

Deuxième partie

Syntaxe et sémantique
formelles

Chapitre 7

La grammaire générative : généralités

Objectifs de connaissance

- Situer la grammaire générative par rapport à la grammaire structurale.
- Expliciter les raisons du changement théorique opéré par la grammaire générative.
- Définir les buts de la théorie linguistique selon Chomsky.
- Définir les notions de grammaire générative et de description structurale.

Sous l'étiquette *grammaire générative* se regroupent depuis soixante ans un grand nombre de travaux descriptifs et théoriques relevant principalement de la syntaxe. À l'origine de ce paradigme scientifique, il y a un linguiste, Noam Chomsky (MIT, Cambridge, Massachusetts), dont les travaux ont à la fois permis un changement de cap dans l'histoire de la linguistique et un retour à la tradition rationaliste de Port-Royal.

Dans ce chapitre, nous allons définir les grandes lignes du modèle chomskien, tel qu'il a été formulé dans les premiers travaux de grammaire générative (nous renvoyons aux chapitres 8, 9 et 10 pour les développements récents). Nous définirons principalement les buts de la grammaire générative, et donnerons la définition de l'adjectif *génératif*. L'adjectif *transformationnel*, qui lui est généralement accolé, sera défini au chapitre 9. Dans un premier temps, nous rappellerons les principes de la grammaire structurale (cf. chapitre 6), ainsi que les principales critiques que lui adresse Chomsky.

La grammaire distributionnelle

La linguistique américaine, à l'époque des premiers travaux de Chomsky (1955-1957), se caractérise par les propriétés suivantes.

Tout d'abord, les grammaires, élaborées pour un grand nombre de langues, sont des **grammaires distributionnelles**. La **distribution** d'une unité est la somme de ses environnements, à savoir l'ensemble des positions dans lesquelles elle peut prendre place. Le linguiste établit ainsi des **classes distributionnelles** (ou paradigmes), qui contiennent l'ensemble des éléments pouvant apparaître dans cette position. La grammaire est donc une **grammaire de listes**.

Ainsi, en (1), on dira que les éléments appartenant à l'ensemble (d) appartiennent à la même classe distributionnelle, car ils peuvent tous apparaître dans la même position en (1a-c) :

- (1) a. le _ chante
 b. un _ joue
 c. le méchant _ est parti
 d. {gamin, garçon}

De même, les exemples (2) indiquent deux positions pour les pronoms clitiques (attachés au verbe en position préverbale) : les pronoms se classent ainsi en trois classes distributionnelles.

- (2) a. Jean {me, te, nous, vous} {le} donne.
 b. Jean {le} {lui, leur} donne.

Ensuite, le modèle d'analyse linguistique est un **modèle taxinomique**. Il s'agit, à partir d'une **procédure de découverte** (l'analyse en constituants immédiats), de classer l'ensemble des constructions d'une langue donnée. L'analyse en constituants immédiats est basée sur le principe suivant : chaque constituant est analysé en constituants de rang immédiatement inférieur, le découpage se poursuivant jusqu'à ce que l'analyse atteigne les catégories lexicales (cf. chapitre 6).

Enfin, la conception taxinomique de la grammaire donne un rôle fondamental au **corpus**. Le linguiste doit dans un premier temps recueillir des données, et ensuite les traiter à l'aide des procédures de l'analyse distributionnelle et de l'analyse en constituants immédiats. L'approche est donc inductive : on part des faits pour induire des généralisations. Par conséquent, la conception de la grammaire est celle d'une **procédure de découverte** : à partir de données, la théorie produit la grammaire. Pour toute langue, il y a donc une et une seule grammaire, et c'est la théorie qui, à partir des données, fournit la grammaire.

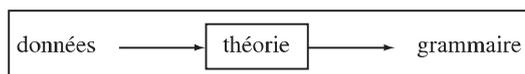


Figure 1. Procédure de découverte

La critique du structuralisme par Chomsky

Chomsky adresse deux critiques principales à la conception taxinomiste de la grammaire.

Première critique : le **corpus** ne contient qu'un **sous-ensemble des phrases grammaticales d'une langue**, de même qu'il contient également un sous-ensemble de phrases agrammaticales.

