

## Chapitre 1

### Les notions de base de l'analyse syntaxique

#### 1. Morphologie et syntaxe, mot et phrase

##### 1.1. Morphologie et syntaxe

La syntaxe, au sens couramment donné à ce terme en linguistique, étudie les régularités que manifestent les combinaisons d'unités dont le rang est compris entre celui du mot et celui de la phrase. La morphologie quant à elle étudie la structure interne des mots.

##### 1.2. La notion de mot : lexème et mot-forme

Le terme de mot recouvre deux notions qu'il convient de distinguer. Nous ne nous référons pas au même sens de 'mot' lorsque nous disons par exemple que le mot *dormez* comprend deux syllabes, ou que *dormez* et *dort* sont deux formes du même mot. Pour éviter cette ambiguïté, on peut utiliser *lexème* pour se référer spécifiquement à ce qui est commun à un ensemble de mots regroupés sous une même entrée lexicale<sup>1</sup> ; on peut de même parler de *mot-forme* pour préciser que c'est à la première valeur qu'on se réfère, c'est-à-dire au mot en tant que séquence de phonèmes ou de lettres. On désigne fréquemment comme *paradigme* l'ensemble des mots-formes considérés comme appartenant à un même lexème<sup>2</sup>.

Le *sens lexical* d'un mot est ce qui est sémantiquement commun à l'ensemble des mots-formes qui constituent le paradigme auquel il appartient. Dans la mesure où le terme de lexical est fortement polysémique, et s'oppose dans certaines de ses acceptions à grammatical (cf. note 1), il importe de souligner que selon cette définition, il n'y a aucune contradiction à parler de sens lexical d'un mot grammatical. Par exemple, le pronom possessif *le mien* est un mot grammatical ; il fait partie d'un paradigme dont les autres éléments sont *la mienne*, *les miens* et *les miennes*, et a donc comme sens lexical 'appartenance à la sphère personnelle de l'énonciateur' ; pour chacun des quatre mots-formes qui constituent le paradigme, à ce sens lexical s'ajoutent une valeur de nombre (singulier / pluriel) et de genre (masculin / féminin).

Le mot (au sens de mot-forme) peut coïncider avec l'unité significative élémentaire (par exemple, la préposition *avec* ne peut pas se décomposer en unités significatives plus petites), mais ce n'est généralement pas le cas. À partir du moment où un paradigme regroupe au moins deux mots, les mots de ce paradigme sont normalement des formes complexes, et même un mot qui constitue à lui seul un paradigme peut être morphologiquement complexe, si sa forme révèle une relation systématique de dérivation avec un autre lexème.

Les combinaisons d'unités significatives reconnues comme mots se caractérisent par un fort degré de cohésion interne et par une relative autonomie par rapport aux autres mots, ce qui pousse à faire l'hypothèse que les régularités concernant les combinaisons d'unités

---

<sup>1</sup> Ceci n'est malheureusement pas la seule acception avec laquelle on peut rencontrer le terme de lexème :

–chez certains auteurs, lexème est synonyme de mot plein, et la notion de lexème exclut les mots grammaticaux (cf. ci-dessous, section 3.3), alors que selon la définition retenue ici elle les englobe ;

–chez d'autres, le terme de lexème s'applique à des unités significatives minimales (ou morphèmes lexicaux, par opposition aux morphèmes grammaticaux).

<sup>2</sup> La question de savoir ce qui peut justifier de regrouper les mots-formes en paradigmes est dans l'immédiat laissée de côté ; elle sera reprise à la section 2.2.

significatives de rang inférieur au mot ne sont pas de même nature (et donc ne relèvent pas des mêmes techniques d'analyse et de description) que celles concernant les combinaisons de mots en unités de rang supérieur.

Afin de pouvoir avancer dans la mise en place des notions fondamentales, nous allons supposer provisoirement résolue la question du découpage de la phrase en mots. Nous reviendrons à la section 6 sur les difficultés que peut soulever ce découpage, du fait de l'existence de formes dont le comportement est tel qu'il est difficile de préciser leur statut du point de vue du découpage de la phrase en mots ; nous évoquerons aussi à cette occasion le fait que le découpage en mots tel que le codifie l'orthographe des langues n'est pas toujours celui auquel on aboutirait en appliquant systématiquement les critères permettant de reconnaître dans une phrase les frontières de mots.

### 1.3. La notion de phrase

En dépit de la définition traditionnelle qui caractérise la phrase comme 'l'expression d'un sens complet', il n'est pas possible de définir la phrase de façon satisfaisante comme unité significative maximale (c'est-à-dire en termes de complétude sémantique ou d'autonomie sémantique). En effet, selon une telle définition, seuls seraient réellement reconnaissables comme phrases les énoncés exprimant des vérités universellement valables, dans laquelle les noms sont pris en valeur générique (*L'homme est mortel, le chat est un mammifère*, etc.). Sémantiquement parlant, les phrases ordinaires sont toujours plus ou moins dépendantes du contexte. C'est en fait un autre type de complétude qui caractérise la phrase.

Si on ne veut pas se contenter d'une approche intuitive, on doit s'appuyer sur les notions de *contenu propositionnel* et d'*opération énonciative*. Un contenu propositionnel (terme qui renvoie à la notion de proposition telle que l'a élaborée la tradition logique), est la représentation d'un état possible du monde (événement, situation) au moyen des lexèmes d'une langue. Quant à la notion d'opération énonciative, elle englobe notamment les notions sur lesquelles se base le classement traditionnel des phrases en déclaratives (ou assertives), interrogatives et impératives :

- les phrases déclaratives affirment l'adéquation (assertion positive) ou l'inadéquation (assertion négative) d'un contenu propositionnel à la description d'une situation de référence ;
- les phrases interrogatives sont de deux types : ou bien l'allocutaire doit assumer l'assertion d'un contenu propositionnel que lui fournit l'énonciateur (questions appelant une réponse par oui ou par non), ou bien l'énonciateur fournit un contenu propositionnel incomplet que l'allocutaire doit compléter (questions à proformes interrogatives) ;
- les phrases injonctives signifient que l'énonciateur met en demeure l'allocutaire de faire en sorte que la réalité devienne conforme à un certain contenu propositionnel.

Le propre du langage est de fournir aux locuteurs la possibilité de construire une variété illimitée de contenus propositionnels et d'explicitier les opérations qu'ils effectuent en manipulant ces contenus propositionnels dans l'interaction communicative, et la phrase est le cadre dans lequel se manifestent systématiquement ces possibilités. Par exemple, les phrases des ex. (1) à (3) ainsi que les groupes de mots entre crochets dans l'ex. (4) renvoient à la même représentation d'un événement conceptualisé comme *réparer* mettant en jeu deux entités désignées respectivement comme *Jean* et *la voiture*, et les variations d'un exemple à l'autre expriment de manière systématique différentes possibilités de manipuler ce contenu propositionnel dans un acte d'énonciation :

- les phrases de l'ex. (1) sont des unités phrastiques indépendantes de type assertif, positif –(1a)– et négatif –(1b) ;

- (1) a. Jean a réparé la voiture

b. Jean n'a pas réparé la voiture

–les phrases de l'ex. (2) sont des unités phrastiques indépendantes de type injonctif –(2a)– et interrogatif –(2b-d) ;

(2) a. Que Jean répare la voiture !

b. Qui est-ce qui a réparé la voiture ?

c. Qu'est-ce que Jean a réparé ?

d. Est-ce que Jean a réparé la voiture ?

–les phrases de l'ex. (3) sont des unités phrastiques qui expriment différentes façons de présenter le même contenu informatif en fonction du contexte discursif ;

(3) a. C'est Jean qui a réparé la voiture (... moi, j'en aurais été incapable)

b. C'est la voiture que Jean a réparé (... pas la moto)

–enfin, les groupes de mots entre crochets dans l'ex. (4) sont des unités phrastiques enchâssées à l'intérieur d'autres unités phrastiques, et leur statut énonciatif dépend de la façon dont elles participent à la construction de l'unité phrastique dans laquelle elles sont enchâssées.

(4) a. Je crois [que Jean a réparé la voiture]

b. Je ne sais pas [si Jean a réparé la voiture]

c. J'ai demandé à Jean [de réparer la voiture]

d. Jean a promis [de réparer la voiture]

e. Tu peux partir avec la voiture [que Jean a réparée]

La phrase est donc une combinaison de mots dont la structuration permet l'expression systématique de l'élaboration énonciative d'un contenu propositionnel ; autrement dit, le propre d'une unité phrastique est de participer à un jeu de correspondances régulières avec d'autres unités phrastiques qui expriment une élaboration énonciative différente d'un même contenu propositionnel<sup>3</sup>.

---

<sup>3</sup> La grammaire transformationnelle des années 60-70 a eu le mérite de mettre pour la première fois fortement l'accent sur cet aspect de l'organisation de la langue, et sur la nécessité d'en donner une description aussi précise et exhaustive que possible. Cette option a largement conditionné le développement ultérieur des recherches en syntaxe, indépendamment des positions que peuvent avoir les différentes théories contemporaines sur les présupposés théoriques de la grammaire transformationnelle et sur le type de formalisation qui est le mieux à même de décrire les phénomènes dont la grammaire transformationnelle a proposé une première systématisation.

