

# Les unités minimales de la syntaxe et de la sémantique : le cas du français

(version corrigée 16/07/08)<sup>0</sup>

Sylvain Kahane  
Modyco, Université Paris 10 & CNRS / Alpage, INRIA  
sylvain@kahane.fr, www.kahane.fr

## 1 Introduction

« De même que le jeu d'échecs est tout entier dans la combinaison des différentes pièces, de même la langue a le caractère d'un système basé complètement sur l'opposition de ses unités concrètes. On ne peut ni se dispenser de les connaître, ni faire un pas sans recourir à elles ; et pourtant leur délimitation est un problème si délicat qu'on se demande si elles sont réellement données. » (Saussure 1916 : 149)

Cet article a comme objectif de caractériser les unités de la langue, en particulier les unités sémantiques et syntaxiques. Le grand problème de la caractérisation de ces unités vient de la non correspondance entre les unités minimales de forme et les unités minimales de sens (par exemple dans *Il dort à poings fermés*, le segment<sup>1</sup> *à poings fermés* est une unité minimale de sens, exprimant l'intensification du verbe DORMIR, et cette unité est construite en utilisant d'autres unités, comme le nom POING, qui dans d'autres contextes sont elles-mêmes des unités de sens). La caractérisation des unités de la syntaxe pose un problème supplémentaire par le fait qu'elles ne se confondent ni avec les unités minimales de forme, ni avec les unités minimales de sens.

La caractérisation des unités de la langue revêt nécessairement un caractère universel. Nous nous intéresserons pourtant ici uniquement au français. Deux raisons au moins permettent de justifier ce choix. La première est que ces unités jouent un rôle central dans le système de chaque langue et que donc l'étude de chaque langue en particulier permet à notre avis de dégager et de caractériser ces unités. Nous pensons que c'est en dégageant pour chaque langue les unités qui sous-tendent sa structure propre et en les comparant avec celles qui sous-tendent la structure de chacune des autres langues que nous pouvons montrer leur universalité. La deuxième raison est que, à travers la caractérisation de ces unités pour le français, on est nécessairement amené à considérer des phénomènes particulièrement intéressants de la syntaxe du français.

## 2 Commutation libre

Le principe de commutation permet d'identifier les unités minimales. Pour reprendre Saussure (1916 : 117), un mot comme *enseignement* entretient des rapports associatifs avec *enseigner*, *renseigner*, etc. d'une part, et *armement*, *changement*, etc. d'autre part. Toujours selon Saussure, il existe un rapport de proportionnalité sémantique *enseignement* : *enseigner* = *changement* : *changer*. Ceci a pu ensuite être reformulé en termes de **commutation** : dans *enseignement*, *enseign(er)* peut **commuter** avec *arm(er)* ou *chang(er)* pour donner *armement* ou *changement*. Autrement dit, *enseignement* se décompose en deux unités *enseign-* + *-ment*. L'impossibilité de décomposer à nouveau ces deux segments — *enseign-* et *-ment* — conduit à les considérer comme des unités minimales de forme.

Nous appellerons les unités minimales de forme les **morphèmes** ou **monèmes**. Un morphème/monème est une collection<sup>2</sup> maximale de signes linguistiques de sens et de formes apparentées qui dans aucun contexte ne peuvent être la combinaison de deux signes.

Nous choisissons de conserver, à côté du terme aujourd'hui standard de *morphème*<sup>3</sup>, le terme *monème* introduit par Martinet 1960. La principale raison de notre choix est qu'il nous permet de construire divers termes que nous utiliserons, comme *mone*, *allomone* et *polymone*. Selon que l'unité minimale de forme est découpée selon la forme, en *morphe*, ou selon la distribution, en *mones*, nous privilierons l'un ou

l'autre terme. En effet, un morphème/monème ne désigne pas un signe linguistique au sens propre (c'est-à-dire l'association d'un sens et d'une forme), mais une collection de signes de formes et de sens apparentés. Ainsi dans *nous enseignons*, *un enseignant* ou *il se renseigne*, nous considérons qu'il s'agit du même morphème *enseign-* en raison de l'identité de forme et de la parenté de sens de chacune de ces occurrences de *enseign-* (par contre, il n'y a pas en synchronie de parenté de sens entre ces occurrences de *enseign-* et celle de *l'enseigne d'un magasin*, qui forme donc un morphème séparé, selon notre terminologie). Ces différentes occurrences de *enseign-* possèdent des distributions très différentes (la première se combine librement avec des flexions verbales, la deuxième est liée à *-ant* et la troisième à *re-*) et donc forment chacune une portion distincte du morphème/monème, que nous appellerons un mone.

Un **mone** est une sous-collection maximale de signes d'un monème distributionnellement compatibles.

Si maintenant nous comparons *enseignement* avec la forme verbale *enseignons*, la combinaison du morphème *enseign-* avec le suffixe à l'intérieur de ces deux expressions est, comme chacun sait, de nature différente : *-ment* est un suffixe dérivationnel et *-ons* un suffixe flexionnel. Cette différence peut être caractérisée par les simples propriétés de la commutation. En effet, dans *enseign+ons*, *enseign-* commute avec tous les verbes (*march-*, *dorm-*, etc.) et *-ons* commute avec toutes les flexions verbales (*-ez*, *-ions*, etc.) et toutes les combinaisons qui en résultent (*marchez*, *marchions*, *dormez*, *dormions*, etc.) sont valides<sup>4</sup>. On dira dans un tel cas que *enseign-* et *-ons* se combinent **librement**. Contrairement à *enseign+ons*, la combinaison *enseigne+ment* est **liée** : en effet, l'ensemble des radicaux qui peuvent se combiner avec ce morphème *-ment* (*chang-*, *comport-*, *avanc-*, mais pas *apprend-*, *march-*, *dorm-*, etc.) ne correspond à la distribution d'aucun autre signe (c'est-à-dire qu'il n'existe aucun suffixe qui se combine avec exactement le même ensemble de radicaux).

Commençons par préciser ce qu'on entend par commutation. Une commutation n'est légitime que si deux conditions sont remplies.

On dit que **A' commute proprement** avec A dans la combinaison A+B si : 1) A et A' s'excluent mutuellement, c'est-à-dire que B ne peut se combiner simultanément avec A et A' et 2) l'interprétation de B n'est pas modifiée par la commutation, c'est-à-dire que le rapport du sens 'A'+B' sur 'A' est le même que le rapport du sens 'A+B' sur 'A' ; en particulier, si A' est synonyme de A, A'+B doit être synonyme de A+B.

Nous pouvons maintenant définir la **commutation libre**.

On dit que **A commute librement** dans la combinaison A+B si 1) l'ensemble des éléments qui commutent proprement avec A est assez régulier et peut notamment se déduire des ensembles d'éléments qui commutent avec A dans d'autres combinaisons et 2) dans l'ensemble des éléments qui commutent proprement avec A dans la combinaison A+B, il existe une proportion importante d'éléments qui ont une distribution similaire à A.

On dit que **A et B se combinent librement** si A et B commutent l'un et l'autre librement dans la combinaison A+B.

Par exemple, dans l'énoncé *Le chat avançait*, chacun des quatre segments *le*, *chat*, *avanç-* et *-ait* commute librement, ainsi que les segments *le chat* et *avançait*, voire le segment *le chat avanç-*<sup>5</sup> :

<i>le</i>	<i>chat</i>	<i>avanç-</i>	<i>-ait</i>
<i>un</i>	<i>chien</i>	<i>grogn-</i>	<i>-e</i>
<i>mon</i>	<i>garçon</i>	<i>chant-</i>	<i>-era</i>
<i>ce</i>	<i>public</i>	<i>march-</i>	<i>-a</i>

Ces commutations sont libres, car chaque commutation est indépendante des autres et toutes les combinaisons qui en résultent sont valides : *mon chat chanta*, *le public grogne*, *ce garçon avancera*, etc.

La commutation libre permet de caractériser certains mones.

Nous appelons **yntaxème simple** toute sous-collection maximale de signes linguistiques d'un morphème qui commutent librement et possèdent une distribution syntaxique équivalente.