Deuxième critique : pour Chomsky, une théorie ne peut pas être une **procédure de découverte**. Au mieux, elle est une **procédure de décision** : elle permet de dire si une grammaire est ou non une grammaire à partir des données fournies. Au pire, elle n'est qu'une **procédure d'évaluation** : elle dit quelle est la meilleure grammaire parmi un ensemble de grammaires possibles.

À l'opposé d'une conception taxinomique, Chomsky défend une conception théorique de la grammaire. Il s'agit pour lui d'envisager la construction de théories générales (ce qu'il appelle des modèles hypothétiques), qui doivent :

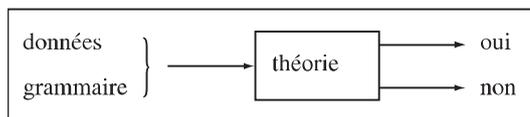


Figure 2. Procédure de décision

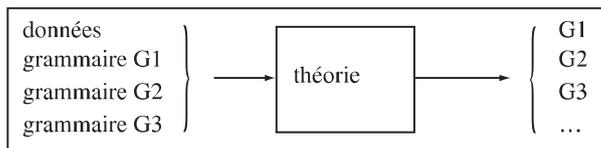


Figure 3. Procédure d'évaluation

- *décrire* les faits connus (satisfaire l'adéquation descriptive),
- *expliquer* ces faits connus (satisfaire l'adéquation explicative) et
- *prédire* de nouveaux faits.

De plus, Chomsky adopte un principe méthodologique courant en science, selon lequel on ne peut tirer que des généralisations négatives des observations ou des expériences. En d'autres termes, on ne pourra jamais dire d'une théorie qu'elle est vraie, mais on pourra essayer de montrer qu'elle est fautive. La valeur d'une théorie ne réside donc pas dans le fait qu'elle est vraie, mais dans le fait qu'elle est réfutable (falsifiable). Elle a donc les attributs suivants : valeur explicative, cohérence interne, compatibilité avec d'autres théories, économie, simplicité et enfin élégance. Pour arriver à ses fins, Chomsky propose la construction de **modèles explicites et précis**. Ces modèles sont les **grammaires génératives**.

Buts de la théorie linguistique

Pour Chomsky, le but de la théorie linguistique est de décrire la **faculté de langage**, à savoir la capacité des sujets parlants à distinguer les phrases grammaticales des phrases agrammaticales. Cette capacité de langage, qui est à l'origine de l'acquisition du langage, fait partie de la compétence des sujets parlants.

Compétence et performance

Chomsky distingue la *compétence* de la *performance*. Si la performance est l'utilisation de la compétence, i.e. la compétence mise en œuvre dans des actes de parole par les sujets parlants, la compétence renvoie au système de règles sous-jacent à l'utilisation et à la compréhension du langage. Plus précisément, la compétence définit un système internalisé de règles (la grammaire) associant des sons à des sens, ou des séquences de signaux acoustiques à des interprétations sémantiques.

Pour Chomsky, la compétence est un héritage biologique, quelque chose d'inné par opposition à quelque chose d'acquis. De plus, la description de la compétence doit indiquer non pas simplement le système de règles propres à chaque langue, mais les universaux du langage (les propriétés universelles communes à toutes les langues). Le but de la grammaire est ainsi d'approcher la description de la **grammaire universelle** (cf. chapitre 8).

Priorité de la compétence sur la performance

Pour Chomsky, l'étude de la compétence est prioritaire par rapport à l'étude de la performance. Quels sont les arguments qui lui permettent de défendre cette thèse ? Tout sujet parlant français, indépendamment du contexte d'emploi, peut exprimer des jugements induits par sa compétence linguistique (Ruwet 1967) :

- (1) est une phrase grammaticale et univoque ;
- (2) est une phrase grammaticale, mais ambiguë ;
- (3) est une phrase grammaticale, mais ininterprétable ;
- (4) est une phrase agrammaticale, mais interprétable.

