ B C & \text{(par la règle (4a))} \\ B A & \text{(par la règle (4b))} \\ B B C & \text{(par la règle (4a))} \\ B B A & \text{(par la règle (4b))} \\ B B B C & \text{(par la règle (4a))} \\ \dots & \dots \end{array}$$

Par leurs propriétés formelles et récursives, ces règles ont la capacité d'engendrer des ensembles infinis, tels que l'ensemble des phrases des langues naturelles.

### 3.2 Abréviations

Les conventions d'abrégations des grammaires : parenthèses, accolades, astérisques, simplifient les grammaires en réduisant le nombre de règles qui leur appartiennent.

Les parenthèses indiquent qu'un symbole donné est optionnel, les accolades marquent un choix disjonctif parmi les éléments qu'elles contiennent, et l'astérisque signale la répétition potentiellement infinie d'un même symbole. Les règles a) et b) des exemples (6) à (8) sont équivalentes, bien que formellement distinctes.

$$(6) \begin{array}{l} \text{a. } A \rightarrow S (C) \\ \text{b. } A \rightarrow B \\ \quad A \rightarrow B C \end{array}$$

$$(7) \text{ a. } A \rightarrow S \left\{ \begin{array}{l} C \\ D \\ E \end{array} \right\}$$

- b. A -> S C
- A -> B D
- A -> B E

- (8) a. A -> B C\*
- b. A -> B C
- A -> B C C
- A -> B C C C
- A -> B C C C C
- A -> B C C C C C
- .....

La métrique d'évaluation des grammaires adoptée dans la TS(É) privilégie parmi des grammaires équivalentes du point de vue de l'adéquation descriptive, celle qui est la plus simple : c'est-à-dire celle qui contient le plus petit nombre de symboles ou de règles. Les conventions d'abréviation contribuent à simplifier les grammaires et à les rendre, par là, plus explicatives.

### 3.3 Les dérivations

La grammaire engendre (décrit explicitement) des dérivations. Une dérivation est une ensemble de suites de symboles décrits explicitement par une grammaire.

Considérons un autre cas très simple de dérivation. Soit la grammaire (9)<sup>[20]</sup>. Cette grammaire se compose de trois règles, et décrit explicitement la structure sous-jacente d'un sous-ensemble de phrases du français en assignant à chacune une ou plusieurs dérivations. Ainsi, la dérivation (10), est une des dérivations possibles que la grammaire (9) assigne à la phrase cet homme écrit un roman.

- (9) a. P -> SN SV
- b. SN -> (DÉT) N (P)
- c. SV -> V (SN)
  
- (10) P
- SN SV
- DÉT N SV
- DÉT N V SN
- DÉT N V DÉT N
- Cet N V DÉT N
- Cet homme V DÉT N
- Cet homme écrit DÉT N
- Cet homme écrit un N
- Cet homme écrit un roman

---

<sup>20</sup>Les règles de cette grammaire sont formulées selon la TS (Chomsky 1965), bien que pour les besoins de l'exposé, elles soient très simplifiées.

Chaque ligne d'une dérivation est obtenue par l'application d'une règle de la grammaire. Une dérivation est terminée lorsque, comme dans la dérivation (10), tous les symboles sont des symboles terminaux. La dernière ligne ne contient pas de symboles qui peuvent être réécrits.

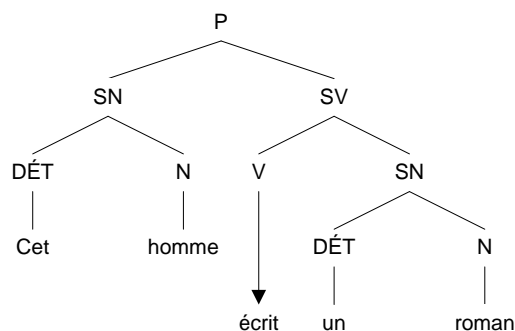
Il est possible de construire plusieurs dérivations pour une même phrase, en modifiant l'ordre dans lequel les symboles se sont développés. En choisissant d'appliquer la règle (9c) avant la règle (9b), ou en appliquant la règle d'insertion lexicale dans un ordre différent, on obtient en (11) une autre dérivation possible de la phrase cet homme écrit un roman à partir de la grammaire (9). Les dérivations (10) et (11) sont équivalentes bien que formellement distinctes.

- (11) P  
 SN SV  
 SN V SN  
 DÉT N V SN  
 DÉT N V DÉT N  
 DÉT homme V DÉT N  
 DÉT homme V DÉT roman  
 DÉT homme écrit DÉT roman  
 Cet homme écrit DÉT roman  
 Cet homme écrit un roman

### 3.4 Les représentations syntaxiques

L'information essentielle d'une dérivation peut être présentée sous forme graphique, par un indicateur syntagmatique (IS) ou par un système de parenthèses. Ainsi, la dérivation (10) par exemple, peut être représentée par l'IS (12) ou par la parenthésisation (13).