Les syntaxèmes qui appartiennent à des classes ouvertes (verbes, noms, adjectifs, adverbes, interjections) sont appelés des **lexèmes**. Les syntaxèmes qui appartiennent à des classes fermées et sont indissociables (voir définition plus loin) d'un lexème sont appelés des **yntaxèmes**, ou **morphèmes, flexionnels** (il se trouve que ces syntaxèmes sont toujours des unités minimales de forme et qu'il est d'usage de les appeler des morphèmes flexionnels). Les autres syntaxèmes (prépositions, conjonctions, pronoms, déterminants, etc.) sont appelés des **lexèmes grammaticaux**.

Ainsi sous le morphème *avanç-* /avãs/, on reconnaît deux syntaxèmes : le lexème verbal *avanç-1* (que l'on note AVANCER) et le lexème nominal *avanç-2* (qui est le nom AVANCE)<sup>6</sup>. Les deux syntaxèmes regroupent bien des signes qui commutent librement. Et il s'agit bien de deux syntaxèmes distincts, car les occurrences de *avanç-1* doivent toujours se combiner avec une flexion verbale telle que *-ait*, tandis que les occurrences de *avanç-2* ne le peuvent jamais, et leurs environnements sont donc toujours incompatibles. Dans AVANCEMENT, on retrouve également le morphème *avanç-*, mais la combinaison entre *avanç-* et *-ment* n'est pas libre et il ne s'agit donc pas d'un syntaxème. Comme pour les autres affixes dérivationnels, le morphème *-ment* de dérivation verbo-nominale ne se combine jamais librement, car l'ensemble des radicaux verbaux avec lesquels il se combine est irrégulier. Les affixes dérivationnels ne sont jamais des syntaxèmes<sup>6bis</sup>. Nous les appelons des morphèmes sous-lexicaux, car ils forment toujours un sous-segment d'un lexème.

La notion de commutation est **graduelle** : plus l'ensemble des éléments qui commutent proprement avec l'élément A dans une combinaison A+B est régulier et plus les éléments de cet ensemble ont une distribution similaire à A, plus on considère que A commute librement. Certains morphèmes sous-lexicaux possèdent une combinatoire très libre qui les rapproche des morphèmes flexionnels. C'est le cas par exemple du morphème *-ment* qui transforme un adjectif en adverbe. Dans de nombreux cas, ce changement de partie du discours ne s'accompagne d'aucun changement de sens (cf. le passage de *un départ rapide à partir rapidement* ou de *une réflexion intense à réfléchir intensément*) et le morphème commute proprement avec *de façon Adj* : 'réfléchir intensément' ≈ 'réfléchir de façon intense'. Mais la dérivation en *-ment* n'est pas systématiquement possible (cf *une réflexion poussée/élaborée* vs *\*réfléchir poussé/élaboré*, *un départ inattendu* vs *\*partir inattendument*) et la distribution de *-ment* est donc irrégulière. Même des adjectifs très courants comme GROS n'ont pas d'adverbe correspondant : *faire une grosse erreur* vs *\*se tromper grossement*. Certains adverbes en *-ment* ne sont pas sémantiquement compositionnels comme *vraiment*, *carrément* ou *vertement*. La combinatoire de *-ment* n'est donc pas suffisamment libre pour le considérer comme un syntaxème.

L'opposition entre combinaison libre et liée marque la frontière entre ce que Saussure nomme la langue et la parole. Les **combinaisons libres** font partie de la **parole**, c'est-à-dire des productions que chaque locuteur est libre de créer à sa guise. Les **combinaisons liées**, au contraire, font partie de la **langue**, c'est-à-dire des connaissances partagées par les locuteurs d'une même langue et qui constituent leur stock lexical. Saussure ne dit pas autre chose dans l'extrait suivant (même si la notion de combinaison libre et liée n'est pas définie formellement comme ici) :

« **Le propre de la parole, c'est la liberté des combinaisons.** [C'est nous qui soulignons]

On rencontre d'abord un grand nombre d'expressions qui appartiennent à la langue ; ce sont les locutions toutes faites, auxquelles l'usage interdit de rien changer, même si on peut distinguer, à la réflexion, des parties significatives.

[...] Mais ce n'est pas tout ; il faut attribuer à la langue, non à la parole, tous les types de syntagmes construits sur des formes régulières. [...] Quand un mot comme *indécorable* surgit dans la parole, il suppose un type déterminé, et celui-ci à son tour n'est possible que par le souvenir d'un nombre suffisant de mots semblables appartenant à la langue (*impardonnable*, *intolérable*, *infatigable*, etc.). Il en est exactement de même des phrases et des groupes de mots établis sur des patrons réguliers ; des combinaisons comme *la terre tourne, que vous dit-il ?*, etc. répondent à des types généraux, qui ont à leur tour leur support dans la langue sous forme de souvenirs concrets.

Mais il faut reconnaître que dans le domaine du syntagme, il n'y a pas de limite tranchée entre le fait de langue, marque de l'usage collectif, et le fait de parole, qui dépend de la liberté individuelle. Dans une foule de cas, il est difficile de classer une combinaison d'unités, parce que l'un et l'autre facteurs ont concouru à la produire, et dans des proportions qu'il est impossible de déterminer. » (Saussure 1916 : 172)

Un syntaxème peut être associé à plusieurs sens et à plusieurs formes : le syntaxème n'est donc pas un signe linguistique à proprement parler, lequel associe un sens à une forme, mais un **faisceau de signes**<sup>7</sup>, c'est-à-dire une association entre un ensemble de sens et un ensemble de formes, où chaque association entre un sens particulier et une forme particulière du syntaxème est possible et forme un signe linguistique. Par exemple, le lexème ALLER possède plusieurs acceptations parmi lesquelles 'aller quelque part', 'aller bien/mal' ou encore son sens de futur et il possède plusieurs formes : *all-*, *v-*, *i-* et *aill-*. Les deux faces, le sens et la forme, sont totalement **indépendantes** : chaque sens peut s'associer avec n'importe quelle forme et chaque forme avec n'importe quel sens. Le syntaxème se découpe selon les deux plans : le découpage selon le plan de la forme donne les **morphes**, le découpage selon le plan du sens donne les **mônes**. Mais la symétrie entre sens et forme n'est pas complète. Le locuteur fait librement le choix d'un sens et ce choix peut se porter sur n'importe laquelle des acceptations du syntaxème. Par contre, le « choix » du signifiant n'en est pas un : il est totalement imposé par l'environnement, c'est-à-dire par les choix adjacents du locuteur. Par exemple, si le locuteur veut parler d'un de ses prochains déplacements, il pourra choisir le faisceau ALLER et produire un énoncé comme « *Je vais à Paris la semaine prochaine.* » Le choix de la forme *v-* du faisceau lui est imposé par le contexte : ce sont les sens 'moi', réalisé par *je*, et 'présent' qui vont sélectionner l'allomorphe *v-* parmi les différents morphes de ALLER. Ce « choix » du signifiant est totalement indépendant du sens par lequel le locuteur a sélectionné le faisceau ALLER.

Les unités de la langue, celles qui sont dans notre cerveau, ne sont donc pas des signes, mais sont plutôt des faisceaux de correspondances entre sens et formes. Lorsqu'un locuteur produit un énoncé, il choisit de tels faisceaux qu'il assemble. Un faisceau donné est choisi parce que l'un des sens du faisceau est un sens que le locuteur veut exprimer. Ce ne sont pas les signes qui sont organisés en faisceaux, mais **les faisceaux qui se réalisent par des signes en parole**. Lorsqu'un faisceau est utilisé dans un énoncé, c'est forcément qu'il a été choisi pour un de ses sens et qu'il est réalisé sous une de ses formes. Donc dans un énoncé, le faisceau apparaît sous la forme d'un signe. Mais dans notre cerveau, c'est bien un faisceau qui a été sélectionné et c'est par l'intermédiaire de ce faisceau qu'un signe a été réalisé.