- (1) Le jeune homme rencontre la vieille dame.
- (2) Pierre aime mieux Paul que Jean.
- (3) Le silence vertébral indispose le voile licite.
- (4) Vous faire moi rigoler.

La récurrence de ces jugements (grammaticalité, ambiguïté, interprétabilité) autorise le linguiste à utiliser son **intuition linguistique** comme support de jugement, plutôt que de recourir à des faits de performance (dépendants de corpus).

Autonomie de la syntaxe

Pour Chomsky, l'approche formelle du langage (l'étude de la syntaxe) est autonome et indépendante par rapport à l'étude de la sémantique. Pour le montrer, il faut distinguer phrases grammaticales et phrases agrammaticales, phrases grammaticales et phrases interprétables, phrases grammaticales et phrases correctes, et enfin phrases grammaticales et phrases fréquentes (Ruwet 1967).

Une phrase agrammaticale n'est pas une phrase **ininterprétable**. Ainsi, dans les exemples (5)-(8), toutes les phrases (b) sont agrammaticales, mais parfaitement interprétables. Les phrases (8) sont intéressantes, car (8a), qui est ambiguë et peut s'interpréter soit comme (8b) soit comme (8c), est la seule réalisation syntaxique possible :

- (5) a. Jean mange une pomme.
- b. * Pomme une mange Jean.
- (6) a. Je n'ai vu personne.
- b. * Je n'ai personne vu.
- (7) a. Que fait Pierre ?
- b. * Quoi Pierre fait-il ?
- (8) a. Pierre m'a proposé de venir.
- b. * Pierre m'a proposé que je vienne.
- c. * Pierre m'a proposé qu'il vienne.

La grammaticalité ne dépend pas de l'**interprétabilité**. Si (9a) est grammaticale, mais ininterprétable (ou en tout cas difficilement interprétable), (9b) est à la fois agrammaticale et ininterprétable.

- (9) a. D'incolores idées vertes dorment furieusement.
- b. Furieusement, dormir idée vert incolore.

La grammaticalité ne recouvre pas la **correction grammaticale**. La phrase (10b) est incorrecte et s'oppose à (10a) sur des critères de niveau de langue. Par contre, les phrases (10c) et (10d) sont agrammaticales, pour des raisons de placement de la particule négative (forclusif) *rien* :

- (10) a. Je n'ai rien vu.

- b. ? J'ai rien vu.
- c. * Je n'ai vu rien.
- d. * J'ai vu rien.

Enfin, la grammaticalité est indépendante de la **fréquence** d'emploi. Dans le contexte *le _ est fragile, gorille* est aussi improbable du point de vue de sa fréquence d'emploi que *de*. Mais les deux phrases (11a) et (11b) n'ont pas le même statut grammatical, seule (11a) étant agrammaticale :

- (11) a. * Le de est fragile.
- b. Le gorille est fragile.

Ainsi, on peut affirmer que la syntaxe, i.e. l'ensemble des principes déterminant la grammaticalité des phrases, est autonome et prioritaire relativement à la sémantique, i.e. l'ensemble des principes déterminant l'interprétation des phrases.

Buts de la théorie linguistique

La capacité de distinguer entre phrases grammaticales et phrases agrammaticales fait partie de la compétence linguistique des sujets parlants. La grammaire comme modèle de la compétence a donc pour première tâche d'énumérer explicitement toutes les phrases qui sont grammaticales (bien formées) et d'exclure toutes les séquences qui sont agrammaticales. On peut dès lors donner la définition d'une grammaire générative :

Une **grammaire générative** est une grammaire explicite, capable d'énumérer toutes et rien que les phrases grammaticales d'une langue.

On notera qu'une grammaire générative n'est pas synonyme de grammaire de production : *générer* signifie *engendrer*, à savoir énumérer explicitement au moyen de règles.

Grammaire générative et grammaire traditionnelle

La critique principale de la grammaire générative à la grammaire traditionnelle tient au fait que la grammaire traditionnelle n'est pas une grammaire explicite (cf. chapitre 1). Elle consiste en la formulation de règles générales, illustrées par des exemples, suivis de listes d'exceptions. Les règles ne sont donc pas des ensembles explicites d'instructions : au contraire, elles permettent d'engendrer des séquences agrammaticales (cf. chapitre 1).