- (12)



- (13) [P[SN[DÉTcetDÉT][NhommeN]SN][SV[VécritV][SN[DÉTunDÉT][NromanN]SN]SV]

Les IS sont des graphes composés de nœuds et de branches. Les nœuds sont étiquetés par les symboles du  $V_n$  de la grammaire et les branches indiquent les relations entre les nœuds. La

séquence terminale de l'IS se compose exclusivement d'éléments qui appartiennent au  $V_t$  de la grammaire. Dans l'IS (12), la séquence terminale est la suite formée des termes suivants : cet homme écrit un roman. La séquence de catégories qui domine immédiatement ces termes est la séquence préterminale de l'IS. C'est à cette dernière que s'applique la règle d'insertion lexicale.

La parenthésisation (13) est équivalente à l'IS (12). Des paires de parenthèses remplacent chaque branche de cette structure, elles contiennent les catégories qui étiquettent les nœuds de l'IS, les termes de la suite terminale de l'IS sont inclus à l'intérieur des parenthèses.

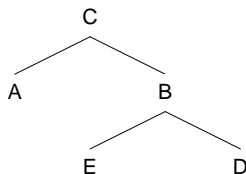
La conversion des structures arborescentes telles que (12) à des structures parenthésisées telles que (13) est simple et peut être décrite comme une procédure qui consiste à ouvrir une parenthèse pour chaque nœud de l'IS à convertir, et à la refermer lorsqu'on a été spécifiés tous les symboles, y compris les symboles terminaux de l'IS.

Ainsi, les dérivations engendrées par les règles du composant catégoriel peuvent être représentées sous forme de graphe arborescent ou encore sous forme parenthésisée. Ces représentations mettent en évidence l'essentiel de la structure sous-jacente d'un sous-ensemble de phrases engendrées par la grammaire.

### 3.5 Les relations structurales

À partir des IS, il est possible d'identifier les relations structurales qui interviennent dans la description des propriétés syntaxiques des phrases. Il s'agit des relations structurales suivantes : « domine », « domine immédiatement », « précède », « c-commande ». L'arborescence (14) nous permettra d'illustrer ces relations avant de les définir explicitement.

(14)



Dans la structure (14), le nœud C domine tous les autres nœuds. Il domine immédiatement les nœuds A et B, alors que les nœuds E et D sont dominés immédiatement par le nœud B. Le nœud E précède le nœud D, et le nœud A précède les nœuds B et D. Le nœud A c-commande les nœuds B, D et E, par contre, les nœuds E et D ne c-commandent pas le nœud A.



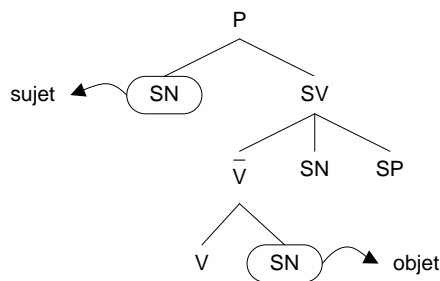
Voici des définitions explicites de ces relations structurales :

- (15) a. Un nœud alpha domine un nœud beta si alpha est supérieur à beta, et s'il y a au moins une branche qui relie alpha à beta.
- b. Un nœud alpha domine immédiatement un nœud beta si alpha est supérieur à beta, et s'il n'y a qu'une seule branche qui relie alpha et beta.
- c. Un nœud alpha précède un nœud beta si alpha est à gauche de beta.
- d. Un nœud alpha c-commande un nœud beta si alpha et beta ne se dominant pas l'un l'autre, et si la première projection maximale<sup>21</sup>  $\Phi$  qui domine alpha domine également beta.

Ces relations structurales ont un rôle central à l'intérieur de la théorie des grammaires. Ce sont des concepts de base qui interviennent dans la description des propriétés syntaxiques des phrases.

Ainsi, par exemple, les fonctions grammaticales sujet et objet sont définies configurationnellement par la relation structurale (15b). Le sujet est le SN immédiatement dominé par le nœud P et l'objet est le SN immédiatement dominé par le nœud V-barre au niveau des structures-P. Les SN encadrés en (16) désignent les fonctions grammaticales.

(16)



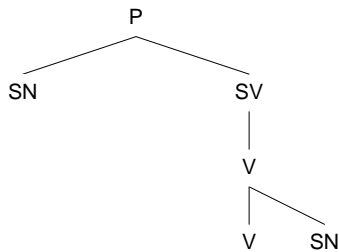
L'identification des relations grammaticales à partir des structures syntaxiques permet de déterminer un aspect de l'interprétation des phrases. À témoin, les phrases (17a) et (17b) qui ont des interprétations différentes, mais la même structure, soit (18). La différence sémantique tient à l'identification des éléments qui occupent les positions sujet et objet : ces éléments correspondent

<sup>21</sup>Les catégories suivantes sont des projections maximales : SN, SV, SA, SP. Voir plus loin pour une définition de catégories maximales ( $X^{\max}$ ).

aux arguments du prédicat sémantique associé au verbe manger. Dans la structure sous-jacente de la phrase (17a) le sujet et/ou le premier argument du prédicat est [SN le chat], l'objet et/ou le second argument du prédicat est [SN la souris], dans le cas (17b) c'est l'inverse.