### 3 Unités significatives

La combinaison libre implique la **compositionnalité** (sémantique), mais l'inverse n'est pas vrai : la combinaison *aspir+ateur* est compositionnelle dans la mesure où l'on peut considérer qu'il existe un signe *-ateur* qui, appliqué à un radical X, signifie 'appareil qui sert à X-er' comme dans *incinérateur* ou *ventilateur*. Autrement dit, il est possible de commuter proprement *aspir-* dans *aspirateur*. Mais cette combinaison n'est pas libre, car le suffixe *-ateur* s'applique à très peu de radicaux et des lexèmes comme *\*lavateur* ou *\*nettoyateur* ne sont pas attestés : l'ensemble des signes qui commutent avec *aspir-* est donc irrégulier.

Une combinaison comme *machine à laver* pourrait être considérée comme libre dans le sens où l'on peut commuter librement sur *machine* comme sur *laver* et construire des combinaisons telles que *appareil à laver* ou *machine à aspirer*. Mais en faisant ce type de commutations, on voit que le statut de la combinaison change : la combinaison *machine à laver*, lorsqu'elle est prise dans son sens idiomatique d'appareil électroménager servant à laver le linge ou la vaisselle, fonctionne comme un tout préconstruit, contrairement aux combinaisons libres *appareil à laver* ou *machine à aspirer*. Il s'agit d'une **combinaison figée**. Ainsi pour *machine à laver*, les Québécois disent *laveuse* et les Français disent aussi *lave-linge* et *lave-vaisselle*, mais à côté de *aspirateur*, nous ne dirons ni *machine à aspirer*, ni *aspire-poussière*. Si *machine à laver* et *laveuse* sont des « paraphrases » dont on préfère arbitrairement une, on peut aussi noter que les combinaisons choisies pour désigner un même objet peuvent exprimer des points de vue différents sur cet objet, comme l'illustre les « synonymes » *téléski*, *remonte-pente* et *tire-fesse*. A

l'inverse, des quasi « paraphrases » comme *ventilateur* et *moulin à vent* désignent des objets très différents (Anscombe 1990). C'est ce côté arbitraire du choix de la combinaison qui manifeste le caractère figé de ces combinaisons, même lorsqu'elles sont sémantiquement compositionnelles, et qui fait qu'on les considère en conséquence comme des combinaisons liées.

Ceci nous amène à la définition des unités minimales de sens, que nous appellerons les **unités significatives**. Comme nous venons de le voir, ce n'est pas la notion de combinaison liée qui permet de les définir (c'est parce que *machine à laver* est une unité significative que nous en déduisons qu'il s'agit d'une combinaison liée et pas l'inverse). Nous définirons les unités significatives à partir de la notion de **choix** introduite par Martinet (1960 : 26, chapitre intitulé *Chaque unité suppose un choix*) :

« Soit un énoncé comme *c'est une bonne bière /setynbonbier/*. [...] Si nous sommes à même de dire quelque chose sur les latitudes combinatoires de /bon/, c'est que ce segment de l'énoncé a été reconnu comme représentant une unité particulière distincte de /yn/ et de /bier/. Pour arriver à ce résultat, il a fallu constater que /bon/, dans ce contexte, correspondait à un **choix** spécifique entre un certain nombre d'épithètes possibles ; la comparaison d'autres énoncés français a montré que dans les contextes où figure /bon/ on trouve aussi /ekselāt/ (*excellente*), /movez/ (*mauvaise*), etc. Ceci indique que le locuteur a, plus ou moins consciemment, écarté tous les compétiteurs qui auraient pu figurer entre /yn/ et /bier/, mais qui ne se trouvaient pas convenir en l'occurrence. Dire de l'auditeur qu'il comprend le français implique qu'il identifie par expérience les choix successifs qu'à dû faire le locuteur, qu'il reconnaît /bon/ comme un choix distinct de /yn/ et de celui de /bier/, et qu'il n'est pas exclu que le choix de /bon/ au lieu de /movez/ influence son comportement. »

Contrairement à Martinet, nous dissocions complètement la notion de monème/morphème de celle d'unité significative. Par exemple, dans l'énoncé *Pierre mange une pomme de terre*, les mots *pomme*, *de* et *terre*, sont des morphèmes, mais aucun d'eux ne résulte d'un choix : *pomme* n'a pas été choisi par opposition à *poire* ou *banane*, *de* n'a pas été choisi par opposition à *ou* *dans*, et *terre* n'a pas été choisi par opposition à *eau* ou *feu*. C'est bien *pomme de terre* dans son intégralité qui a été choisi par opposition à *carotte*, *chou-fleur* ou *haricot vert*. Ainsi, *pomme de terre* est ici une unité significative et les morphèmes qui la composent n'en sont pas (dans cet énoncé).

Nous appelons **unité significative** tout signe qui résulte d'un choix unique et indivisible ; une unité significative commute généralement avec des syntaxèmes simples, mais aucune de ses composantes ne peut commuter avec des syntaxèmes simples sans changer fortement le sens.<sup>7bis</sup>

Prenons un autre exemple : l'énoncé *La moutarde me monte au nez* pris avec le sens ‘je sens la colère m'envahir’. La production de cet énoncé suppose quatre choix et cet énoncé contient donc quatre unités significatives. La première est *me*, choisie par opposition à *te*, *lui*, etc. Plus exactement, il s'agit du lexème MOI, la forme clitique *m+e* étant imposée par le prédicat verbal qui réalise son premier actant comme objet indirect ; dans une paraphrase comme *Je commence à être en colère*, MOI sera réalisé par la forme sujet *je*. La deuxième unité significative est la locution *la moutarde mont- au nez*, que nous notons ‘LA MOUTARDE MONTER AU NEZ’. Comme dans le cas de *pomme de terre*, aucun des morphèmes qui composent cette locution ne peut être commuté individuellement et le choix est unique et indivisible. La troisième unité significative est le présent, réalisé par *-e* (zéro à l'oral) et choisi par opposition à l'imparfait ou au futur (*La moutarde me montait au nez*, *La moutarde me montera au nez*). La quatrième unité significative est non segmentale : c'est la déclarativité. Notre énoncé s'oppose ainsi à *La moutarde me monte-t-elle au nez ?* La déclarativité s'exprime ici par l'utilisation d'une proposition avec un verbe à l'indicatif et une courbe intonative descendante en fin de phrase, typique des énoncés déclaratifs.

Le cas des régimes mérite une discussion supplémentaire. Dans *Pierre parle à Marie*, nous considérons que la préposition *À* n'est pas une unité significative, mais fait partie du régime de PARLER. Autrement dit, la préposition *À* ne constitue pas un choix séparé de celui de PARLER : plus exactement, le choix fait ici est celui d'une acceptation particulière de PARLER qui se construit avec un complément d'objet indirect introduit par *À*, le verbe et sa construction formant un tout indissociable. La version extrême de cette position est de considérer que *À* est un **mot vide**. Le terme *mot vide* est dû à Lucien Tesnière (1959 : chapitre 28 — *Mots pleins et mots vides*) :

« 1. — Il y a deux espèces de mots essentiels, les mots **pleins** et les mots **vides**.

2. — Les mots **pleins** sont ceux qui sont **chargés d'une fonction sémantique**, c'est-à-dire ceux dont la forme est associée directement à une idée, qu'elle a pour fonction de représenter et d'évoquer. [...]