## 2. La structure interne des mots et ses relations avec la syntaxe

### 2.1. La notion de morphème et ses limites

La notion de structure morphologique des mots découle du fait qu'on peut trouver entre les mots des ressemblances partielles de forme régulièrement associées à des ressemblances partielles de signification. Ce sont d'ailleurs ces relations qui permettent de considérer que deux mots-formes appartiennent au même lexème, ou que deux lexèmes appartiennent à la même famille. Dans les cas simples, comme celui de l'ex. (5), on peut rendre compte de ces relations en découpant les mots-formes en segments qui véhiculent chacun une partie de la signification.

(5)	aim-er	aim-ons	—	aim-able	—
	trouv-er	trouv-ons	—	trouv-able	in-trouv-able
	lav-er	lav-ons	lav-eur	lav-able	—
	cass-er	cass-ons	cass-eur	cass-able	in-cass-able

A partir de là, il est tentant de poser comme principe général que, de même que la phrase est une chaîne de mots, le mot est une chaîne de segments significatifs minimaux, ou *morphèmes*, et le morphème est à son tour une chaîne de phonèmes (ou de lettres).

Il n'y a toutefois rien d'exceptionnel à ce que les relations régulières entre mots prennent des formes qui se prêtent mal à une telle analyse, et il est préférable de reconnaître la possibilité d'opérations morphologiques autres que la simple adjonction d'affixes à une base.

En particulier, dans une morphologie parfaitement concaténative, on peut identifier le lexème à la base à laquelle s'attachent les affixes flexionnels. Dans cette optique, il serait cohérent de désigner par exemple comme 'le lexème verbal *trouv-*' ce qu'on désigne plus communément comme 'le verbe *trouver*'. Une telle pratique, qui aurait l'avantage de réduire les risques de confusion entre mot en tant que lexème et mot-forme, est malheureusement difficilement généralisable aux mots (comme par exemple en français le verbe *savoir*) dont la flexion met en jeu des opérations plus complexes que la simple adjonction d'affixes à une base qui reste constante tout au long de la flexion. C'est ce qui justifie de s'en tenir à la pratique traditionnelle qui consiste à désigner les lexèmes variables (c'est-à-dire dont le paradigme comporte au moins deux formes) au moyen d'une de leur formes conventionnellement choisie (par exemple l'infinitif pour les verbes<sup>4</sup>, le masculin singulier pour les adjectifs, etc.).

Dans une conception 'lexicaliste' de la syntaxe, qui pose une séparation stricte entre les règles de formation des mots et les règles de concaténation des mots en phrases, il est inutile d'approfondir plus cette question, car il n'y a là aucun problème particulier : dans cette conception, les règles de syntaxe tiennent bien sûr compte de caractéristiques des mots qui leur sont conférées par leur structure interne, mais la façon précise dont ces caractéristiques se manifestent dans la structure du mot n'est pas pertinente pour la syntaxe (par exemple, il peut être pertinent pour une règle de syntaxe qu'un nom soit à la forme du singulier ou à la forme du pluriel, mais une règle de syntaxe n'a pas à prendre en considération le fait que le pluriel d'un nom soit formé par préfixation, suffixation, redoublement, antériorisation d'une voyelle,

---

<sup>4</sup> Remarquons au passage que la convention d'utiliser la forme d'infinitif pour se référer aux lexèmes verbaux est loin d'être universelle. Selon les traditions, d'autres formes sont utilisées (par exemple en hongrois, la troisième personne du singulier du présent), et d'ailleurs la convention d'utiliser la forme d'infinitif pour se référer aux lexèmes verbaux n'est pas généralisable pour la simple raison que l'infinitif en tant que type particulier de forme grammaticale n'existe pas dans toutes les langues.

changement tonal, etc.). C'est d'ailleurs pourquoi, dans les exemples qui illustrent ce cours, les mots ne sont pas systématiquement segmentés en morphèmes, même dans les cas où cela ne poserait aucun problème majeur. L'important est que les gloses fassent clairement apparaître les traits syntaxiquement pertinents (comme par exemple *singulier*, *première personne du pluriel*) que possède chaque mot du fait de sa structure interne.

## 2.2. Dérivation et flexion

La morphologie est traditionnellement subdivisée en *morphologie dérivationnelle*, qui étudie la relation entre lexèmes qui peuvent être considérés comme appartenant à une même 'famille de mots', et *morphologie flexionnelle*, qui étudie la relation entre les différentes formes d'un même lexème. Cette définition oblige à poser la question de la justification théorique du regroupement des mots-formes en paradigmes, et des critères utilisables pour établir ce regroupement.

Dans la pratique, il n'est pas difficile de reconnaître comme relevant de la dérivation les processus morphologiques qui changent la catégorie syntaxique des mots (*chanter* → *chanteur*), par contre la question des critères permettant de caractériser comme flexionnels ou dérivationnels les processus morphologiques qui ne changent pas la catégorie syntaxique des mots auxquels ils s'appliquent est une question délicate. De tous les critères qui ont pu être proposés, celui qui semble le plus intéressant est celui qui fait référence à l'interaction entre la structure morphologique du mot et son comportement syntaxique. Selon ce critère, seules les caractéristiques flexionnelles du mot peuvent intervenir dans la formulation de règles syntaxiques, et il ne doit pas exister de règle syntaxique spécifique à tel ou tel type de mot dérivé. Autrement dit, la dérivation peut modifier les propriétés syntaxiques du mot, mais le comportement syntaxique d'un mot dérivé doit pouvoir se prédire en le rattachant simplement à une catégorie à laquelle appartiennent aussi des mots non dérivés. Par contre, une règle de syntaxe peut comporter des conditions nécessaires sur les caractéristiques flexionnelles des mots auxquels elle s'applique.

Par exemple, selon ce critère, l'infinitif du français est une formation flexionnelle, car les propriétés syntaxiques des infinitifs ne se retrouvent dans aucun mot qui ne présente pas la terminaison caractéristique des infinitifs ; par contre, le nom d'action est une formation dérivationnelle, car les noms d'action ont les mêmes propriétés syntaxiques que les noms qui ne sont pas dérivés de verbe. Le temps verbal est une catégorie flexionnelle, car aucun mot dépourvu de variation en temps ne peut se substituer au verbe conjugué. Le pluriel des noms est aussi en français une catégorie flexionnelle, du fait de l'existence de règles d'accord en nombre ; par contre, dans une langue qui aurait une formation morphologique de pluriel des noms ne donnant lieu à aucun phénomène d'accord, on devrait envisager la possibilité de reconnaître le pluriel comme une catégorie dérivationnelle.

## 3. Classes de mots

### 3.1. Critères possibles de classement

Une répartition des mots en classes est nécessaire pour pouvoir formuler les règles selon lesquelles, dans une langue, certaines combinaisons de mots sont des phrases possibles, et d'autres pas. Parmi les systèmes de classement des mots qu'on peut *a priori* imaginer, il faut dans la perspective syntaxique s'efforcer d'établir celui qui permet de formuler le plus simplement possible des règles de syntaxe ayant un maximum de généralité.

Les mots peuvent être classés selon leur *distribution*, c'est-à-dire en observant le fait que seuls certains mots peuvent occuper une position donnée dans une construction donnée.

- (6) a. Cet homme —  
           chante  
           dort  
           \*médecin  
           \*intelligent

- b. Cet homme est —  
           \*chante  
           \*dort  
           médecin  
           intelligent

Les mots peuvent être classés selon les possibilités de faire varier leur structure interne, et, comme cela a déjà été évoqué, certaines au moins des caractéristiques morphologiques du mot ont un lien direct avec son comportement syntaxique. Par exemple, en français, le fait que *dort* et *intelligent* ont une distribution différente est corrélé au fait que *dort* peut varier en personne (*dort* / *dormons*) ou en temps (*dort* / *dormait*), alors que *intelligent* n'a pas ces possibilités, tandis qu'inversement, *intelligent* peut varier en genre (*intelligent* / *intelligente*), mais pas *dort*.

Les définitions de la grammaire traditionnelle suggèrent aussi la possibilité d'utiliser des critères sémantiques, mais la question de la relation entre classes grammaticales de mots et types de significations est à poser une fois le classement grammatical des mots établi selon des critères strictement formels, car il n'y a pas une relation nécessaire entre le comportement grammatical et le sens dénotatif des mots pris individuellement, même si au niveau du lexique pris globalement on peut reconnaître des relations entre les classes de mots et certains prototypes sémantiques. Par exemple, il est banal qu'une même notion puisse apparaître à la fois sous forme de nom et sous forme de verbe, le nom et le verbe qui signifient une même notion pouvant selon les cas être formellement apparentés –ex. (7a)– ou totalement différents –ex. (7b).

- (7) a. aimer / amour  
       pleuvoir / pluie

- b. tomber / chute  
           tuer / meurtre

### 3.2. Discussion

Le critère morphologique n'est jamais suffisant à lui seul pour reconnaître toutes les classes de mots qu'il est nécessaire de distinguer du point de vue syntaxique. Notamment, il est impossible de justifier sur la base du seul critère morphologique la répartition des mots invariables en plusieurs classes différentes (cf. par exemple la distinction entre conjonctions et prépositions en français). En français, le critère morphologique est aussi mis en défaut en ce qui concerne la distinction entre noms et adjectifs, car noms et adjectifs varient également

en nombre, et beaucoup de noms (par exemple *boulangier* / *boulangère*) ont des variations en genre<sup>5</sup>.