Par contre, en tant qu'ensemble d'instructions explicites applicables mécaniquement (algorithmes), la **grammaire** doit être capable :

- d'**engendrer un ensemble infini de phrases grammaticales** à partir d'un ensemble fini d'éléments (catégories, unités lexicales, règles) ;
- de leur associer automatiquement une **description structurale**.

Les aspects de la description structurale

La description structurale d'une phrase fournit un ensemble de renseignements qui déterminent d'une part la représentation phonétique et d'autre part la représentation sémantique de la phrase. Ruwet (1967) relève les aspects suivants de la description structurale.

• Structure syntaxique et forme de surface

Certaines phrases sont différentes du point de vue superficiel alors qu'elles ont une même structure syntaxique. La grammaire doit donc leur assigner un niveau de représentation structurale identique. Ainsi, les phrases (12a) et (12b) ont, à un certain degré de généralité, une même structure syntaxique (construction *SN-V-SN*, cf. les structures (12')) :

- (12) a. Pierre aime Marie.
 b. Le petit vieillard alerte qui habite en face de chez nous a perdu les lunettes qu'il avait achetées hier soir.
- (12') a. [_S [_{SN} Pierre] [_{SV} [_V aime] [_{SN} Marie]]]
 b. [_S [_{SN} le petit vieillard alerte [_S qui habite en face de chez nous]] [_{SV} [_V a perdu [_{SN} les lunettes [_S qu'il avait achetées hier soir]]]]]

• Structure profonde et structure de surface

Certaines phrases sont semblables du point de vue superficiel, mais ont des structures profondes différentes. Ainsi, (13a) et (14a) se ressemblent du point de vue superficiel (un seul mot a été modifié, *tailleur vs fils*), mais ont des structures profondes différentes, (13b) et (14b) :

- (13) a. J'ai fait faire un veston à mon tailleur.
 b. [j'ai fait [mon tailleur faire un veston]]
- (14) a. J'ai fait faire un veston à mon fils.
 b. [j'ai fait [x faire un veston à mon fils]]

Il est donc fondamental de distinguer le **structure profonde**, définie comme le lieu de représentation de la description structurale et produite par les règles de la syntaxe et l'information lexicale, de la **structure de surface**, définie comme l'entrée de la représentation phonétique, et produite par des **règles de transformation** appliquées à la structure profonde (cf. chapitre 9).

• Phrases ambiguës

Le troisième aspect de la description structurale est lié aux ambiguïtés. Une phrase ambiguë est une phrase dont la structure de surface est le produit de deux (au moins) structures profondes. Ainsi, la phrase (15a) reçoit les deux lectures (15b) et (15c), qui correspondent à deux structures profondes simplifiées :

- (15) a. J'ai lu la critique de Chomsky.
 b. [j'ai lu [x a critiqué Chomsky]]
 c. [j'ai lu [Chomsky a critiqué x]]

On notera qu'une manière de désambigüiser (15) consiste à lui ajouter un syntagme prépositionnel, dont le choix n'est pas libre :

- (16) a. J'ai lu la critique de Chomsky du livre de Postal.
 b. J'ai lu la critique de Chomsky par Postal.

La description structurale doit donc être capable de fournir, pour toute phrase ambiguë syntaxiquement, une **structure profonde** correspondant à chacune de ses interprétations sémantiques.

• Fonctions grammaticales

La description structurale doit aussi expliquer les relations ou fonctions grammaticales entre les éléments des phrases (cf. chapitres 6 et 8), bien que les fonctions grammaticales

ne soient pas des primitives linguistiques. Par exemple, *Pierre* et *Marie* ont les mêmes relations avec le verbe *aimer* en (17a) et (17b), mais pas en (17a) et (17c) :

- (17) a. Pierre aime Marie.
 b. Marie est aimée de Pierre.
 c. Marie aime Pierre.