3. — Les mots **vides** sont ceux qui ne sont pas chargés d'une fonction sémantique. Ce sont de simples **outils grammaticaux** (en note : Damourette et Pichon disent très heureusement des « struments », d'autres ont proposé « mots-charnières ») dont le rôle est uniquement d'indiquer, de préciser ou de transformer la catégorie des mots pleins et de régler leurs rapports entre eux. »

La position inverse est au contraire de considérer que le PARLER de *Pierre parle à Marie* est le même que le PARLER de *Pierre parle en dormant* ou *Pierre parle anglais* et que la nuance de sens qu'il y a entre ces trois occurrences de PARLER est due à la **construction** avec laquelle PARLER se combine (au sens des grammaires de construction ; cf. par ex. Goldberg 1995). Autrement dit, lorsque *Pierre parle à Marie*, le sens sous-jacent qui est que ‘Pierre s’adresse à Marie’ vient davantage de la construction « X V à Y », dans laquelle S’ADRESSER, PARLER, TELEPHONER ou ECRIRE peuvent occuper la position V, que du verbe qui occupe cette position. Bien sûr cette construction signifiant en elle-même ‘s’adresser à quelqu’un’ n’est compatible qu’avec des verbes du type S’ADRESSER et il est donc difficile de savoir d’où vient exactement la contribution sémantique. En admettant que PARLER se combine avec une construction comme il se combine avec un syntaxème, il reste que la combinaison d’une lexie et d’une construction est une combinaison liée : par exemple, si on reprend les trois verbes ECRIRE, PARLER et TELEPHONER qui se combinent tous avec « X V à Y », on voit qu’on peut dire *parler anglais*, mais pas *\*téléphoner anglais*, ni *\*écrire anglais* (on dit *écrire en anglais*). On peut dire *écrire quelque chose à quelqu’un*, mais pas *\*parler quelque chose à quelqu’un*. Par contre, on peut *parler de quelque chose*, mais pas *\*écrire de quelque chose* ou *\*téléphoner de quelque chose*. En conclusion, il peut être envisageable de considérer la construction du verbe comme un signe à part entière. On peut même considérer que cette construction est elle-même la combinaison de plusieurs signes, chacun correspondant à la relation entre le verbe et un de ses actants. De ce point de vue, la préposition n’est donc pas ici un mot vide mais le signifiant d’un signe indiquant une relation de destinataire entre l’action décrite par le verbe et son dépendant : dans *Pierre parle à Marie*, Marie est le destinataire des paroles de Pierre. Néanmoins la combinaison entre le verbe et sa construction est une combinaison liée et les deux forment ensemble une seule unité significative.

Il est assez courant qu’une unité significative soit réalisée par une combinaison de mones ; il s’agit même de la majorité des unités significatives.

Une unité significative composée de plusieurs mones sera appelée un **polymone**.

Nous préférons ce terme (à ne pas confondre avec le terme *polynôme* des mathématiques) à ceux proposés respectivement par Martinet et Mel’čuk de *synthème* et de *phrasème*, car ici le suffixe *-ème* ne se justifie pas du tout. Un polymone est une unité significative dont le signifiant est formé de plusieurs mones et il nous semble que le terme *polymone* est de loin le plus cohérent. Le terme *synapsie* proposé par Benveniste est également cohérent avec nos notations.

La présence de polymones dans la langue complexifie le système, puisqu'il n'y a plus correspondance systématique entre les segments minimaux du point de vue du sens (les unités significatives) et les segments minimaux du point de vue de la forme (les morphèmes). Mais d'un autre côté, l'utilisation de polymones offre une certaine économie en réduisant le caractère **arbitraire** du lien entre la forme et la sens. Ainsi la lecture littérale d'un signe tel que *fer à repasser* comme ‘objet en fer qui sert à repasser’ n'est pas très éloigné du sens exact de la locution et le sens de la locution est donc relativement **compositionnel** et **transparent**. Cette relative **motivation** du signifiant des polymones a été bien dégagée par Saussure (1916 : 180) :

« Le principe fondamental de l’arbitraire du signe n’empêche pas de distinguer dans chaque langue ce qui est radicalement arbitraire, c'est-à-dire immotivé, de ce qui ne l'est que relativement. Une partie seulement des signes est absolument arbitraire ; chez d’autres intervient un phénomène qui permet de reconnaître des degrés dans l’arbitraire sans le supprimer : *le signe peut être relativement motivé*.

Ainsi *vingt* est immotivé, mais *dix-neuf* ne l'est pas au même degré, parce qu'il évoque des termes dont il se compose et d'autres qui lui sont associés, par exemple *dix*, *neuf*, *vingt-neuf*, *dix-huit*, *soixante-dix*, etc. ; pris séparément *dix* et *neuf* sont sur le même pied que *vingt*, mais *dix-neuf* présente un cas de motivation relative. Il en est de même de *poirier*, qui rappelle le mot simple *poire* et dont le suffixe *-ier* fait penser à *cerisier*, *pommier*, etc. ; pour *frêne*, *chêne*, etc., rien de semblable. »

Les polymones comprennent les lexèmes complexes, comme ENSEIGNE+MENT, et les locutions, mais aussi les polymones grammaticaux comme le conditionnel  $(-r+i-)$ , qui est un polymone dans la plupart de ses emplois (*Nous aimerions vous rencontrer*, *Il aurait tué sa femme*), ou encore le passé composé (*AVOIR + -é*). Les polymones peuvent avoir des signifiants surprenants, comme l'intensifieur *POUR ÊTRE X, ... EST X*<sup>1</sup> qui suppose une duplication de l'adjectif : *Pour être bizarre, elle est bizarre* 'Elle est vraiment très bizarre'. Certains polymones sont polymorphes : c'est bien sûr le cas lorsque les mones sont des syntaxèmes polymorphes comme *'S'EN ALLER'*, où le réfléchi (noté SE) comme ALLER ont de nombreux allomorphes : *je m'en vais, nous nous en irons*, etc. Mais il y a aussi des polymones qui ne s'expriment pas de manière univoque par une seule combinaison de mones : c'est le cas par exemple de *'CHIER/FAIRE DANS SON FROC/PANTALON'* 'manifester une peur intense' avec deux alternances possibles et donc quatre formes en tout. (Les délimiteurs  $'\dots'$  servent à délimiter les polymones, en particulier les locutions.)

Les unités significatives commutant avec des lexèmes sont appelées des **unités lexicales** ou **lexies**. Nous distinguons très nettement lexème et lexie. D'une part, une lexie peut être la combinaison de plusieurs syntaxèmes (alors que le lexème est par définition un syntaxème), mais surtout un lexème est potentiellement polysémique (alors qu'une lexie est monosémique par définition).

Les polymones ne doivent pas être confondus avec les collocations. Une combinaison libre est nécessairement la combinaison de deux unités significatives, mais l'inverse n'est pas vrai : il existe des combinaisons non libres qui résultent de plusieurs choix. Telles sont les collocations.

Une **collocation** est une combinaison liée de deux unités significatives.

Dans l'énoncé *Pierre a peur*, le choix de PEUR s'oppose à ceux de FAIM, FROID ou ENVIE, tandis que le choix de AVOIR s'oppose à ceux de PRENDRE ou FAIRE. Par contre, PEUR ne peut pas commuter dans cet énoncé avec ses synonymes CRAINTE ou EFFROI et à la place de *\*avoir crainte* ou de *\*avoir effroi*, on dira plutôt *craindre* ou *être effrayé*. De même, le verbe AVOIR ne peut pas commuter ici avec un autre verbe que PRENDRE ou FAIRE et la combinaison est donc liée.

Une collocation est toujours dissymétrique : l'un des choix est dépendant de l'autre. Par exemple, dans AVOIR + PEUR, c'est PEUR qui est **choisi librement**, tandis qu'AVOIR est choisi pour permettre à PEUR de jouer un rôle prédictif et ce **choix est dépendant** de PEUR, dans le sens où un autre nom de sentiment entraînerait un choix différent (cf. *être en colère*, *éprouver de la joie*, etc.). Des deux parties d'une collocation, on appelle **base** l'unité significative choisie librement et **collocatif** l'unité significative choisie en fonction de la première. Le collocatif étant choisi en fonction de la base, on peut modéliser les collocations par des fonctions appelées **fonctions lexicales** : à chaque sens qui s'exprime par des collocatifs, on fait correspondre une fonction lexicale qui associe à chaque unité lexicale les collocatifs exprimant ce sens (Mel'čuk *et al.* 1995). Remarquons que, dans une collocation, la base possède son sens usuel, tandis que le collocatif est généralement dénaturé et exprime avant tout le sens de la fonction lexicale. Il en résulte que si l'on prend une collocation comme *une colère aveugle*, on voit qu'il est facile de commuter le collocatif avec de nombreux éléments collocationnels ou non (*une colère froide*, *une colère sourde*, *une colère homérique*, etc.) et en particulier avec une expression périphrastique : *une colère qui s'exprime sans contrôle* ; il est par contre difficile de commuter proprement la base (*aveugle* se combine surtout avec des noms de sentiments proches de COLÈRE comme HAINES), bien que la fonction lexicale puisse se combiner avec d'autres noms de sentiment et donner d'autres valeurs comme *une peur panique*.