Par ailleurs, seule une partie de la structure interne des mots est directement pertinente pour la syntaxe : si la morphologie flexionnelle est par définition même à prendre en considération dans un classement des mots orienté vers la possibilité de formuler le plus simplement possible les règles syntaxiques, par contre la prise en compte de la morphologie dérivationnelle dans l'établissement des classes de mots ne ferait qu'introduire des complications inutiles.

En ce qui concerne maintenant le critère distributionnel, il ne peut fonctionner de manière satisfaisante qu'en faisant abstraction des limitations aux combinaisons de mots qui tiennent à des incompatibilités sémantiques (par exemple, n'importe quel adjectif ne peut pas qualifier n'importe quel nom). Le critère distributionnel est indispensable pour tester des hypothèses de classement (au sens où on doit rejeter toute hypothèse de classement des mots qui ne serait pas confirmée par des observations sur leur distribution), mais il serait illusoire de penser pouvoir établir le classement des mots d'une langue en appliquant brutalement, sans aucune hypothèse préalable, un principe selon lequel deux mots qui ne peuvent pas commuter entre eux dans au moins un contexte devraient être reconnus comme appartenant à deux classes différentes.

Une application mécanique de la méthode distributionnelle est par ailleurs difficilement compatible avec le fait que les formes grammaticales d'un même mot (par exemple les formes casuelles du nom, dans les langues où les noms varient en cas) peuvent avoir des distributions différentes.

L'application des tests distributionnels doit enfin tenir compte de l'existence possible de lexèmes à statut multiple, c'est-à-dire aptes à donner naissance à des formes qui peuvent relever de plusieurs types différents, ainsi que de la possibilité d'homonymie entre des formes de type différent issues d'un même lexème à statut multiple. Par exemple, un mot ne se reconnaît pas de manière générale comme nom ou verbe dans l'absolu, mais relativement à une construction à laquelle il participe et au paradigme auquel il se rattache dans la construction en question. Dans la plupart des langues, et notamment en anglais –ex. (8)– et en français –ex. (9), l'homonymie entre formes verbales et formes nominales sémantiquement apparentées est un phénomène banal. Dans de tels cas, il n'est pas difficile d'observer que deux mots à la fois formellement identiques et sémantiquement apparentés ne manifestent pas les mêmes possibilités de varier et de se combiner avec des modificateurs selon qu'ils ont le statut de verbe ou celui de nom – ex.(10). Mais une application brutale de la méthode distributionnelle aboutirait là à une complication inutile, car elle conduirait au moins dans un premier temps à traiter de tels mots comme formant une classe spéciale, alors qu'il est beaucoup plus simple de leur reconnaître le double statut de noms et de verbes.

(8) a. Usually, we break for coffee at ten

b. She rang me during the break

(9) a. La dernière limite est fixée au 15 mai

b. Cette nouvelle loi limite les pouvoirs des préfets

---

<sup>5</sup> L'argument traditionnel selon lequel les variations en genre sont de nature flexionnelle pour les adjectifs et dérivationnelle pour les noms est passablement circulaire, car d'un strict point de vue morphologique les variations en genre des noms ne sont pas différentes de celles des adjectifs.

(10) a. Je-me-limit-er-ai à l'essentiel

b. Je lui ai indiqué l-es-limite-s à ne pas dépasser

Un autre aspect du classement des mots qui échappe totalement à une application mécanique de la méthode distributionnelle est la décision de considérer que deux mots qui n'ont pas exactement la même distribution appartiennent à deux classes différentes ou à deux sous-classes d'une même grande classe.

En conclusion, il n'existe pas de méthode pour régler de façon satisfaisante la question de la répartition des mots en classes sur la seule base de l'observation des variations morphologiques des mots et de leurs possibilités de commutations. Il est indispensable de prendre en considération des hypothèses générales sur des aspects jugés particulièrement importants de l'organisation de la phrase, et de chercher à partir de là la cohérence qui s'établit, de manière variable d'une langue à l'autre, entre ces aspects de l'organisation de la phrase et les caractéristiques morphologiques et/ou distributionnelles des mots. C'est cette démarche qui sera développée aux chapitres 2 à 6.

### 3.3. Mots pleins et mots grammaticaux

Comme beaucoup des notions de base de l'analyse linguistique, la distinction entre mots pleins et mots grammaticaux est une distinction unanimement admise mais qui s'avère difficile à préciser au moyen de tests à la fois opératoires et théoriquement justifiés permettant de distinguer les uns des autres mots pleins et mots grammaticaux<sup>6</sup>.

Le test le plus généralement invoqué pour opérer cette distinction est que les mots pleins constituent des classes ouvertes et peu stables, alors que les mots grammaticaux constituent des classes fermées et relativement stables, mais l'application de ce critère est parfois problématique, et la question de sa justification théorique reste entière.

Une définition qui semble particulièrement intéressante, bien que rarement envisagée, consiste à poser que les classes de mots pleins sont celles où vont se ranger les lexèmes qui sémantiquement ont un sens dénotatif indépendant de toute référence à la construction d'un discours (par exemple, un objet est reconnaissable comme *table* ou *chaise* quelles que soient les conditions dans lesquelles on en parle), alors que la signification des mots grammaticaux est relative à la construction d'un discours (par exemple, il n'y a pas d'objet qui soit intrinsèquement *celui-ci* ou *le mien*). Cette définition ne résout pas toutes les difficultés, mais elle permet au moins d'éviter un certain nombre de faux problèmes.

Par exemple, quels que soient les critères invoqués pour distinguer les mots pleins des mots grammaticaux, les prépositions apparaissent difficiles à classer selon cette distinction. Or, la définition ci-dessus permet sans problème de reconnaître par exemple que la préposition *sur* du français, même si on peut dans certaines constructions lui reconnaître des emplois grammaticalisés, a fondamentalement le statut de mot plein : dans *A est sur B* (par

---

<sup>6</sup> Ces termes sont repris ici faute de mieux pour une distinction dont il semble impossible de se passer, mais pour laquelle n'existe aucune solution terminologique pleinement satisfaisante. Il convient de s'attacher à leur définition plus qu'à la signification que suggère leur étymologie. Le terme de *mot plein* correspond dans la terminologie anglaise à *content word*. Il a bien sûr l'inconvénient de suggérer que les mots qui ne sont pas 'pleins' seraient d'une manière ou d'une autre 'vides', ce qui n'est évidemment pas le cas. On peut être tenté de lui préférer *mot lexical*, mais ce terme implique une opposition entre lexical et grammatical difficile à articuler avec la notion courante de lexique comme ensemble des mots d'une langue, ainsi qu'avec la décision d'utiliser de manière générale le terme de lexème pour se référer à ce qui est commun aux mots-formes regroupés sous une même entrée lexicale. En ce qui concerne *mot grammatical*, le terme anglais correspondant est *function word*, qu'on traduit parfois en français par *mot fonctionnel* ; on trouve parfois aussi le terme de *mot-outil*.



exemple, *Le journal est sur la table*) *sur* signifie entre deux entités A et B une relation dont la reconnaissance est purement un problème de conceptualisation d'une situation de référence.

Beaucoup de difficultés auxquelles on se heurte dans la description grammaticale des langues tiennent au fait que sont constamment à l'œuvre des processus de grammaticalisation par lesquels :

- les mots pleins présentant certaines caractéristiques sémantiques tendent à devenir des mots grammaticaux ;
- les mots grammaticaux tendent à perdre leur autonomie et à devenir des affixes.

## 4. La structure en constituants de la phrase

### 4.1. La notion de constituant

En première approximation, la phrase se présente comme une suite de mots. Mais les locuteurs ont l'intuition de l'existence de 'groupes de mots' qui fonctionnent comme un seul bloc à un certain niveau de la structure de la phrase. Cette intuition peut être précisée par différents tests, notamment en observant des modifications de l'unité phrastique qui impliquent la présence de mots ou groupes de mots en une position différente de leur position canonique (c'est-à-dire différente de celle qu'ils occuperaient dans une phrase indépendante de type assertif minimalement marquée du point de vue discursif). L'ex. (11) montre qu'en français, la focalisation peut fournir des indications sur la structure en constituants de la phrase.