- (17) a. Le chat mange la souris.  
 b. La souris mange le chat.

(18)

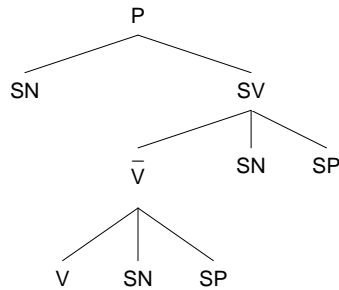


Certaines transformations syntaxiques ont comme effet de modifier la position des éléments sujet et objet, c'est le cas des structures (19) qui sont très proche du point de vue de l'interprétation sémantique, mais qui diffèrent notamment par la position des SN sujet et objet. Nous verrons au chapitre 4 que la théorie des traces permet d'identifier les fonctions grammaticales en structure-S, après que les transformations se sont appliquées.

- (19) a. [<sub>P</sub> le chat mange la souris]  
 b. [<sub>P</sub> la souris est mangée par le chat]

La c-commande est une relation structurale fondamentale à l'intérieur de la TSÉ. Elle permet de définir formellement plusieurs propriétés des structures de phrases. En particulier, le liage entre éléments grammaticaux (voir chapitre 4) est basé sur la c-commande. En outre, cette relation structurale permet de différencier le domaine des compléments associés aux catégories lexicales, du domaine du sujet. Un V c-commande ses compléments (sous-catégorisants et non sous-catégorisants), mais il ne c-commande pas son sujet. En effet, le V et ses compléments sont dominés par la projection maximale SV, ce qui n'est pas le cas du SN en position sujet :

(20)



La relation de c-commande, et les autres relations structurales définies sur les structures syntaxiques, domine, domine immédiatement, précède, ont un rôle central à l'intérieur de la théorie grammaticale, et permettent de décrire explicitement plusieurs propriétés syntaxiques.

## 3.6 La théorie X-Barre

### 3.6.1 Justification

Telles que nous les avons définies jusqu'ici, les règles catégorielles sont très peu contraintes. La théorie X-barre, un des principes de la théorie des grammaires, vient restreindre leur forme. Elle limite la classe des grammaires accessibles à l'enfant en imposant des contraintes sur la forme des règles catégorielles et sur les propriétés du  $V_n$  des grammaires.

La théorie X-barre a été proposée dans Chomsky (1970) pour combler les lacunes de théories non restrictives telles que la théorie de la sémantique générative (TSG) et la théorie standard (TS). Ces modèles n'imposaient pas de restrictions sur le  $V_n$  ni sur les règles syntaxiques. Ces dernières étaient devenues des mécanismes très puissants, capables d'effectuer plusieurs opérations différentes et d'engendrer des structures qui n'appartenaient à aucune langue naturelle. En outre, ces modèles ne pouvaient décrire simplement, c'est-à-dire avec peu de symboles ou de règles, les phénomènes linguistiques puisqu'ils n'incluaient pas de théorie restrictive des catégories grammaticales. Considérons ces lacunes de plus près.

Dans la TS, il n'y avait pas de restrictions sur la classe des catégories grammaticales propres aux langues naturelles, de sorte que n'importe quelle catégorie pouvait être postulée. Dans la TSG, les catégories lexicales N, A, V, Prép, Aux, etc. étaient dérivées de super-catégories « N » et « V » par une série de transformations. La TSG voulait par là exprimer les parallélismes entre catégories, mais elle compliquait les grammaires.

La théorie X-barre impose des restrictions sur la classe des catégories grammaticales possibles et permet d'exprimer les parallélismes entre les catégories sans compliquer les

grammaires. Ainsi, par exemple, dans Chomsky (1970, 1981) les catégories lexicales majeures : V, N, A<sup>[22]</sup> sont définies par les traits syntaxiques [± N] et [± V] comme suit :

$$(21) \quad \begin{array}{l} V : \begin{pmatrix} +V \\ -N \end{pmatrix} \\ N : \begin{pmatrix} -V \\ +N \end{pmatrix} \\ A : \begin{pmatrix} +V \\ +N \end{pmatrix} \end{array}$$

Ceci permet d'exprimer formellement la distinction que les grammairiens traditionnels ont établie entre les prédicats (V) et les substantifs (N). En outre, le système de traits [± N] [± V] présente une analyse abstraite des catégories lexicales majeures.