Une collocation A+B se distingue d'une locution par la possibilité de commutation sur le collocatif (comme on l'a vu ci-dessus pour *avoir peur* et *colère aveugle*) et/ou par la possibilité de modifier

indépendamment A et B (cf. *ne pas avoir peur* vs *avoir une belle peur*). La compositionnalité ne permet pas par contre de les distinguer. Comparons par exemple les expressions *noyer le poisson* et *avoir les boules* : à première vue, aucune des deux expressions n'est très transparente, si l'on s'en tient au sens usuel des différents lexèmes qu'elles contiennent. Pourtant, alors que la première ne permet aucune commutation propre, la deuxième permet des commutations propres sur AVOIR : *foutre les boules à qqn*, *C'est les boules !*, *Putain, les boules !*. Comme par ailleurs, *les boules* ne permet pas de commutation de *les* dans ce contexte (#*avoir une boule*, #*avoir des boules*), on en déduit que 'LES BOULES' est une unité significative et que AVOIR est un collocatif. Au final, *avoir les boules* apparaît comme compositionnelle, tandis que 'NOYER LE POISSON' est vraiment une locution.

Si la distinction entre locution et collocation est réelle<sup>8</sup>, la notion d'unité significative reste néanmoins insuffisamment définie à ce stade. La comparaison de la collocation *faire un pacte* avec le lexème *pact+iser* permet de s'en convaincre. Les deux sont tout aussi compositionnels (-*iser* est un suffixe causatif productif), on considère néanmoins que *faire un pacte* est la combinaison de deux unités significatives (où FAIRE est un collocatif de PACTE), tandis que PACTISER est une unique unité significative (cf. Aronoff 1974 ou Kerleroux 2004, pour qui le lexème est l'unité de la morphologie constructionnelle et le morphème sous-lexical n'existe pas en tant qu'unité significative). La principale raison est que *faire un pacte* est un syntagme, ce qui se traduit par le fait que les deux éléments de la combinaison ont une plus grande indépendance et peuvent par exemple être modifiés séparément. Nous développerons ce point dans la section suivante.

Nous allons conclure cette section en soulignant un point fondamental concernant les unités significatives et la syntaxe : il n'y a pas de mode spécifique de combinaison des unités significatives. Le comportement des différents mots d'une unité significative ne sera en général pas très différent du comportement de mots appartenant à des unités significatives séparées. Par exemple, si on compare l'unité significative 'BRISER LA GLACE' 'dissiper la gêne' avec la combinaison libre BRISER + GLACE, on voit que les deux combinaisons acceptent d'entrer dans des constructions analogues : négation (*Pierre n'a pas brisé la glace*), clivage du sujet (*C'est Pierre qui a brisé la glace le premier*) ou passif (*La glace a été brisée*).

Autrement dit, le fait pour deux syntaxèmes qui sont combinés de former ou non ensemble une unité significative n'a pas en général d'incidence sur les propriétés syntaxiques de leur combinaison. C'est ce qu'on nomme l'**autonomie de la syntaxe** par rapport à la sémantique.

Les unités significatives sont des unités sémantiques et pas des unités syntaxiques. Leur comportement syntaxique dépend des unités syntaxiques qui les composent et l'articulation de ces unités entre elles. De même, les morphèmes sont les unités de la morphologie constructionnelle et pas de la syntaxe. Morphèmes et unités significatives structurent le lexique de la langue : ce sont respectivement les plus petites et les plus grandes unités de première articulation. C'est le fait que les unités minimales du point de vue du signifiant (les morphèmes) ne correspondent pas aux unités minimales du point de vue du sens (les unités significatives) qui fait que ni les unes ni les autres ne sont les unités de la syntaxe. Nous allons donc pouvoir maintenant présenter les unités de la syntaxe et cette longue digression sur morphèmes et unités significatives permettra, nous l'espérons, de bien faire comprendre ce qui est et n'est pas une unité de la syntaxe.

## 4 Syntagme et syntaxème

Nous avons déjà présenté le **yntaxème simple** : il s'agit d'un morphème qui commute librement, c'est-à-dire qui est à la fois un morphème et une unité significative. C'est là situation idéale et plutôt rare, car de très nombreuses unités significatives sont des polytones. Il convient néanmoins de distinguer parmi les unités significatives, celles qui se comportent comme des syntaxèmes simples et celles qui se comportent plutôt comme une combinaison de syntaxèmes simples. Nous appelons les premières des syntaxèmes et les secondes des syntagmes. Nous allons illustrer notre propos par deux exemples avant de donner une définition.

Prenons un exemple de signe comme *brûlure*, qui se décompose en deux morphèmes *brûl-* + *-ure*. La combinaison possède une très grande cohésion, tant et si bien que les locuteurs peuvent ne pas avoir toujours conscience d'utiliser ici un polymone. Ces deux morphèmes sont linéairement inséparables (rien ne peut venir se glisser entre eux) et cette combinaison ne s'apparente à aucune combinaison libre de syntaxèmes. En effet, *brûl-* est ici apparenté au lexème verbal BRÛLER, mais aucun syntaxème ne peut se combiner librement à un verbe pour en faire un nom. Ce qui y ressemble le plus serait la combinaison d'un lexème verbal avec l'infinitif, qui est bien une combinaison libre de syntaxèmes, mais qui n'a pas exactement les propriétés distributionnelles d'un nom : la forme infinitive peut commuter avec un nom (*Pierre veut manger* vs *Pierre veut une banane*), mais elle n'accepte pas la flexion en nombre, ni les articles<sup>9</sup> et elle peut se combiner avec des pronoms clitiques ou régir un complément d'objet direct à la différence du nom. Par ailleurs, le morphème *-ure* ne peut jamais commuter librement. Il s'agit d'un **morphème fondamentalement lié**<sup>10</sup>. Une combinaison comme *brûl+ure*, qui ne s'apparente pas à une combinaison libre, est ce que nous appelons un **yntaxème complexe**. Le syntaxème complexe est un cas particulier de polymone.

Comparons avec une combinaison comme *pomme de terre*. Du fait qu'il s'agit d'une unité significative, ses composants ne se combinent pas librement et il est également impossible de les séparer. Pourtant, nous ne souhaitons pas considérer *pomme de terre* comme un mot (ou un syntaxème), car *pomme de terre* a la même structure que par exemple *part de gâteau*, qui lui est une combinaison libre et est séparable (cf. *une part énorme de ce délicieux gâteau*). En effet, *pomme de terre* est construit de manière **analogue** à *part de gâteau*, puisque *pomme* appartient à la même classe distributionnelle que *part* (celle des noms), *terre* à la même que *gâteau* (celle des noms encore) et le complexe lui-même, *pomme de terre* a la même distribution que *part de gâteau* (celle des noms toujours). Comme *part de gâteau* est clairement une combinaison libre de syntaxèmes (*part/tranche/morceau/... de gâteau/tarte/rôti/...*), nous considérons que *pomme de terre* est également une combinaison de syntaxèmes, par analogie. Le même raisonnement pourra s'appliquer à la locution *donner sa langue au chat* qui est construite de manière analogue à la combinaison libre *donner son devoir au professeur* et est donc une combinaison de syntaxèmes et pas un syntaxème complexe.

La définition du syntaxème repose sur la notion d'**analogie structurelle**.

Une combinaison A+B est dite (**structurellement**) **analogue** à une combinaison A'+B' si A a une distribution équivalente à A', B à B' et A+B à A'+B'.

Cette définition suppose que non seulement les composants de la combinaison sont distributionnellement équivalents, mais que le résultat est de même nature. Par exemple, dans *Pierre lave le linge* et *Pierre a rempli le lave-linge*, on peut admettre que *lave* et *linge* sont les mêmes morphèmes (et sont donc par conséquent distributionnellement équivalents), mais les combinaisons *lave le linge* et *lave-linge* n'ont pas du tout la même distribution (la première commute avec un verbe et la deuxième avec un nom) et donc les combinaisons ne sont pas analogues. De même, dans *Pierre met un livre en terre* et *Pierre enterre un livre*, nous avons une combinaison *en + terre* avec les mêmes morphèmes, mais les combinés possèdent des distributions totalement différentes (le premier est locatif et le deuxième est un verbe) et on ne peut donc pas considérer qu'il s'agit de combinaisons analogues.