- (11) a. Jean a parlé [de son projet de thèse] [au directeur du département]
- b. C'est [de son projet de thèse] que Jean a parlé — au directeur du département
- c. C'est [au directeur du département] que Jean a parlé de son projet de thèse —
- d. \*C'est [au directeur] que Jean a parlé de son projet de thèse — du département
- e. \* C'est [de son projet] que Jean a parlé — de thèse au directeur du département
- f. \* C'est [de thèse au directeur] que Jean a parlé de son projet — du département

En dehors des tests reposant sur la possibilité de déplacer en bloc les groupes de mots qui forment des constituants, les manipulations suivantes sont généralement considérées comme susceptibles de fournir des indices de l'existence de constituants syntaxiques :

- possibilité / impossibilité de ne pas répéter certains mots ou groupes de mots (ellipse) : *Si Jean veut [venir avec nous au cinéma], il peut —* ;
- possibilité / impossibilité de substituer à un groupe de mots une proforme : *Jean m'[en] a déjà parlé, [de son projet de thèse]* ;
- possibilité de coordonner un constituant avec d'autres constituants du même type : *Jean veut [vendre sa vieille voiture] et [acheter une moto]* ;

Ces tests ne donnent malheureusement pas toujours des résultats dont l'interprétation est évidente, et des tests différents appliqués au même groupe de mots peuvent donner des indications contradictoires. Il faut notamment être très prudent dans la manipulation des tests de coordination, car le fonctionnement de la coordination peut s'accompagner de phénomènes d'ellipse qui ont comme résultat que souvent, cela n'a pas de sens de vouloir comparer directement le deuxième terme d'une coordination à la partie de la phrase qui précède. Par

exemple, la possibilité d'ellipse qu'indiquent les parenthèses dans des phrases comme *Jean a posé le journal sur la table et (a posé) le colis sur la chaise* ou *Jean est parti en voiture et Marie (est partie) à pied* a pour effet de faire apparaître comme deuxième terme d'une coordination les groupes de mots *le colis sur la chaise* et *Marie à pied*, qui dans ces constructions représentent respectivement la forme elliptique d'un groupe verbal et d'une phrase, et qu'il serait donc absurde de vouloir utiliser tels quels pour dégager la structure en constituants de *Jean a posé le journal sur la table* et *Jean est parti en voiture*.

#### 4.2. Types majeurs de constituants

Il est parfois difficile de prouver la pertinence des divers types de regroupements de mots en constituants postulés par les théories syntaxiques qui font un usage particulièrement systématique de cette notion<sup>7</sup>.

Les seuls constituants syntaxiques qui se laissent généralement isoler sans difficulté (et les seuls dont il est crucial de reconnaître l'existence dans les analyses syntaxiques) sont :

- les constituants nominaux, formés par un nom accompagné d'un nombre variable de termes de nature diverse qu'on peut désigner de manière générale comme 'modificateurs du nom' ;

- les constituants phrastiques, c'est-à-dire des fragments de phrase ayant eux-mêmes une structure interne de type phrastique, comme les 'propositions subordonnées' des analyses syntaxiques traditionnelles ;

- les constituants prépositionnels (ou selon les langues, postpositionnels) formés par une préposition et un constituant nominal (ou par un constituant nominal et une postposition) ;

- les constituants adjectivaux et adverbiaux, formés par un adjectif ou un adverbe accompagné d'un nombre variable de modificateurs (*Jean est [très fier de sa nouvelle voiture]*, *Jean court [presque aussi rapidement que Paul]*).

Il faut toutefois noter que :

- souvent, les prépositions ou postpositions s'avèrent avoir de manière plus ou moins nette des propriétés de morphèmes liés (cf. section 6), ce qui fait que des constituants traditionnellement reconnus comme 'groupes prépositionnels (ou postpositionnels)' peuvent être réanalysables comme des groupes nominaux dont le premier mot (ou le dernier) s'affixe un morphème qui marque la fonction du constituant nominal dans la construction à laquelle il participe ;

- les notions de constituant adjectival et de constituant adverbial ne sont intéressantes que dans les langues où adjectifs et adverbes ont la propriété de pouvoir prendre comme compléments des groupes nominaux ou prépositionnels, ce qui n'est pas universel.

Et il est important d'avoir à l'esprit qu'on se heurte très souvent à des problèmes qui dans beaucoup de langues ne semblent pas avoir de solution satisfaisante :

- lorsqu'on veut justifier la reconnaissance d'un groupe verbal formé par le verbe et les termes de sa construction autres que le sujet (cf. section 4.3 ci-dessous) ;

- lorsqu'on veut analyser systématiquement comme une hiérarchie de constituants emboîtés les uns dans les autres la structure interne des constituants nominaux et des groupes verbaux.

---

<sup>7</sup> C'est pour cette raison que dans ce cours, on n'aura recours qu'exceptionnellement à la représentation de la structure en constituants des phrases sous forme d'arbre. Les représentations arborescentes ont en effet l'inconvénient d'obliger à tout instant à prendre position sur des aspects problématiques de la structure en constituants qui n'ont aucune relation directe avec les questions discutées. On s'en tiendra généralement à l'utilisation de crochets, qui lors de la discussion de questions particulières a l'avantage de permettre d'explicitier seulement les aspects pertinents de la structure en constituants.

### 4.3. La notion de groupe verbal

Parmi les types de constituants dont les analyses syntaxiques récentes posent souvent l'existence, la notion de groupe verbal, bien qu'indiscutablement utile pour décrire de nombreux mécanismes syntaxiques, est par bien des aspects problématique, surtout dans certaines langues. Il est notamment difficile de soutenir l'existence d'un groupe verbal réunissant le verbe et les termes de sa construction autres que le sujet dans les langues qui placent systématiquement le verbe en tête de phrase et le sujet immédiatement après le verbe. Ce problème cesse toutefois d'en être un si on admet que, de même que les constituants nominaux ont en principe pour tête un nom, les phrases et constituants phrastiques ont en principe pour tête un verbe ; en effet, dans cette optique (qui est celle qu'adopte ce cours), le groupe verbal n'a pas à être défini de manière indépendante, car ce n'est ni plus ni moins qu'une phrase non saturée (c'est-à-dire à laquelle il manque un constituant nominal en fonction de sujet pour être une unité phrastique complète). Ce qui sur un plan théorique crée des difficultés, c'est la conception selon laquelle le sujet est extérieur à la construction maximale ayant pour tête le verbe. Cette position a été défendue par les premières versions de la grammaire générative, mais elle est de plus en plus abandonnée.

### 4.4. Structure en constituants des phrases complexes par subordination

A propos des constituants phrastiques (ou phrases enchâssées), un reproche qu'on peut faire à la grammaire traditionnelle est qu'elle présente systématiquement les phrases complexes comme des enchaînements de 'propositions', y compris dans des cas de phrases complexes par subordination où il est évident que la construction est plutôt à décrire comme un enchâssement, car la subordonnée fonctionne comme constituant de la principale.

Par exemple, la grammaire traditionnelle découpe (12a) en une proposition principale *je vois* et une proposition subordonnée (*que*) *vous êtes bien arrivé* et présente la conjonction *que* comme 'reliant' ces deux unités phrastiques. Mais en réalité, cet énoncé constitue globalement une unité phrastique, à l'intérieur de laquelle une autre unité phrastique fonctionne comme un constituant : *que vous êtes bien arrivé* dans l'ex. (12a) est tout aussi intégré à l'unité phrastique qui a pour noyau le verbe *voir* que *un enfant* dans l'ex. (12b) : l'impossibilité de (12c) ou (12d) montre que le verbe *voir* peut ainsi prendre un complément phrastique ou un complément nominal qui par ailleurs n'ont pas exactement les mêmes propriétés syntaxiques, mais qui sont néanmoins équivalents au sens où ils saturent également l'une des valences du verbe *voir*. Seule la notion d'enchâssement d'unités phrastiques –qu'on peut figurer soit au moyen de représentations arborescentes, soit au moyen de parenthèses emboîtées, comme en (12e)– peut correctement rendre compte de la relation entre les deux structures phrastiques impliquées dans la construction de (12a).

- (12) a. Je vois [que vous êtes bien arrivé]  
       cf. Ce que je vois, c'est [que vous êtes bien arrivé]
- b. Je vois [un enfant]  
       cf. Ce que je vois, c'est [un enfant]
- c. \*Je vois [un enfant] [que vous êtes bien arrivé]
- d. \*Je vois [que vous êtes bien arrivé] [un enfant]

e. [<sub>S1</sub> Je vois [<sub>S2</sub> que vous êtes bien arrivé <sub>S2</sub>] <sub>S1</sub>]

(et non pas [<sub>S1</sub> Je vois <sub>S1</sub>] que [<sub>S2</sub> vous êtes bien arrivé <sub>S2</sub>])

#### 4.5. La structure interne des constituants : tête et dépendants

Parmi les mots qui forment un constituant, la tête est celui qui détermine le statut syntaxique du constituant pris en bloc. Ainsi, le constituant nominal a en principe pour tête un nom, et la phrase a en principe pour tête un verbe. Les autres mots ou groupes de mots qui entrent dans la formation du constituant peuvent être désignés du terme générique de *dépendants* (on dit aussi *modifieurs*, et ce terme sera largement utilisé ici, mais il faut être attentif au fait que certains auteurs donnent à ce terme un sens plus retreint).

La notion de tête est fondamentalement le reflet syntaxique du fait que, sémantiquement, les dépendants précisent un sens dénotatif qui est donné par la tête : *le nouveau directeur du département* dénote un individu catégorisé comme *directeur*, et *Jean m'a parlé de toi* se réfère à un événement catégorisé comme *parler*.