De même, *Pierre* ne joue pas le même rôle en (18) (il n'a pas la même relation avec *vivre* et *comprendre*), comme le montrent les structures profondes (19) :

- (18) a. Pierre est difficile à vivre.
 b. Pierre est difficile à comprendre.
 (19) a. [il est difficile [x vivre avec Pierre]]
 b. [il est difficile [x comprendre Pierre]]

• *Sous-catégorisation et relations de sélection*

La description structurale doit aussi permettre de comprendre la manière dont les différentes langues naturelles **sous-catégorisent** leurs unités lexicales (cf. chapitre 6). Une sous-catégorie est le sous-type d'une catégorie. Ainsi, les **noms** du français se classent en différentes sous-catégories :

- nom masculin vs féminin ;
- nom animé vs non animé ;
- nom comptable vs massif.

Les **verbes** sont sous-catégorisés en :

- verbe intransitif vs transitif ;
- verbe transitif direct vs indirect ;
- verbe transitif avec vs sans suppression de l'objet.

Le rôle des **sous-catégorisations** apparaît dans les **relations de sélection**. On définira les relations de sélection de la manière suivante (Ruwet 1967) :

Relation de sélection

Un élément de catégorie *A* (verbe), qui a dans son environnement un élément de catégorie *B* (nom), requiert, s'il appartient à une sous-catégorie *a*, que *B* appartienne à une sous-catégorie *b*.

Par exemple, les sous-catégories de verbes données en (20) imposent des relations de **sélection** ou **restrictions sélectionnelles**, d'une part sur leur sujet (verbes du groupe (20a)), d'autre part sur leur objet (verbes du groupe (20b)), ce que montrent les phrases (b), agrammaticales, de (21)-(23) :

- (20) a. { admirer, redouter, ... } sujet [+ animé]
 b. { effrayer, intriguer, étonner, ... } objet [+ animé]
 (21) a. Pierre admire la sincérité.
 b. * La sincérité admire Pierre.
 (22) a. Les soldats redoutent le danger.
 b. * Le danger redoute les soldats.
 (23) a. La sincérité effraie Pierre.
 b. * Pierre effraie la sincérité.

• *La distinction animé/inanimé*

Enfin, la description structurale doit tenir compte de la distinction animé/inanimé dans la syntaxe, afin de rendre compte du choix des morphèmes interrogatifs d'une part, et des pronoms de troisième personne d'autre part :

- (24) a. Qui vois-tu ? {Pierre, *un livre}
 b. Que vois-tu ? {*Pierre, un livre}
- (25) a. Je pense à lui lui = {à Pierre, *à ce travail}
 b. J'y pense y = {à ce travail, * à Pierre}

Ainsi, une grammaire générative ne doit pas simplement produire l'ensemble des phrases grammaticales d'une langue à l'aide de règles syntaxiques explicites. Elle doit de plus en fournir une **description structurale**, qui conditionne à la fois leur **interprétation phonétique** et leur **interprétation sémantique**.

Lectures conseillées

On se reportera prioritairement à l'œuvre fondatrice de Chomsky publiée en 1957 (cf. Chomsky 1969a pour la traduction française), qui est un condensé de sa thèse de 1955, publiée en 1975 (cf. Chomsky 1975a). Ruwet (1967) reste une excellente introduction toujours à la grammaire générative, et a le grand avantage de situer la grammaire générative par rapport à la linguistique structurale. Pour les relations entre la grammaire générative et Port-Royal, voir Chomsky (1969b). Pour une définition de la faculté de langage et de la grammaire universelle, voir Chomsky (1977a) et Chomsky (1986). Pour une présentation complète du modèle standard, on renvoie à Chomsky (1971).

À retenir

- La méthode scientifique de la grammaire générative est hypothético-déductive et s'oppose à la méthode inductive de la grammaire structurale.
- Le but de la grammaire générative est de décrire la compétence des sujets parlants et d'approcher la description de la faculté de langage propre à l'espèce humaine.
- La syntaxe est définie comme prioritaire et autonome par rapport à la sémantique.
- Une grammaire est générative si elle est capable, à partir d'un ensemble explicite de règles, de générer l'ensemble des phrases grammaticales d'une langue donnée et de leur assigner une description structurale.