Nous pouvons maintenant définir **syntagmes** et **yntaxèmes**.

Nous définissons le **syntagme** comme une combinaison analogue à une combinaison libre. Une unité significative qui est aussi un **syntagme** est appelée une **locution**.

Notre définition du **syntagme** est compatible avec celle de Saussure (1916), qui se limite au paragraphe suivant (p. 170) :

« Dans le discours, les mots contractent entre eux, en vertu de leur enchaînement, des rapports fondés sur le caractère linéaire de la langue, qui exclut la possibilité de prononcer deux éléments à la fois. Ceux-ci se rangent les uns à la suite des autres sur la chaîne de la parole. Ces combinaisons qui ont pour support l'étendue peuvent être appelées **syntagmes**. Le syntagme se compose donc toujours de deux ou plusieurs unités consécutives (par exemple : *re-lire* ; *contre tous* ; *la vie humaine* ; *Dieu est*

*bon ; s'il fait beau temps, nous sortirons, etc.). Dans un syntagme, un terme n'acquiert sa valeur que parce qu'il est opposé à ce qui précède ou ce qui suit, ou à tous les deux. »*

La dernière phrase semble suggérer que les composantes d'un syntagme permettent la commutation. Saussure donne un seul exemple de syntagme qui n'est pas clairement une combinaison libre : *re-lire*. La grammaire traditionnelle considère généralement *re-* comme un préfixe, c'est-à-dire un morphème lié. Si *re-* est un préfixe en français, il s'agit alors du plus libre d'entre eux. On peut le combiner avec tous les verbes pleins et même avec des locutions (*'il a renoyé le poisson* ; *'il a repris le taureau par les cornes*), le dupliquer (*rerefaire, rerelire*) et l'on trouve des erreurs de production comme *'Je revais lui dire* au lieu de *Je vais lui redire*. Il se combine également avec des noms en emploi prédictif : *il a re-envie de venir, Re-bonjour !* Il s'agit donc plutôt d'un clitique préverbal et *re-lire* est alors bien un syntagme.

Notons deux différences entre notre usage du terme syntagme et d'autres usages. D'une part, nous considérons que l'unité de base de la syntaxe est le syntaxème et nous appelons bien syntagme toute combinaison de syntaxèmes, et pas seulement les combinaisons de mots. En particulier, nous considérons qu'un mot comme *avançait* qui combine librement au moins deux syntaxèmes est un syntagme. Par contre, le signe *décomposition* (qui combine plusieurs morphèmes) est un syntaxème complexe, mais pas un syntagme. D'autre part, nous ne considérons pas, à la différence de l'école anglo-saxonne (qui utilise le terme *phrase*, traduit en français par *syntagme*) qu'un syntagme doive être saturé, c'est-à-dire contenir tous les dépendants de chaque portion du syntagme. Ainsi dans *Pierre doit chercher sa montre*, nous considérons que *doit chercher* est un syntagme au même titre que *chercher sa montre* ou *Pierre doit*.

Un **syntaxème** est une collection maximale de signes 1) de formes et de sens apparentés possédant la même distribution et 2) maximaux parmi les signes analogues à aucun syntagme (c'est-à-dire analogues à aucune combinaison libre). Nous appelons **syntaxème simple** tout syntaxème appartenant à un morphème et **syntaxème complexe** un syntaxème polymorphe.

**Maximal** signifie qu'il n'existe aucun segment plus grand qui ne soit analogue à aucun syntagme. Autrement dit, toute combinaison de deux syntaxèmes est un syntagme et tout syntagme est une combinaison de plusieurs syntaxèmes.

En général, tout syntaxème commute librement dans certains environnements, mais il existe des cas extrêmes comme *FUR*, qui n'apparaît que dans la locution 'AU FUR ET À MESURE'<sup>111</sup>. Le fait que *au fur* soit assez clairement analogue à une combinaison libre, en raison notamment de la coordination avec *à mesure*, oblige à considérer *fur* comme un syntaxème, puisque tout segment plus large est analogue à un syntagme.

Un syntaxème peut avoir plusieurs acceptations, mais il possède une classe distributionnelle homogène par définition. Dans *Pierre a mis des dalles sur sa terrasse* et *Pierre dalle sa terrasse*, nous considérons qu'il s'agit d'un même morphème, mais de deux syntaxèmes différents, car, bien que leurs formes et leurs sens soient apparentés, leurs distributions sont incompatibles (le premier est un nom et le deuxième un verbe). Par contre, dans *Pierre a la dalle*, bien que 'AVOIR LA DALLE' soit une locution, nous considérons qu'il s'agit du même lexème nominal DALLE que dans *Pierre a mis des dalles sur sa terrasse*.

Si nous reprenons la paire *Pierre met un livre en terre* et *Pierre enterre un livre*, nous voyons que dans les deux cas il s'agit d'une combinaison préposition + nom, mais la première 1) possède la distribution normale d'une combinaison libre préposition + nom et commute par exemple avec *Pierre met un livre sur le sol* ou *sous la table* et 2) est séparable (*Pierre met un livre en pleine terre*), tandis que la deuxième se comporte comme un verbe et est absolument inséparable (même avec un clitique ou un auxiliaire : *Pierre l'enterre* ; *Pierre a enterré un livre*). La première est donc un syntagme et la deuxième un syntaxème.

De la même façon, *lave-linge* dans *Pierre a rempli le lave-linge* est un syntaxème, puisque *lave* et *linge* sont linéairement inséparables et qu'il n'existe aucune combinaison libre verbe + nom qui se comporte comme un nom. La construction illustrée par *lave-linge* est très productive (cf. la très longue liste de composés VN en annexe de Villoing 2002), mais elle n'est pas libre, car on ne peut pas faire commuter les morphèmes lexicaux avec des synonymes : \**nettoie-linge*, \**lave-tissu*, etc.

Nos deux derniers exemples de lexèmes (*enterrer* et *lave-linge*), tout en étant bien des syntaxèmes, c'est-à-dire en n'étant analogue à aucune combinaison libre, laissent apparaître une structure syntagmatique figée, comme *un rendez-vous*, *un va et vient*, (*une idée*) *à la mors-moi le nœud*, *un je-m'en-foutiste*, *je ne sais* (*quelle idée*), *il embarque*, *il atterrit*, *une bonne femme*, *parce que*, etc. Dans chacun de ces lexèmes, on reconnaît une structure syntaxique (*il va et vient*, *je m'en fous*, *en barque*, *à terre*, etc.), mais la combinaison ne leur est pas analogue, car elle ne possède pas la distribution qu'aurait le syntagme d'origine. Nous appellerons de tels syntaxèmes des syntaxèmes syntagmatiques.

Un **syntaxème syntagmatique** est une combinaison A+B qui n'est analogue à aucun syntagme, mais qui est la version figée d'un syntagme A+B ; autrement dit, il existe des combinaisons libres A'+B', où A et A' comme B et B' sont de distributions équivalentes, mais pas A+B et A'+B'. Si A est un syntaxème dans la combinaison libre A+B, il est appelé un **quasi-syntaxème** quand il apparaît dans le syntaxème syntagmatique A+B.

Les composants d'un syntaxème syntagmatique sont fortement ressentis par les locuteurs : dans *lave-vaisselle*, la proximité du segment *vaisselle* avec le lexème *vaisselle* est très importante et justifie de parler de quasi-lexème. C'est moins clair lorsque l'un des éléments correspond à un lexème grammatical comme dans *en+terrer*. Et encore moins quand la combinaison liée n'est plus compositionnelle comme dans *ce+pendant*.

Il existe également des cas limites de combinaisons totalement atypiques, comme *à qui mieux mieux* ou *cucul la praline*, qui semblent n'obéir à aucun des procédés de construction de la syntaxe ou de la morphologie, mais où on reconnaît des combinaisons d'éléments lexicaux, ou encore des combinaisons qui attestent de constructions syntaxiques disparues comme l'ordre objet-verbe possible en ancien français : (*aller*) *tambour battant, sans coup férir, maintenir, ce faisant*.