D'un point de vue strictement syntaxique, la reconnaissance d'un mot comme tête d'un constituant est évidente lorsque ce mot a la même distribution syntaxique que le constituant dont il est la tête : lorsque le constituant nominal peut se réduire à un nom, et lorsqu'une forme verbale peut constituer une phrase à elle seule, il n'y a pas de difficulté à reconnaître le nom comme tête du constituant nominal, et le verbe comme tête de la phrase. Par contre, le caractère obligatoire de certains dépendants peut poser problème à la reconnaissance d'une relation tête-dépendant ; par exemple, dans beaucoup de langues, le sujet est un terme obligatoire de la phrase, ce qui explique que la reconnaissance du verbe comme tête de la phrase et du sujet comme dépendant du verbe a mis longtemps à s'imposer, et ne fait toujours pas l'unanimité chez les linguistes.

La reconnaissance d'une relation tête-dépendants est particulièrement problématique lorsque le nom ou le verbe sont nécessairement associés à des mots grammaticaux (déterminants dans le cas du nom, auxiliaires dans le cas du verbe) avec lesquels ils partagent les propriétés grammaticales communément considérées comme caractéristiques des têtes de constituants<sup>8</sup>.

Parmi les critères morphosyntaxiques susceptibles de justifier la relation tête-dépendant, on peut être tenté de poser comme principe qu'une tête peut déterminer certaines caractéristiques formelles de ses dépendants, mais pas l'inverse. En réalité, d'un point de vue typologique, l'application systématique d'un tel critère ne peut aboutir qu'à des incohérences, notamment du fait de la variété des phénomènes d'accord observés dans les langues. Il est préférable de considérer que les marques morphologiques d'une relation tête-dépendant, lorsqu'elles existent, peuvent aussi bien apparaître sur la tête que sur le dépendant (ou éventuellement sur les deux à la fois, comme par exemple en turc où, dans un constituant nominal comme *ev-in kapı-sı* 'la porte de la maison', le dépendant *ev* 'maison' est au cas génitif, et la tête *kapı* 'porte' est à la forme possessive –à lui seul, le mot *kapı-sı* signifierait 'sa porte').

#### 5. L'accord

Certaines langues (par exemple le bambara) ignorent totalement les phénomènes

<sup>8</sup> C'est ce qui a conduit les linguistes de l'école chomskyenne à une révision radicale de la notion de tête, selon laquelle le constituant nominal et la phrase sont la projection de 'têtes fonctionnelles' abstraites, qui ne sont pas nécessairement apparentes dans la phrase réalisée, et qui lorsqu'elles se manifestent peuvent prendre la forme d'affixes flexionnels.

grammaticaux traditionnellement décrits en termes d'accord, alors que dans d'autres langues (par exemple les langues bantoues) ces phénomènes interviennent massivement dans la construction de la phrase et du constituant nominal.

Il semble difficile sinon impossible de donner une définition générale de l'accord qui s'applique exactement aux phénomènes couramment rangés à cette rubrique : toutes les définitions qui viennent à l'esprit s'avèrent, ou trop larges, ou trop restrictives. Une solution prudente est de considérer que le terme d'accord regroupe trois types de mécanismes distincts bien que présentant un certain nombre de points communs :

- l'accord entre le nom et ses dépendants dans le cadre du constituant nominal,
- l'accord entre le constituant nominal et des mots appartenant à la même phrase qui sont avec le constituant nominal dans une relation syntaxique particulière (souvent, mais pas toujours, il s'agit de mots avec lesquels le constituant nominal est dans une relation de dépendance directe),
- l'accord des pronoms avec un constituant nominal susceptible de représenter le même référent, phénomène fondamentalement discursif qui n'implique aucune relation syntaxique particulière et qui n'implique même pas que le constituant nominal avec lequel le pronom s'accorde soit effectivement présent dans le discours.

L'accord entre le nom et ses dépendants dans le cadre du constituant nominal consiste en ce que plusieurs mots parmi ceux qui entrent dans la formation du constituant nominal présentent un phénomène de co-variation par lequel ils contribuent conjointement à l'expression de caractéristiques grammaticales ou sémantiques du constituant nominal ou de son référent (nombre, cas, genre). L'accord entre le constituant nominal et un mot avec lequel il entretient une relation syntaxique particulière (par exemple en français l'accord entre sujet et verbe, ou l'accord entre sujet et adjectif attribut) consiste en ce que un mot qui est dans une relation déterminée avec le constituant nominal présente des variations qui reflètent certaines caractéristiques grammaticales du constituant nominal ou certaines caractéristiques sémantiques de son référent.

Les phénomènes d'accord (accord sujet-verbe, accord nom-modifieur, etc.) constituent un domaine qui permet particulièrement bien de montrer la différence entre une approche *dérivationnelle* de la syntaxe, systématisée et popularisée par la grammaire transformationnelle des années 60, qui considère les combinaisons de mots effectivement observées comme le produit de règles de transformation appliquées à des structures syntaxiques abstraites, et une approche qui s'impose de rendre directement compte des contraintes sur les combinaisons de mots bien formées, sans avoir recours à des structures syntaxiques abstraites.

Les formulations de la grammaire traditionnelle suggèrent une orientation des phénomènes d'accord : on ne dit pas traditionnellement que le sujet et le verbe s'accordent, mais plutôt que le verbe s'accorde avec son sujet, que l'adjectif épithète s'accorde avec le nom, etc. Ces formulations impliquent que les traits donnant lieu à un accord entre deux termes d'une construction appartiennent fondamentalement à l'un des deux termes de la construction (le sujet dans la relation sujet-verbe, le nom dans la relation nom-épithète), et que leur manifestation sur l'autre terme de la construction résulte d'une règle de copie ; par exemple, un syntagme *sujet + verbe* avec un accord en nombre entre le sujet et le verbe sera considéré comme ayant une 'structure profonde' *Sujet-NOMBRE<sub>x</sub> Verbe* qu'une règle de copie transforme en *Sujet-NOMBRE<sub>x</sub> Verbe-NOMBRE<sub>x</sub>*, comme cela est illustré en (13).

- (13) a. enfant-DEF-SG venir-FUT → enfant-DEF-SG venir-FUT-SG  
           'L'enfant viendra'

- b. enfant-DEF-PL venir-FUT —> enfant-DEF-PL venir-FUT-PL  
 ‘Les enfants viendront’

Autrement dit, dans une telle approche de la syntaxe, les caractéristiques flexionnelles des mots peuvent être manipulées par les règles de syntaxe.

Dans une approche non dérivationnelle, on rendra compte du même phénomène en considérant que, pour chaque mot ou syntagme susceptible d’entrer dans une relation syntaxique qui donne lieu à un accord, la morphologie flexionnelle détermine un trait de nombre qui peut prendre les valeurs *SG* (singulier), *PL* (pluriel) ou rester non spécifié (pour les mots ou syntagmes dont la forme à elle seule ne permet pas de spécifier une valeur de nombre, comme par exemple en anglais le groupe nominal *the salmon* ‘le(s) saumon(s)’ ou la forme verbale *swimmed* ‘nagea / nagèrent’) ; ensuite, la construction d’un syntagme mettant en jeu des phénomènes d’accord se caractérise par la contrainte que les deux termes du syntagme ne doivent pas avoir deux spécifications différentes du même trait –ex. (14).

- (14) a. [l’enfant]<sub>SG</sub> [viendra]<sub>SG</sub> : combinaison bien formée  
 b. \*[l’enfant]<sub>SG</sub> [viendront]<sub>PL</sub> : combinaison mal formée  
 c. \*[les enfants]<sub>PL</sub> [viendra]<sub>SG</sub> : combinaison mal formée  
 d. [les enfants]<sub>PL</sub> [viendront]<sub>PL</sub> : combinaison bien formée

A première vue, l’approche dérivationnelle a l’avantage de fournir une explication d’un phénomène que l’approche non dérivationnelle se borne à décrire. Mais la question est justement de savoir s’il est raisonnable de chercher une explication purement syntaxique au phénomène de l’accord (c’est-à-dire de voir dans l’accord le résultat de la diffusion de certains traits d’un élément de la structure de la phrase à d’autres), ou s’il ne vaut pas mieux s’orienter vers une explication fondamentalement sémantique selon laquelle les contraintes syntaxiques d’accord, du point de vue du fonctionnement global de la langue, tiennent à ce que certaines informations sur un même référent peuvent se trouver répétées sur les deux termes d’un syntagme, ou, plus généralement, réparties entre les deux termes d’un syntagme. Dans cette optique, on dira par exemple que dans l’ex. (14), l’explication des incompatibilités constatées en (b) et (c) est que, dans le système du français, le verbe *venir* qui représente une action et le sujet qui se réfère à l’agent de cette action doivent tous les deux porter des marques morphologiques qui spécifient si l’agent de l’action de venir est envisagé comme constitué d’un individu unique ou d’un groupe d’individus.

L’inconvénient majeur de l’approche dérivationnelle de l’accord (et donc d’une théorie purement syntaxique de l’accord) est qu’elle ne fournit un traitement simple que lorsque les traits supposés se diffuser d’un terme A à un terme B d’une construction apparaissent toujours maximalement spécifiés dans la morphologie du terme A, c’est-à-dire si les informations que livre la morphologie du terme B en ce qui concerne le trait donnant lieu à accord sont toujours redondantes avec des informations présentes sans ambiguïté dans la morphologie du terme A. Mais ce n’est pas toujours le cas, comme on peut le voir à partir de l’ex. anglais (15).