Il y a des avantages à définir ainsi les catégories lexicales par des traits syntaxiques : en plus de fournir une définition explicite des catégories, cette théorie permet de limiter la classe des catégories lexicales possibles pour toutes les langues naturelles.

En effet, la théorie X-barre prédit que dans les cas non-marqués, toutes les langues ont des catégories lexicales majeures, puisque les traits [± N] et [± V] appartiennent à la théorie des grammaires, ce qui n'est pas le cas pour les catégories lexicales mineures : déterminant (DÉT), complémenteur (COMP), clitique (CL), modal (M), quantificateur (Q), etc... Cette prédiction se vérifie dans les faits. En hindi par exemple, il n'y a pas de DÉT, il n'y a pas de CL en anglais, et en guarani (une langue parlée en Amérique du sud) il n'y a pas de COMP. Par contre, la majorité des langues possèdent des catégories lexicales majeures. On voit ici le pouvoir explicatif d'une théorie grammaticale restrictive qui inclut la théorie X-barre.

Cette approche diffère du point de vue structuraliste selon lequel les catégories sont définies par leur contexte d'occurrence, c'est-à-dire par rapport à l'occurrence d'autres catégories à l'intérieur des phrases. L'approche générativiste effectue un changement de perspective par rapport à la définition des catégories grammaticales. Elle permet le passage d'une approche empiriste à une approche mentaliste du V<sub>n</sub> des grammaires.

---

<sup>22</sup>Dans ce cadre, la catégorie Prép ne fait pas partie de l'ensemble des catégories lexicales majeures contrairement à d'autres analyses dans le cadre de la grammaire générative (p. ex. Jackendoff (1977) qui inclut les Prép. Ces dernières pourraient être définies par la matrice à la Chomsky [-V -N], ou par d'autres traits syntaxiques à la Jackendoff : [-subj. +obj. +comp.]

La théorie X-barre inclut en outre le concept de « type de catégorie » qui permet de limiter la classe des catégories syntaxiques possibles. Les types de catégories sont désignés par une variable X et par des barres. La variable peut être remplacée par n'importe quelle constante catégorielle : N, V, A... et les barres indiquent des types de catégories syntaxiques. Ainsi, X désigne la classe des catégories de type lexical, et X-barre, X-deux-barres, X-trois-barres, ... X<sup>max</sup> désignent des classes de catégories de type syntaxique. Cette notation permet de faire l'économie des catégories syntagmatiques : SN, SV, SA, etc.<sup>[23]</sup>, et d'uniformiser la notation des éléments du V<sub>n</sub> des grammaires. Elle a d'autres avantages, dont les suivants. La notation X-barre permet notamment d'exprimer le fait que les catégories syntaxiques sont des projections des catégories lexicales. Cette généralisation n'était pas mise en relief par les grammaires structuralistes, ni par la TS.

En outre, la notation X-barre permet de simplifier les grammaires. Elle permet de désigner, par très peu de symboles, des classes naturelles de catégories, c'est-à-dire des ensembles de catégories qui subissent les mêmes règles. Par exemple, dans la grammaire du français, les catégories V et Prép assignent des Cas aux SN qu'elles régissent. La notation X-barre désigne la classe naturelle des catégories V et Prép par le symbole [-N], ce qui simplifie la formulation de la règle d'assignation de Cas<sup>[24]</sup> :

(22) Une catégorie [-N] assigne un Cas au SN qu'elle régite.

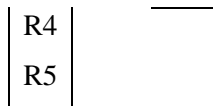
Les règles des grammaires peuvent ainsi exprimer des généralisations transcategorielles, c'est-à-dire des généralisations qui concernent plusieurs catégories. Ce résultat est appréciable du point de vue de la métrique d'évaluation des grammaires qui, de deux grammaires G<sub>1</sub> et G<sub>2</sub> décrivant les mêmes phénomènes, fera préférer celle qui contient le moins de règles ou de symboles. Dans la représentation abstraite de ces deux grammaires en (23), G<sub>2</sub> est plus simple que G<sub>1</sub>, elle est supérieure à G<sub>1</sub> du point de vue de l'adéquation explicative.

(23)

G11	R1	G22	R1
	R2		R2
	R3		R3

<sup>23</sup>Les catégories syntagmatiques sont néanmoins largement utilisées dans la TSÉ, et sont interprétées comme des catégories maximales (X<sup>max</sup>). Nous les utiliserons également en ce sens.

<sup>24</sup>La règle (22) est une formulation simplifiée de la règle d'assignation des Cas qui sera présentée plus loin.



Ainsi, la théorie X-barre restreint la classe des catégories grammaticales par une notation en traits et en types syntaxiques. Elle permet en outre de restreindre la classe des règles catégorielles et de combler certaines lacunes de la TS.