L'exemple le plus problématique en français de construction à la frontière de la morphologie et de la syntaxe est certainement la construction NN. D'un côté, il existe des combinaisons fondamentalement liées comme la **construction NN coordonnée**, qui associe deux N de manière assez symétrique : *un chien-loup*, *un enseignant-chercheur*, *une moissonneuse-batteuse*, *un hôtel-restaurant*, *une fille-mère*, *un enfant-martyr*, etc. Ces combinaisons sont bien liées : si on a *un canapé-lit*, on n'a pas *\*un fauteuil-lit* ou *\*un canapé-couchette* ; si on a *la physique-chimie*, on n'a pas *\*la chimie-physique*. De l'autre côté, il existe des combinaisons parfaitement libres, comme la **construction NN nominative**, qui associe un grand nombre de noms communs avec n'importe quel nom propre : *le docteur Mabuse*, *le général Lee*, *les frères Coen*, *l'avenue Victor Hugo*, *la bibliothèque François Mitterrand*, *les usines Renault*, *la station Châtelet*, *l'affaire Dreyfus*, jusqu'à des constructions où le nom propre désigne une époque comme *une table Louis XV*. Certaines paraissent plus figées : à côté de *la région Auvergne*, on n'a pas *\*le département Cantal* ou *\*la ville Saint-Flour*.

Entre les deux, on trouve d'autres constructions NN plus ou moins libres. Commençons par la **construction NN modificative**, qui apparaît comme la réduction d'une construction N Prép N (Fradin, à paraître) : *un accès pompiers*, *une borne incendie*, *une manif étudiants*, *un fauteuil relax* (= pour la relaxation), *des pommes vapeur*, *un steak frites*, *un coin fumeurs*. Dans cette construction dissymétrique, l'un des deux noms va potentiellement pouvoir se libérer : par exemple, à côté de *accès pompiers*, on trouvera *accès handicapés*, *accès visiteurs*, *accès personnel* (= du personnel) et donc le nom *accès* devient ainsi un nom susceptible de régir un N nu, et ce régime est également possible pour des noms sémantiquement similaires comme *entrée* ou *porte*. De même des noms comme *espace* ou *coin* régissent potentiellement un N nu : *coin repas*, *espace repos*, *coin fumeurs*, *espace enfants*, *coin télé*, etc. A noter qu'on est passé d'une construction modificative où le N2 est un modifieur à une construction où le N1 est recteur. Mais il y aussi des cas où c'est le N2 qui devient productif : par exemple à côté d'*accès handicapés*, on a aussi *un fauteuil handicapés*, *une rampe handicapés*, *un ascenseur handicapés*, etc. Pourtant même lorsque le N2 devient un modifieur assez libre comme *maison* dans *une confiture maison* ou *une tarte aux pommes maison*, on voit que la construction reste confinée à un N particulier et on n'a pas *\*une confiture restaurant* ou *\*une confiture usine*.

Les constructions NN coordonnées ne sont pas toujours symétriques : *un poisson-chat* est un poisson qui a l'allure d'un chat et pas l'inverse. Fonctionnent de manière similaire *un requin-marteau*, *une guerre éclair*, *une justice escargot* ou *un discours fleuve*. Fradin (à paraître) les nomme des constructions NN subordonnées, mais il nous semble qu'elle relève du même procédé de construction que les autres NN coordonnées, avec seulement un usage métaphorique de N2 : un discours fleuve est un discours au sens propre et un fleuve au sens figuré. Dans les **constructions NN coordonnées dissymétriques**, N2 peut également devenir assez libre : *satellite espion*, *discours fleuve*, *personnage clé*, *grammaire jouet*, *un livre/film culte/phare/événement*.

La frontière entre les différentes constructions NN n'est pas toujours évidente. On peut penser que c'est le NN coordonné *stylo-plume* qui va faciliter le passage de *stylo à bille* au NN modificatif *stylo-bille*. Ou que c'est le NN nominatif *rayon laser* qui va permettre le passage de *imprimante à laser* au NN modificatif *imprimante laser* ou au NN coordonné *sabre laser*.

Parmi les constructions qui deviennent totalement productives, outre la construction nominative dont nous avons parlé au début, citons la construction « viande légume », qui est un cas particulier de la construction NN modificative : *une saucisses frites*, *un steak salade*, *une truite pommes vapeur*, *un côte de porc haricots verts*. On a aussi *un steak sauce au poivre* ou même *un steak sauce poivre*. Les ingrédients sont des N2 assez productifs ; mais si on a *une crêpe chocolat* ou *une gaufre confiture*, on n'aura pas \**un gâteau pommes* ou \**une glace fraise*. Par contre avec le doublement de l'ingrédient, on a naturellement *une glace vanille-fraise* ou même des combinaisons très complexes (et très naturelles) comme *une glace deux boules citron vert-chocolat amer*, où les N modificateurs sont eux-mêmes modifiés. Les matériaux, comme les ingrédients, s'utilisent assez librement comme modificateurs : *une peinture métal*, *une montre or*, *une toiture ardoise*, *une finition bois*, *un revêtement pierre*. Comme pour les ingrédients, on ne dira pas \**un pantalon coton*, mais on aura *un pantalon lin-coton*. On trouve enfin à côté de la construction nominative des combinaisons libres similaires comme *un bébé phoque* ou *une mère kangourou*.

Que conclure de tout ça ? Lorsqu'une combinaison NN est libre, il s'agit soit d'une construction particulière, soit d'un N particulier. Lorsque N2 devient un modificateur complètement libre, comme *une fenêtre standard*, *un cas limite* ou *une chaise marron*, il tend à passer dans la catégorie adjetivale même s'il reste invariable : il peut alors s'employer avec la fonction d'attribut et être modifié par un adverbe (*cette fenêtre est parfaitement standard*, *ce cas est très limite*, *cette chaise est complètement marron*). La combinaison NN ne peut donc pas être considérée comme libre de manière générale. Et une combinaison comme *chien-loup* ne peut pas être considérée comme une combinaison analogue à une combinaison libre. Nous considérons donc qu'il s'agit bien dans ce cas d'un procédé morphologique — la composition nominale — et que le résultat est un syntaxème syntagmatique. Mais on voit que pour décider si les combinaisons liées NN sont des syntaxèmes ou des syntagmes, ce sont avant tout les combinaisons NN libres que nous devons étudier pour voir s'il y a ou non analogie entre les deux. Cette démarche nous distingue nettement des précédentes études sur le caractère morphologique ou non de constructions de ce type qui se concentrent sur les seules combinaisons liées (cf. notamment Corbin 1992, Villoing 2002, Fradin, à paraître).

Nous allons maintenant étudier une autre construction du français qui pose problème pour notre définition du syntagme. Il s'agit des adjectifs utilisés comme compléments de verbe, comme dans l'énoncé *Marie pèse lourd*. Il ne s'agit pas d'une combinaison libre, puisqu'aucun adjectif ne peut commuter avec LOURD, même pas LÉGER. Les combinaisons du type PESER + LOURD sont très irrégulières et toujours liées : *coûter cher*, *parler fort*, *rire jaune*, *sonner creux*, *chanter juste*, *s'habiller jeune*, *voter utile*, etc. (Le Goffic 1993 : 367). Il s'agit donc d'une construction qui n'est analogue à aucune combinaison libre du français, au sens strict où nous avons défini l'analogie. Il existe des combinaisons libres verbe-adjectif du type *Marie se trouve lourde*, mais elles sont d'un autre type, puisque l'adjectif est un attribut du nom, qu'il s'accorde avec le nom et qu'il commute librement avec d'autres adjectifs. La combinaison PESER + LOURD est pourtant bien un syntagme. Si l'adjectif LOURD ne peut commuter avec un autre adjectif, il peut commuter avec divers groupes nominaux, qui eux commutent librement : *Marie pèse 70kg*, *Marie pèse un sacré poids*, etc. Par ailleurs, PESER et LOURD peuvent être séparés et

surtout ils peuvent être modifiés indépendamment l'un de l'autre : *Marie pèse très lourd*, *Marie ne pèse pas lourd*. La combinaison PESER + LOURD est donc la combinaison de deux unités significatives et son comportement est analogue à celui des combinaisons libres et non à celui des syntaxèmes, qui même lorsqu'ils sont complexes, ne peuvent être modifiés que comme un tout. Il ne s'agit donc plus de l'analogie directe avec un syntagme précis, mais d'une analogie de comportement avec les syntagmes en général.