- (15) a. The whale is swimming  
 b. The whales are swimming

- c. The whale swimmed
- d. The whales swimmed
- e. The salmon is swimming
- f. The salmon are swimming
- g. The salmon swimmed

En effet, dans les phrases de l'ex. (15) :

- en (a/b), la distinction singulier / pluriel est marquée à la fois sur le sujet et sur le verbe ;
- en (c/d), la distinction singulier / pluriel est marquée sur le sujet, mais pas sur le verbe ;
- en (e/f), la distinction singulier / pluriel est marquée sur le verbe, mais pas sur le sujet ;
- en (g), la distinction singulier / pluriel n'est marquée, ni sur le sujet, ni sur le verbe.

Par conséquent, l'hypothèse de la diffusion du trait de nombre du sujet au verbe rend immédiatement compte des phrases (a) à (d), mais pose un problème pour les phrases (e) à (g). Pour les phrases (e) et (f), l'hypothèse d'une diffusion du trait de nombre du sujet au verbe oblige à admettre qu'il n'y a pas en anglais un nom *salmon* 'saumon(s)' non marqué pour le nombre, mais deux homonymes, *salmon* et *salmon-PL*, avec une marque de pluriel invisible qui gouverne néanmoins l'accord du verbe. Le traitement de la phrase (g) est encore plus problématique, puisque dans cette phrase, il n'y a même pas la possibilité d'observer dans le verbe une marque de nombre qui pourrait être analysée comme la manifestation indirecte de la marque de pluriel invisible qu'on suppose portée par le nom sujet.

Observons aussi les données de l'ex. (16) sur l'accord en personne entre sujet et verbe en espagnol:

- (16) a. Nosotros    somos            jóvenes  
      nous            être.TAM.1P    jeune.PL  
      'Nous, nous sommes jeunes'
- b. Vosotros    sois            muy amabes  
      vos            être.TAM.S2P    très    aimabe.PL  
      'Vous, vous êtes très aimables'
- c. Ellos son            muy simpáticos  
      eux    être.TAM.S3P    très    sympathique.PLM  
      'Eux, ils sont très sympathiques'
- d. Somos            jóvenes  
      être.TAM.S1P    jeune.PL  
      'Nous sommes jeunes'
- e. Sois            muy amabes  
      être.TAM.S2P    très    aimabe.PL  
      'Vous êtes très aimables'
- f. Son            muy simpáticos  
      être.TAM.S3P    très    sympathique.PLM  
      'Ils / elles sont très sympathiques'

- g. Los            ciudadanos    queremos    más    justicia  
 DEF.PLM    citoyen.PLM    vouloir.TAM.S1P    plus    justice  
 ‘Nous les citoyens, nous voulons plus de justice’ (litt. ‘Les citoyens voulons plus de justice’)
- h. Los            hombres    sois            todos    iguales  
 DEF.PLM    homme.PL    être.TAM.S2P    tous.PLM    pareil.PL  
 ‘Vous les hommes, vous êtes tous pareils’ (litt. ‘Les hommes êtes tous pareils’)
- i. Los            españoles    son            así  
 DEF.PL M    Espagnol.PLM    être.TAM.S3P    ainsi  
 ‘Les Espagnols sont comme ça’

Dans les phrases (a) à (c) de l'ex. (16), il n'y a aucune difficulté à supposer qu'une partie de l'information livrée par le sujet est copiée sur le verbe. Par contre, les phrases (d) à (f) ne comportent aucun sujet, et la seule façon de maintenir l'idée que les marques de personne du verbe lui sont imposées par son sujet consiste à admettre l'existence d'un sujet 'invisible' ou 'effacé' capable d'intervenir dans les mécanismes syntaxiques exactement comme un mot concret ; par exemple, on devra admettre que la phrase (16f) comporte comme sujet un pronom invisible de 3ème personne du pluriel masculin : *pro<sub>3PM</sub> son muy simpáticos*, alors que (16d) comporte un pronom invisible de première personne du pluriel masculin, et (16e) un pronom invisible de deuxième personne du pluriel masculin, ce qui est déjà une complication.

Une difficulté supplémentaire apparaît avec les phrases (g) à (i), qui à la différence des phrases (d) à (f) comportent un sujet, mais dans lesquelles le verbe manifeste des possibilités de variation en personne indépendantes du groupe nominal en fonction de sujet. Ici encore, la seule façon de sauver l'idée que les marques de personne du verbe lui sont imposées par son sujet consiste à admettre que le 'véritable' sujet n'est pas celui qui se voit, mais plutôt un pronom invisible par rapport auquel le sujet apparent constitue une sorte d'apposition.

En résumé, la conception de l'accord comme copie de traits d'un terme d'une construction sur un autre, qui à première vue semble fournir une explication à la fois simple et élégante du phénomène, conduit nécessairement à postuler des structures syntaxiques abstraites incluant notamment des éléments 'invisibles' dont la seule justification est précisément de permettre de rendre certaines données compatibles avec une hypothèse sur le fonctionnement des mécanismes syntaxiques. Par contre, si on ne décide pas d'orienter les mécanismes d'accord et qu'on considère plutôt que les mots ou constituants impliqués dans un mécanisme d'accord partagent certains traits qui concernent un même référent, il n'y a plus aucune nécessité d'avoir recours à des mots ou morphèmes invisibles pour décrire les mécanismes d'accord. Dans cette optique, la particularité des langues qui ont un accord sujet-verbe est que certaines informations sur l'un des participants au procès signifié par le verbe apparaissent dans la morphologie verbale ; selon les langues, il peut arriver que le participant en question soit en outre nécessairement représenté par un mot ou groupe de mots (le sujet), et dans ce cas il est possible que l'information présente dans la morphologie verbale soit redondante avec des informations spécifiées au niveau du mot ou groupe de mots en fonction de sujet, mais il n'y a pas de raison de considérer les situations où ce n'est pas le cas comme des situations exceptionnelles dont l'explication nécessiterait le recours à des structures syntaxiques abstraites.



## 6. Formes libres et formes liées, cliticisation et composition

### 6.1. Considérations générales

Le principal problème que pose la notion de mot est qu'il est souvent difficile de distinguer les morphèmes grammaticaux liés (clitiques ou affixes) des mots grammaticaux. Les mots grammaticaux manifestent universellement une tendance à perdre leur autonomie pour devenir des morphèmes grammaticaux liés, mais cette perte d'autonomie se fait progressivement, et il est souvent délicat de caractériser de façon satisfaisante des éléments grammaticaux qui sont déjà engagés dans ce processus sans être encore parvenus à l'étape finale. Les linguistes qui décrivent des langues sont constamment confrontés à de telles situations, et il convient en particulier d'être attentif à ne pas se laisser exagérément influencer par les traditions grammaticales et les conventions orthographiques, qui ne traitent pas toujours cette question de façon parfaitement cohérente. Par exemple, l'orthographe française et la grammaire traditionnelle du français traitent comme mots à part les unités suivantes, qui ont pourtant plus ou moins clairement des propriétés de morphèmes liés :

- les pronoms personnels conjoints (*je, tu, me*, etc.), qui s'attachent au verbe ;
- le morphème de négation *ne*, qui s'attache au premier mot du groupe verbal<sup>9</sup> ;
- les prépositions *de, à* et *en*, les articles, les déterminants démonstratifs et les déterminants possessifs, qui s'attachent au premier mot du constituant nominal.

Le terme de clitique est un terme dont il est difficile de donner une définition précise, et qui est utilisé de façon très variable d'une tradition à une autre, mais qui répond au souci de préciser le statut de morphèmes ou séquences de morphèmes qui n'ont, ni tout à fait le comportement caractéristique d'un mot, ni tout à fait celui d'un affixe ou d'une séquence d'affixes.

Mais il convient aussi d'aborder dans cette section la question de la composition, car un aspect de cette question est que la combinaison de deux mots pleins peut aboutir à une forme ayant globalement les propriétés d'un mot unique.

### 6.2. Formes libres et formes liées

Une forme (morphème ou séquence de morphèmes) peut être caractérisée comme libre ou liée selon qu'elle peut ou non constituer à elle seule un mot. Une forme libre peut être qualifiée de minimale si toute suppression d'un de ses éléments constitutifs lui enlève son caractère de mot.

Plusieurs types de critères peuvent être utilisés pour reconnaître qu'une limite entre deux segments significatifs sépare l'une de l'autre deux formes libres, ou qu'au contraire au moins l'une des deux formes situées de part et d'autre de cette limite est une forme liée.

Un premier type de critère repose sur la notion d'indépendance prosodique des formes libres. Une forme qu'il est naturel d'énoncer en isolation, par exemple en répondant à une question, est nécessairement une forme libre. Si une phrase est interrompue par des pauses, les pauses tendent à coïncider avec des limites entre formes libres. Il en va de même pour l'insertion d'incises. Les formes libres se prêtent plus facilement que les formes liées à une mise en relief par l'intonation. Dans les langues à accent de mot, seules les formes libres peuvent constituer des unités accentuelles autonomes ; si une forme liée porte un accent, cet accent ne lui appartient pas en propre, il est déterminé dans le cadre de l'unité accentuelle à laquelle la forme liée en question est rattachée.