Soulignons tout d'abord que les règles catégorielles proposées dans la TS sont soumises à très peu de contraintes : les éléments qui les constituent appartiennent au  $V_n$  de la grammaire et un seul élément peut figurer à gauche de «  $\_>$  », comme dans le schéma suivant :

$$(24) \quad X \rightarrow \dots Y \dots \quad \text{où } X, Y \text{ appartiennent à } V_n.$$

D'après ce schéma, il est impossible de distinguer des règles telles que (25) des règles telles que (26) qui n'appartiennent à aucune grammaire. Ces dernières peuvent être postulées dans la TS, puisqu'aucun principe ne vient limiter la nature des catégories X et Y dans le schéma de règle (24)

$$(25) \quad \begin{aligned} P &\rightarrow SN SV \\ SV &\rightarrow V (SN) (SP) \\ SN &\rightarrow (DÉT) N (SP) \\ SP &\rightarrow Prép (SN) \end{aligned}$$

$$(26) \quad \begin{aligned} P &\rightarrow DÉT \\ SV &\rightarrow SN (DÉT) \\ SN &\rightarrow SV COMP \\ SP &\rightarrow (SA) N \end{aligned}$$

La théorie X-barre limite la forme générale des règles catégorielles pour les grammaires particulières et exclut des règles telles que (26) qui n'appartiennent à aucune grammaire.

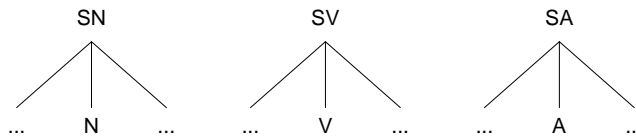
Une seconde lacune de la TS vient du fait que le schéma de règle (24) n'exprime pas les propriétés essentielles des constituants syntaxiques. En particulier, le fait que tout constituant est formé d'un élément lexical obligatoire de même nature syntaxique que le constituant qui le contient. On observe en (27) que le SV, le SN et le SA contiennent respectivement un V, un N et un A. En outre, d'autres constituants sont associés à l'élément lexical, soit des spécifieurs : des DÉT ou des DEG, soit des compléments : des SN ou des Sprép. Ces régularités formelles entre constituants de nature différente ne sont pas exprimées par le schéma (24).

- (27) a. Il [<sub>SV</sub> aime l'Opéra]  
 b. Son [<sub>SN</sub> amour de l'Opéra]  
 c. Il est [<sub>SA</sub> amoureux de l'Opéra]

La théorie X-barre inclut une notation qui décrit très simplement ces régularités et impose une forme symétrique aux règles catégorielles, de sorte que les parallélismes de structure observés sont alors la conséquence des principes de la théorie des grammaires. Les régularités formelles entre structures différentes se trouvent ainsi expliquées.

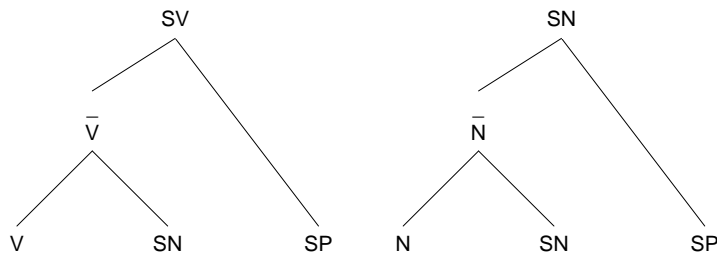
Une troisième lacune de la TS réside dans la définition des projections syntaxiques. Les règles catégorielles formulées dans le cadre de la TS engendrent des projections telles que (28) où les catégories lexicales : N, V, A sont immédiatement dominées par les catégories syntagmatiques : SN, SV, SA. Ces structures sont nettement insuffisantes pour décrire les propriétés des constituants syntaxiques.

(28)



Plusieurs faits, observables en anglais et dans d'autres langues configurationnelles, justifient l'existence de catégories syntaxiques intermédiaires entre les catégories lexicales et les catégories syntagmatiques ( $X^{\max}$ ). En particulier, les différences empiriques entre les compléments qui sous-catégorisent et les non sous-catégorisants justifient l'existence d'une catégorie intermédiaire entre les catégories lexicales et les catégories syntagmatiques, comme dans les représentations suivantes :

(29)



À cet effet, la notation X-barre inclut différents types de catégories syntaxiques (X, X-barre, ..., X<sup>max</sup>).

Mentionnons une dernière lacune des théories grammaticales non restrictives qui a conduit à la formulation de la théorie X-barre. Il s'agit de l'analyse des structures apparentées, comme celles des exemples (30). Les analyses proposées dans le cadre de la TSG (cf. Lakoff (1970, McCawley (1968)) et de la TS (cf. Lees (1960)) sur les nominalisations étaient très compliquées. Des transformations très élaborées, comme les règles (31) ont été postulées pour dériver les structures nominales de structures propositionnelles<sup>25</sup>].

- (30) a. John discovers a treasure.  
 (John découvre un trésor.)  
 b. John's discovery of a treasure.  
 (La découverte d'un trésor par John.)  
 c. He draws a picture.  
 (Il dessine un tableau.)  
 d. His drawing of a picture.  
 (Son dessin d'un tableau.)

(31) Nominaux d'action

$$\begin{array}{l}
 \text{(GT7)} \\
 \left. \begin{array}{l} X - T - N_a - T \\ \text{Nom} - \text{Tns} - [\text{V}_x + \text{V}_{in} + \text{Nom}'] - (\text{Adj} - \text{Ly}) - \text{Z} \end{array} \right\} \Rightarrow \\
 \text{P} \\
 X - \text{Nom} + \text{Gen} - (\text{Adj}) \left\{ \begin{array}{l} \text{Ing} \\ \text{Mml} \end{array} \right\} [\text{V}_x + \text{V}_{in} + \text{Nom}'] - \text{Z} - \text{Y} \\
 \text{P}
 \end{array}$$

$$\begin{array}{l}
 \text{(GT8)} \\
 \left. \begin{array}{l} X - T - N_a - T \\ \text{Nom} - \text{Tns} - \text{V}_t - \text{Nom}' - (\text{Adj} - \text{Ly}) - \text{Z} \end{array} \right\} \Rightarrow \\
 \left\{ \begin{array}{l} \text{Ing} \\ \text{Mml} \end{array} \right\} \\
 X - \text{Nom} + \text{Gen} - (\text{Adj}) \left\{ \begin{array}{l} \text{Ing} \\ \text{Mml} \end{array} \right\} \text{V}_t + \text{of} + \text{Nom}' - \text{Z} - \text{Y}
 \end{array}$$

Ainsi, l'hypothèse transformationnelle pour la dérivation des structures nominales requiert un nombre impressionnant de règles complexes. Toutefois, il ne pouvait en être

<sup>25</sup>Les règles GT7 et GT8 sont proposées dans Lees (1960). Ces règles dérivent des nominaux d'action tels que his drawing of a picture. Lees postule autant de transformations qu'il existe de types syntaxiques et sémantiques de nominaux. Ainsi, la GT7 dérive des nominaux d'action qui sont liés à des verbes intransitifs (V<sub>in</sub>), et la GT8 dérive des nominaux d'action liés à des verbes transitifs (V<sub>t</sub>). D'autres règles sont postulées pour dériver des nominaux agentifs ou infinitifs, etc.



autrement dans un modèle de grammaire comme celui de Lees ou de Lakoff, où les transformations sont le seul type de règles capable d'exprimer les similitudes de sélection et de sous-catégorisation entre les V et leurs dérivés N et A.

Pour pallier l'inélégance de l'hypothèse transformationnelle, Chomsky (1970) propose l'hypothèse lexicaliste qui limite la capacité descriptive des transformations au profit des règles catégorielles et du lexique. Selon l'hypothèse lexicaliste, les transformations ne peuvent pas modifier la morphologie des éléments lexicaux, de sorte que les nominaux ne peuvent être dérivés de structures propositionnelles par transformations. La parenté des V et de leurs dérivés N et A est traitée à l'intérieur du lexique, où les mêmes traits de sous-catégorisation et de sélection leurs sont associés. Les parallélismes de structures sont exprimées par une extension des règles catégorielles qui engendrent directement les structures nominales et adjectivales, au même titre que les structures verbales. Ces parallélismes sont très nets en anglais : dans les exemples (30) on observe que des constituants de même nature, soit des SN, figurent à gauche et à droite des N et des V. Les structures (30a) et (30b) sont parallèles, mis à part les formes of et 's qui figurent uniquement dans les structures nominales [<sup>26</sup>]. Ceci permet de postuler des règles catégorielles symétriques, où des constituants de même nature figurent à gauche et à droite des V et des N [<sup>27</sup>].

(32) V-deux-barres → N-deux-barres – V-barre      V-barre → V – N-deux-barres  
 N-deux-barres → N-deux-barres – N-barre      N-barre → N – N-deux-barres

### 3.7 Formulations

Plusieurs formulations de la théorie X-barre ont été proposées dans la TSÉ. Nous présentons brièvement les propositions de Chomsky (1970, 1981) et de Jackendoff (1977).

La première formulation de cette théorie est présentée dans Chomsky (1970). Soit les schémas de règles (33) qui expriment, au niveau de la théorie des grammaires, les régularités transcatégorielles observées dans les structures V, N, A, et imposent une forme canonique aux règles catégorielles des grammaires particulières. La forme générale des structures engendrées par les métarègles est présentée en (34).

<sup>26</sup>Il est à noter que bien qu'il y ait parallélisme en français pour des structures équivalentes, il est moins net. Outre la présence obligatoire de la préposition de en (ii), c'est un pronom possessif ou un article, et non un SN plein comme Paul ou cet homme, qui peut figurer à gauche des N.