---

<sup>9</sup> En français, le premier mot du groupe verbal est presque toujours le verbe, mais la nécessité de préciser que *ne* se rattache au premier mot du groupe verbal plutôt qu'au verbe apparaît à l'infinitif –cf. *ne-pas le-lui-donner*.

Un deuxième type de critère repose sur le fait que les possibilités d'insertion d'éléments supplémentaires entre un morphème lié et le reste du mot dont il fait partie sont limitées à un petit nombre de morphèmes qui sont eux-mêmes des morphèmes liés. Ainsi, *-ra* dans *L'avion atterri-ra* et *va* dans *L'avion va atterrir* apportent également une information de type temporel sur l'événement représenté par le verbe *atterrir* et une information sur la personne et le nombre du sujet de ce verbe, mais le fait que *-ra* est suffixé à la base verbale *atterri-*, alors qu'il serait incorrect de considérer *va* comme préfixé à l'infinitif, se manifeste notamment par la différence de comportement lors de l'insertion d'adverbes –ex. (17).

(17) a. L'avion va probablement bientôt atterrir

b. \*L'avion atterri- probablement bientôt -ra

Un troisième type de critère repose sur l'impossibilité de traiter isolément les morphèmes liés dans des opérations auxquelles se prêtent les mots ou groupes de mots, telles que la focalisation par une construction clivée ou la coordination, cette impossibilité étant particulièrement remarquable dans des cas où le morphème qui donne lieu à tel blocage contribue à la signification globale de la phrase de façon équivalent à un mot ou groupe de mots se prêtant aux opérations en question – ex. (18).

(18) a. Marie aime Jean —> C'est Jean que Marie aime

b. Marie m'aime / \*C'est me que Marie aime

c. Marie connaît Jean —> Marie connaît Jean et Nathalie  
Marie connaît Nathalie

d. Marie me connaît / \*Marie me et te connaît  
Marie te connaît

Un quatrième type de critère repose sur la possibilité d'interactions phonologiques spécifiques (c'est-à-dire de processus phonologiques autres que ceux qui se produisent régulièrement aux frontières de mots) entre les morphèmes liés et les autres éléments du mot dont ils font partie : allomorphie *dé-* ~ *dés-* en (19a), élision du *a* de l'article féminin singulier *la* avec une partie des noms commençant par voyelle en (19b).

(19) a. dé-militariser / dés-armer

b. la fourchette / la hache / l'assiette

Mais l'application de ces critères donne parfois des résultats peu nets, ou difficiles à interpréter. Il faut notamment être très prudent avec les tests d'insertion, qui conduisent souvent à des raisonnements circulaires, car leur interprétation est largement dépendante d'une hypothèse sur ce qui pourrait constituer la base de rattachement d'un morphème lié.

En cas de doute, l'existence d'une interaction phonologique spécifique doit être considérée comme décisive pour reconnaître à un élément grammatical le statut de forme liée. Par exemple, on peut hésiter à reconnaître comme forme liée l'article défini du français, car si on considère comme base de rattachement potentielle le nom qui est la tête du groupe nominal, on doit conclure que le test d'insertion donne un résultat négatif (cf. par exemple *la triste histoire*, où un adjectif s'intercale entre l'article et le nom) ; mais l'existence de processus

phonologiques spécifiques entre l'article défini et ce qui lui succède immédiatement (notamment l'élision possible de *a* au féminin singulier : *l'étonnante histoire*) conduit à reconsidérer la question, et à voir dans l'article défini un morphème lié dont la base de rattachement est le premier mot du constituant nominal.

### 6.3. Mots grammaticaux, clitiques et affixes

Le problème à envisager maintenant est que les morphèmes grammaticaux (ou séquences de morphèmes grammaticaux) reconnaissables selon les critères précédents comme formes liées ne manifestent pas toujours au même degré des caractéristiques d'intégration à un mot qui les engloberait. Notamment, il est évident qu'on compliquerait inutilement la description en reconnaissant le statut d'affixe à des formes liées qui occupent dans la linéarité de la phrase la même place qu'une forme libre sémantiquement équivalente. Dans de tels cas, on a un traitement beaucoup plus simple en décrivant de telles formes liées comme étant fondamentalement des mots dont la position dans la phrase est déterminée par les règles de la syntaxe, mais des mots susceptibles de se *cliticiser*, c'est-à-dire de perdre leur autonomie prosodique pour former de ce point de vue une unité avec un mot plein auquel ils sont adjacents. C'est par exemple le cas en anglais pour 've forme réduite de *have*, ou 's forme réduite de *is*.

Inversement, ne pas reconnaître comme affixe une forme grammaticale liée qui s'attache invariablement à un mot plein d'un type précis a toutes les chances de compliquer inutilement la description.

Toute la question est de savoir comment traiter les formes grammaticales liées qui ne relèvent pas de l'une de ces deux situations extrêmes, et à partir de quel moment il est préférable de considérer une forme grammaticale liée, non pas comme la forme cliticisée d'un mot dont la place dans la phrase est déterminée par la syntaxe au même titre que celle de n'importe quel autre mot, mais comme un affixe qui n'est concerné par les règles de la syntaxe que dans la mesure où sa présence peut avoir une incidence sur les propriétés syntaxiques du mot dont il fait partie. En effet, étant donné une forme grammaticale liée qui n'est de manière évidente, ni la variante cliticisée d'un mot grammatical existant par ailleurs sous forme libre, ni un affixe qui n'a d'existence que comme élément constitutif d'un type précis de mot, on peut toujours envisager, soit de traiter cette forme comme un mot grammatical qui aurait la particularité de toujours se cliticiser sur le mot plein au contact duquel le placent les règles de la syntaxe, soit de la traiter comme un *affixe de constituant* qui peut entrer dans la constitution de mots de types variés, pourvu que ces mots occupent la première ou la dernière position dans un constituant d'un type donné.

Par exemple, le morphème de génitif 's de l'anglais est indiscutablement une forme liée (ne serait-ce que du fait qu'il est impossible de le détacher dans la prononciation), mais ce n'est pas non plus un affixe typique, puisqu'il peut s'attacher à n'importe quel type de mot, pourvu que ce mot termine un constituant nominal (cf. par exemple *the man we saw yesterday's car*). Il est possible *a priori*, soit de l'analyser comme une postposition qui se cliticise nécessairement sur le dernier mot du constituant nominal avec lequel elle se combine, soit comme un suffixe susceptible de s'attacher à n'importe quel mot qui termine un constituant nominal, et qui détermine la façon dont le constituant nominal au dernier mot duquel il est attaché peut participer à la construction d'une phrase.

Il n'est pas question de vouloir trancher dans une introduction élémentaire à la syntaxe une question aussi complexe et aussi controversée, aussi on se limitera dans ce qui suit à indiquer les critères utilisables pour classer les formes grammaticales liées, selon leur degré d'intégration à leur base de rattachement, comme clitiques (faiblement intégrés) ou affixes (fortement intégrés).

Le fait qu'une forme liée dépende phonologiquement de manière spécifique de la base à laquelle elle est rattaché n'est pas un critère de classement, car il s'agit d'une propriété générale des formes liées. Par contre, le fait que la base de rattachement elle-même subisse du fait du rattachement de la forme liée des processus phonologiques autres que ceux qui se produisent régulièrement à une frontière de mots peut être considéré comme caractéristique d'un processus d'affixation. Le moindre degré d'intégration d'une forme grammaticale liée reconnue comme clitique doit donc se manifester par le fait que la base à laquelle cette forme s'attache est une forme libre, c'est-à-dire un mot déjà constitué, et que cette base ne subit en présence du clitique aucune modification spécifique. Par contre, il peut arriver que les affixes aient pour base une forme qui n'existe pas par elle-même comme mot, ou que la forme de la base soit influencée de façon spécifique par l'adjonction d'un affixe. Par exemple en finnois –ex. (20), l'harmonisation vocalique selon le trait antérieur / postérieur s'applique aussi bien au suffixe possessif *-nsa* qu'au clitique interrogatif *-ko*, mais *-ko* se rattache toujours à une base qui forme par elle-même un mot et qui ne subit aucune modification du fait du rattachement de *-ko*, alors qu'avec au moins certains mots, *-nsa* s'attache à une base qui n'existe pas comme mot.

(20) a.	koira	/	koira-ko ?	/	koira-nsa
	'chien'		chien-INTER		chien-3S
			'un chien ?'		'son chien'
b.	ystävä	/	ystävä-kö ?	/	ystävä-nsä
	'ami'		ami-INTER		ami-3S
			'un ami ?'		'son ami'
e.	nainen	/	nainen-ko?	/	<u>naise</u> -nsa
	'femme'		femme-INTER		femme-3S
			'une femme ?'		'sa femme'

Il y a de ce point de vue en anglais une différence nette entre le morphème de pluriel *-s* (qui est un affixe du nom, et qui peut imposer une modification de la base) et le morphème génitif *'s* (qui peut s'attacher à n'importe quel mot qui termine un constituant nominal, et qui n'entraîne jamais de modification de ce mot) –ex. (21).

(21)	knife	'couteau'	pl. knives	gén. knife's
	life	'vie'	pl. lives	gén. life's
	scarf	'écharpe'	pl. scarves	gén. scarf's

Ces phénomènes propres à l'affixation se manifestent de façon particulièrement radicale dans des cas où l'expression d'une valeur grammaticale normalement véhiculée par un affixe fait apparaître une forme qui phonologiquement ne peut pas s'analyser comme résultant de la combinaison de l'affixe qui véhicule en principe cette valeur et de la base à laquelle devrait en principe s'ajouter l'affixe en question, comme dans certaines formes de pluriel de l'anglais –ex. (22c).

(22) a.	book 'livre'	pl. book-s
b.	knife 'couteau'	pl. knife-s
c.	mouse 'souris'	pl. mice

Dans le même ordre d'idées, dans les langues à accent de mot, on peut s'attendre à ce que l'adjonction d'un clitique n'introduise aucun changement dans l'accentuation du mot auquel il s'attache, alors que les affixes sont à prendre en considération dans les règles qui déterminent la place de l'accent de mot.

Un autre critère qui peut s'avérer utile est qu'une forme grammaticale liée qui a plusieurs rattachements possibles à l'intérieur d'une même phrase, comme par exemple le clitique objet du polonais *-go* –ex. (23), est à reconnaître comme clitique plutôt que comme affixe.

- (23) a. Tak bardzo-go      chciałbym      spotkać      w Krakowie  
                  tellement-O3SM      vouloir.TAM.S1S      rencontrer.INF      à Cracovie.CAS
- b. Tak bardzo      chciałbym-go      spotkać      w Krakowie  
                  tellement      vouloir.TAM.S1S-O3SM      rencontrer.INF      à Cracovie.CAS
- c. Tak bardzo      chciałbym      spotkać-go      w Krakowie  
                  tellement      vouloir.TAM.S1S      rencontrer.INF-O3SM      à Cracoviee.CAS

'J'aurais tellement voulu le rencontrer à Cracovie'

On mentionne aussi parfois comme critère possible que les clitiques peuvent connaître des variations de positionnement par rapport à leur base de rattachement qu'ignorent en principe les affixes. Ce critère est notamment régulièrement mis en avant par les linguistes francisants qui se refusent à analyser les pronoms personnels conjoints du français comme affixes et acceptent tout au plus d'en faire des clitiques, mais il n'est très intéressant dans la mesure où dans de très nombreuses langues, y compris parmi les mieux documentées, il y a consensus pour analyser comme affixes des morphèmes grammaticaux qui en fonction de divers facteurs peuvent apparaître tantôt comme préfixes, tantôt comme suffixes (c'est le cas par exemple des marques de personne du verbe en arabe, basque, géorgien, etc.).

Mais, sans que ceci remette en question l'intérêt de ces critères, il faut rester conscient du fait que leur application rigoureuse et systématique à une langue donnée peut révéler une situation plus complexe qu'une simple dichotomie entre une classe de morphèmes liés peu intégrés (clitiques) et une classe de morphèmes liés fortement intégrés (affixes). Et surtout, l'application rigoureuse et systématique de ces critères montre qu'il n'y a aucune cohérence dans les décisions prises d'une tradition grammaticale à une autre quant au traitement de formes grammaticales qui ne sont, ni des mots tout à fait normaux, ni des affixes tout à fait typiques. Ceci se reflète d'ailleurs dans les conventions orthographiques, dont il n'est pas exceptionnel qu'elles fassent apparaître comme mots à part des formes liées suffisamment intégrées pour être analysées comme affixes, alors que par ailleurs il leur arrive d'enregistrer le rattachement de formes liées qui sont de manière évidente la variante cliticisée d'un mot existant par ailleurs sous forme libre.

On peut notamment reprendre ici le cas des pronoms personnels conjoints du français. Les grammaires françaises répartissent en effet en deux sous-ensembles les pronoms personnels. Il ne fait de doute pour personne que les pronoms personnels disjoints sont des mots autonomes qui ont une distribution syntaxique de constituants nominaux. Quant aux pronoms personnels conjoints, ils s'écrivent la plupart du temps comme des mots à part, mais il est facile de montrer qu'à la différence des pronoms disjoints, ils ne sont pas dans une relation de commutation avec les constituants nominaux. A la différence des pronoms disjoints, les pronoms conjoints sont des formes liées, comme le prouve par exemple l'impossibilité qu'il y a à les coordonner –ex. (24)– ou à les séparer du verbe par une incise –ex. (25).

(24) a. Je te connais

b. Je le connais

c. \*Je te et le connais

d. Je vous connais, toi et lui

(25) a. Jean, à ce que j'ai entendu dire, a l'intention de changer de voiture

b. \*Il, à ce que j'ai entendu dire, a l'intention de changer de voiture

On peut remarquer que les mêmes critères appliqués aux pronoms personnels de l'anglais permettent de constater qu'il s'agit dans tous les cas de formes libres, contrairement à ce qui se passe en français, où seuls les pronoms personnels disjoints sont des formes libres, et où les pronoms personnels conjoints ont très nettement des propriétés de formes liées – ex. (26), à comparer avec (24) et (25).

(26) a. You also have a house in Paris

b. You and I know what should be done

Le caractère lié des pronoms conjoints se manifeste aussi par le fait qu'ils donnent lieu à des phénomènes morphophonologiques qui n'ont pas leur équivalent lorsque la base verbale est immédiatement précédée d'autre chose que d'un pronom conjoint : liaison obligatoire, possibilité d'élision du *u* de *tu* devant voyelle, etc.

Mais les descriptions courantes du français reconnaissent tout au plus aux pronoms conjoints le statut de clitiques, alors qu'en poursuivant l'analyse, on peut montrer que beaucoup de leurs propriétés sont des propriétés généralement considérées comme typiques des affixes, et qu'aucune de leurs propriétés ne s'oppose réellement à ce qu'on leur reconnaisse le statut d'affixes.

Le point crucial est que les pronoms conjoints ont totalement perdu en français contemporain la liberté de sélection d'une base qui caractérise normalement les formes liées identifiées comme clitiques : ils s'attachent forcément au verbe, ce qui est un argument essentiel pour les considérer comme affixes. Un autre argument important est la tendance du français parlé à rendre obligatoire l'utilisation des pronoms conjoints sujets : les morphèmes liés qui accompagnent obligatoirement un type donné de base s'analysent en principe comme affixes flexionnels, et leur identification comme clitiques est en contradiction avec le principe selon lequel la base à laquelle s'attache un clitique doit être un mot déjà constitué. Toutefois, pour des raisons qui tiennent largement au poids de la tradition, la reconnaissance des pronoms personnels conjoints comme affixes verbaux n'est admise que par une partie des linguistes travaillant sur le français.

#### 6.4. Syntagmes et composés

Le terme de composition s'utilise couramment en linguistique avec deux sens qui ne se confondent pas, bien qu'il puisse y avoir entre eux une relation, en particulier du point de vue diachronique :

—on peut parler de *composition sémantique* lorsqu’une combinaison de mots pleins qui a l’apparence d’un syntagme formé selon les règles productives de construction de syntagmes est régulièrement utilisée avec un signifié qu’on ne peut pas prédire en appliquant simplement les règles sémantiques qui en principe permettent de calculer le signifié d’un syntagme à partir du signifié des mots qui le constituent —ex. (27) ;

- (27) a. gros bonnet ‘personnage important ou influent’  
b. avoir du pain sur la planche ‘avoir beaucoup de travail à faire’  
c. rester les bras croisés ‘ne pas intervenir’

—on peut parler de *composition morphologique* lorsque deux mots pleins forment une combinaison qui a globalement (notamment en ce qui concerne les possibilités de variation des marques flexionnelles et les possibilités d’adjonction d’expansions) les caractéristiques d’un mot unique —ex. (28).

- (28) a. bébé-éprouvette, chou-fleur (N + N → N)  
b. coffre-fort, rond-point (A + N ou N + A → N)  
c. aide-mémoire, brise-glace (V + N → N)  
d. va-et-vient, cache-cache, savoir-faire (V + V → N)  
e. maintenir (N + V → V)

La relation entre composition sémantique et composition morphologique est que :

—les constituants de syntagmes régulièrement utilisés avec un signifié qui ne peut pas être prédit par la simple application des règles qui permettent normalement de calculer le signifié d’un syntagme tendent à perdre la relative autonomie qu’ont normalement les constituants de syntagmes, et à se comporter de plus en plus comme s’ils constituaient les formants d’un mot unique ;

—les mots formés par composition morphologique tendent à se spécialiser dans l’expression de signifiés qui ne peuvent pas se calculer automatiquement à partir du signifié de leurs éléments.

Toutefois, cette relation entre composition sémantique et composition morphologique n’est pas absolue, et les deux notions doivent être traitées comme distinctes, car :

—d’une part le figement d’un groupe de mots régulièrement utilisé avec un signifié qui ne peut pas se prédire à partir du signifié des mots qui le forment ne va pas nécessairement jusqu’au stade où le groupe de mots peut se réanalyser comme un mot unique ;

—d’autre part beaucoup de langues (l’allemand par exemple) ont des mécanismes hautement productifs de formation de composés morphologiques (et notamment de noms composés), ce qui n’est guère concevable sans un certain degré de ‘transparence sémantique’ de la majorité de tels composés.