- (i) Paul aime la musique
- (ii) Son amour de la musique
- (iii) \*Paul amour de la musique

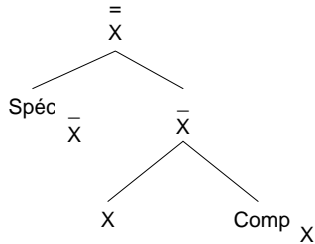
<sup>27</sup>Les formes of et 's qui appartiennent exclusivement aux structures nominales sont insérées dans ces structures par des transformations tardives.

(33) X-deux-barres  $\rightarrow$  Spéc<sub>X-barre</sub> – X-barre

X-barre  $\rightarrow$  X - Comp<sub>X</sub>

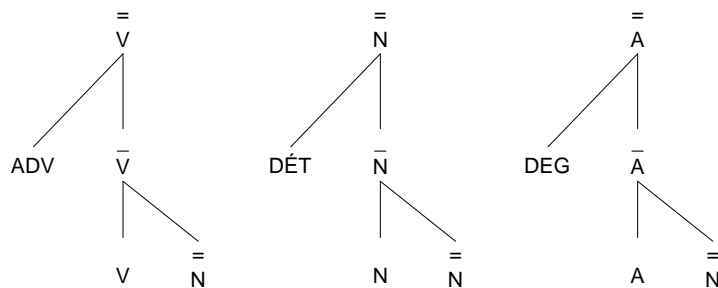
Où X : V, N, A.

(34)



Les métarègles (33) sont des abréviations de plusieurs règles, obtenues par la substitution des constantes catégorielles à la variable X. En outre, les catégories spécificateur (Spéc) et complément (Comp) désignent différents types de constituants. Ainsi, Spéc N-barre = DÉT, Spéc V-barre = AUX, Spéc A-barre = DEG, Comp X-barre = N-deux-barres, Prép-barre, P-barre. Les catégories Spéc et Comp figurent dans les métarègles (33) mais non dans les règles des grammaires particulières qui, soumises aux principes de la théorie X-barre, engendrent des structures telles que :

(35)



On trouve dans Chomsky (1983) une autre formulation de la théorie X-barre, soit le schéma de règle (36). Cette formulation a l'avantage d'être plus simple que la première. Les deux schémas de règles (33) sont réduits à un seul, où ne figure pas les catégories Spéc et Comp.

En outre, on remarque en (36) que le nombre de barres est représenté par une variable :  $\underline{n}$ . Ceci permet de rendre compte de la variation paramétrique entre les langues qui diffèrent du point

de vue de la valeur des projections maximales : en walpiri  $X^{\max}=1$ , en français  $X^{\max} = 2$ , et en anglais  $X^{\max} = 3$  selon l'analyse de Jackendoff (1977)<sup>[28]</sup>. En outre, le schéma (36) a l'avantage de ne pas imposer un ordre aux constituants, contrairement à la première formulation (18) où les têtes sont précédées de Spéc et suivies de Comp. Le paramètre de l'ordre des mots est alors fixé dans les grammaires particulières et le principe (36) de la théorie des grammaires peut garder toute sa généralité.

$$(36) \quad X^n \rightarrow \dots X^{n-1} \dots$$

Où  $1 \leq n \leq \max,$   
 $X : V, N, A, \dots : X^{\max}$

Le schéma de règle (36) stipule toutefois que toute catégorie syntaxique  $X^n$  est formée d'un élément obligatoire  $X^{n-1}$  qui partage les traits syntaxiques de  $X^n$  : cet élément est la tête de  $X^n$ . Cette généralisation s'applique à toutes les langues, qu'elles soient configurationnelles ou non, et quel que soit l'ordre des mots qui les caractérise.

La présence des têtes à l'intérieur des structures syntaxiques du français est déterminante pour leur bonne formation. À témoin, l'agrammaticalité des structures (37) où l'on note l'absence de certaines têtes.

- (37) a. \*[Le de Pierre aime la musique]  
 b. \*[Marie regarde vers la]  
 c. \*[Les rencontrent un]

En outre, la tête d'un constituant détermine l'accord grammatical. Les exemples (38) et (39) montrent que le verbe s'accorde avec la tête du SN sujet et non ses compléments.

- (38) a. Un animal à trois têtes est sorti de l'eau.  
 b. \*Un animal à trois têtes sont sortis de l'eau.  
 (39) c. Les frères de Pierre sont arrivés hier.  
 d. \*Les frères de Pierre est arrivé hier.

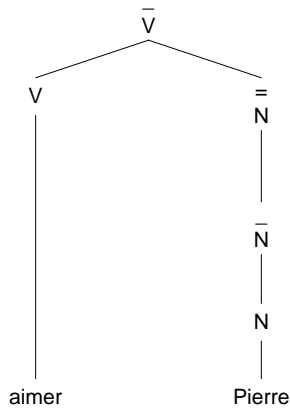
Ainsi, l'hypothèse selon laquelle les constituants syntaxiques ont une tête obligatoire a un contenu empirique en français. Des faits semblables dans d'autres langues viennent confirmer cette hypothèse.

---

<sup>28</sup>Jackendoff (1977) propose une autre formulation de la théorie X-barre : l'hypothèse des trois niveaux uniformes. Voir di Sciullo (1981a, b) pour une critique de cette hypothèse et une autre formulation de la théorie X-barre.

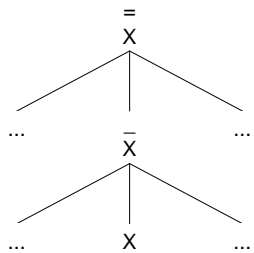
La théorie distingue les têtes lexicales des têtes syntaxiques. Les têtes lexicales sont des catégories de type X. Dans les structures (35) les V, N, A sont des têtes lexicales. Ces éléments dominant immédiatement des éléments terminaux, comme dans la structure suivante :

(40)



Lorsque les catégories de type X sont des têtes lexicales, les têtes syntaxiques sont des catégories de type  $X^n$ . Le schéma de règle (36), et les représentations qu'il engendre telle que (41), expriment la généralisation selon laquelle les catégories syntaxiques  $X^n$  sont des projections des catégories lexicales. La notion de projection syntaxique permet d'exprimer formellement le fait que les termes d'un même constituant ont des propriétés communes : ils partagent des traits syntaxiques.

(41)



### 3.8 Règles catégorielles $G_f$

Les principes de la théorie X-barre nous conduisent à formuler des règles telles que (42) pour la grammaire du français ( $G_f$ ). Chacune des règles contient une tête et des constituants

associés [<sup>29</sup>]. En outre, les règles qui développent des catégories V-barre, N-barre et A-barre sont symétriques : des compléments de même nature leur sont associés, soit N-deux-barres.

- (42) a. P-barre → COMP P  
 b. P → N-deux-barres INFL V-deux-barres  
 c. V-deux-barres → (ADV) V-barre  
 d. V-barre → V (N-deux-barres)  
 e. N-deux-barres → (DÉT) N-barre  
 f. N-barre → N (N-deux-barres)  
 g. A-deux-barres → (DEG) A-barre  
 h. A-barre → A (N-deux-barres)

La grammaire (42) n'est pas exhaustive, elle ne décrit qu'un sous-ensemble de structures-P du français, après l'application de la règle d'insertion lexicale. Elle servira néanmoins à illustrer la capacité descriptive et explicative des grammaires limitées par les principes de la théorie X-barre.

Nous avons mentionné au premier chapitre que l'ordre des mots est un paramètre qui varie selon les langues. En hindi par exemple, le COMP peut se trouver en début ou en fin de phrase, le sujet et l'objet précèdent le verbe, et les compléments précèdent généralement les noms et les verbes auxquels ils sont associés. Par contre, l'ordre des mots des phrases du français n'est pas celui qui s'observe en hindi et la  $G_f$  (42) permet de le décrire en partie [<sup>30</sup>].

La règle (42a) spécifie que les complémenteurs (COMP) se trouvent au début de phrase [<sup>31</sup>]. La règle (42b) indique que le sujet N-deux-barres précède le constituant verbal V-deux-barres. La catégorie inflexion (INFL) inclut le trait [ $\pm$  Temps] et permet de décrire les phrases à temps plein [+ Temps] et les phrases à temps non-fini [- Temps]. Les règles (42c) à (42h) expriment des généralisations sur la position des compléments et des spécificateurs par rapport aux têtes. Alors que les compléments suivent les têtes, les spécificateurs les précèdent. Cette

<sup>29</sup>Nous supposons, sans le discuter, que INFL est la tête de P, et que COMP est la tête de P-barre.

<sup>30</sup>Les règles de base décrivent un sous-ensemble de structure de phrases d'une langue donnée, les autres structures sont dérivées par transformation.

<sup>31</sup>La structure de COMP contient deux positions, l'une qui peut être remplie par un constituant [+qu] déplacé dans COMP, et une autre position [ $\pm$  qu], où [- qu] : {que, qui} est l'élément qui se trouve dans les relatives et les complétives, et [+qu] : {qui, si} est l'élément abstrait qui figure dans les questions directes et indirectes :

- (i) Je sais que Pierre est intelligent.  
 L'homme que tu connais.  
 L'homme qui est logicien.  
 (ii) Qui connaît Pierre ?  
 Je me demande qui connaît Pierre.  
 Je ne sais pas si Pierre est logicien.

grammaire décrit la structure de phrases telles que (43) et ne peut engendrer des suites telles que (44) qui sont malformées.

(43) a. z

## **Table des matières**

Remerciements

Introduction

Chapitre 1  
L'investigation linguistique

Chapitre 2  
Le lexique

Chapitre 3  
La base

Chapitre 4  
Les transformations

Chapitre 5  
La forme logique