Un dernier exemple nous permettra de rappeler que l'analogie structurelle est une notion graduelle. Le signe *bonne femme*, dans un énoncé comme *Une vieille bonne femme me parlait*, est un bon exemple de segment qui est à la limite entre syntagme et syntaxème. Il s'agit clairement d'une unité significative et donc d'une combinaison liée. La question est alors de savoir s'il existe des combinaisons analogues qui sont libres. A première vue, l'unité significative 'BONNE FEMME' est analogue à la combinaison libre BON + FEMME (*Pierre a rencontré une mauvaise et une bonne femme*). Il existe pourtant une différence distributionnelle assez subtile : en français, le déterminant indéfini pluriel DES possède une forme *de* qui s'utilise devant un adjectif, tandis que la forme *des* est obligatoire devant un nom : *Pierre a acheté des oranges* vs *Pierre a acheté de bonnes oranges*. Or dans un énoncé comme *Pierre a rencontré de bonnes femmes*, seule la combinaison libre BON + FEMME est acceptable. L'unité significative 'BONNE FEMME' n'a donc pas exactement la distribution d'une combinaison libre adjectif + nom, mais plutôt celle d'un nom véritable. Si l'on tient compte de cette différence distributionnelle, il s'agit donc d'un syntaxème et pas d'un syntagme.

Parmi les propriétés des syntaxèmes, l'inséparabilité linéaire est certainement la plus remarquable. Elle est souvent utilisée comme propriété définitoire, notamment lorsque les syntaxèmes sont définis à partir des mots.

Une combinaison A+B est (**linéairement**) **séparable** s'il existe une classe non fermée de syntaxèmes qui peuvent venir s'intercaler entre A et B sans changer la nature de la combinaison entre A et B.

L'inséparabilité linéaire des syntaxèmes ne peut pas être démontrée. Elle peut simplement être constatée<sup>12</sup>. Elle découle très logiquement du fait qu'il n'y a aucune raison qu'un élément vienne séparer un syntaxème : en effet, pour s'intercaler ainsi dans un segment, il faudrait que l'élément intercalaire dépende d'une seule des deux parties du segment et que donc ce segment soit une combinaison de deux unités significatives ou se comporte de manière analogue. Ce qui devrait donc nous surprendre n'est pas qu'un syntaxème soit inséparable, mais qu'une unité significative soit séparable ; par exemple, qu'une locution comme 'NOYER LE POISSON' puisse être séparable : dans *Pierre noyait toujours le poisson*, le syntaxème flexionnel *-ait* et l'adverbe *toujours* se sont intercalés. Ceci montre le caractère particulier des locutions : l'ensemble forme une unité sémantique, mais seule une partie, ici le verbe NOYER, est syntaxiquement accessible et donc les syntaxèmes qui se combinent avec 'NOYER LE POISSON' vont en fait se combiner syntaxiquement au seul verbe NOYER et se positionner par rapport à lui seul.

C'est ce caractère inséparable et cette absence d'indépendance de ses composantes qui rend le syntaxème complexe assez différent d'une locution ou d'une collocation.

Le caractère inséparable des locutions N de N, comme *pomme de terre* ou *maison de campagne*, est parfois utilisé pour conclure qu'il s'agit de lexèmes (on dit plus familièrement des mots composés) plutôt que de syntaxèmes. Nous avons déjà dit qu'il existait des combinaisons libres N de N, comme *part de gâteau*, et que donc les combinaisons figées N de N ne sont pas des syntaxèmes. Mais il est vrai que les combinaisons N de N sont très cohésives et inséparables. Ainsi si un nom possède plusieurs compléments de nom, c'est toujours le complément en de N (avec N non déterminé) qui est en tête : *une chemise de coton pour homme* vs *\*une chemise pour homme de coton, une chemise d'hom*

*e en coton* vs *?une chemise en coton d'homme*. Ce complément en de N ne peut même pas être précédé d'un adjectif : *\*une chemise orange d'homme* vs *une chemise d'homme orange*. Dans Gerdes & Kahane 2006, nous avons appelé ces séquences des amas nominaux et les avons rapprochées des séquences de

verbes, les amas verbaux, qui sont également très cohésives. L'amas se situe ainsi sur l'échelle de cohésion juste après le mot.

## 5 Mots

Le mot représente l'un des niveaux maximal de cohésion d'un groupe de syntaxèmes. Cette cohésion se caractérise par deux propriétés. La première est l'**indissociabilité**, qui, on le verra plus loin, est équivalente à l'**autonomie faible**.

Deux classes distributionnelles sont **indissociables** si un élément de l'une ne peut jamais apparaître sans un élément de l'autre. Deux syntaxèmes sont dits **indissociables** s'ils appartiennent à des classes distributionnelles indissociables l'une de l'autre. Un signe est dit **indissociable** si ses seules décompositions possibles font apparaître des éléments indissociables l'un de l'autre.

Par exemple, *avanç-* et *-ait* sont indissociables dans *Le chat avançait* : *avanç-* doit nécessairement se combiner avec une flexion verbale, tandis que *-ait* doit nécessairement se combiner avec un lexème verbal. Plus généralement, la classe des lexèmes verbaux est indissociable de celle des flexions verbales.

La notion d'indissociabilité utilise fortement la notion de morphème zéro. L'introduction d'un morphème zéro permet de modéliser le caractère obligatoire du choix d'un autre morphème : c'est parce que les énoncés *Je pars lundi* et *Je partais lundi* situent respectivement le procès le lundi suivant et le lundi précédent, que l'absence du *-ai-* est signifiante et doit être considérée comme un morphème zéro (le présent).

Il existe en français deux cas limite d'indissociabilité : le verbe et ses clitics, le nom et l'article.

Commençons par le cas bien connu des clitics préverbaux et tout particulièrement celui des pronoms sujet. Le verbe à l'indicatif en français demande obligatoirement un sujet, comme dans *Marie mange*. Le sujet n'est pourtant pas indissociable du verbe, puisque si un nom propre comme *Marie* peut être le sujet d'un verbe, ce nom ne requiert pas d'être toujours associé à un verbe en tant que sujet. Noms et verbes liés par une relation sujet ne sont donc pas indissociables. Plus complexe est le cas d'une combinaison comme *je mange*. Le pronom *je* doit toujours être associé à un verbe fini accordé à la première personne du singulier et inversement une telle forme verbale doit toujours avoir un sujet et, qui plus est, lorsque cette forme verbale est accordée à la 1<sup>ère</sup> personne du singulier, le sujet ne peut être que *je*. Le lexème verbal et le pronom ne sont pour autant pas indissociables, puisque le pronom *je* peut aussi commuter avec *Marie*, qui, comme on vient de le voir, n'est pas indissociable du lexème verbal. On pourrait plus légitimement considérer que le pronom *je* est indissociable du morphème d'accord de la 1<sup>ère</sup> personne du singulier de la flexion verbale. Néanmoins, s'il y a certes une certaine indissociabilité des deux syntaxèmes, il n'y a pas indissociabilité des classes distributionnelles auxquels ils appartiennent, puisque les accords de la 3<sup>ème</sup> personne ne sont pas indissociables d'un pronom. Nous en concluons donc que *je* n'est indissociable ni du lexème verbal, ni de sa flexion. Mais l'on voit qu'il est indispensable de définir l'indissociabilité pour les classes distributionnelles et pas pour les syntaxèmes isolés.

Passons au cas des articles. Certains déterminants ne peuvent apparaître sans nom, en particulier les déterminants définis *le*, *son* et *ce* ou les indéfinis *un* et *des* : *\*son est venu*, *\*des sont repartis*. La chose est un peu moins nette pour les autres déterminants indéfinis qui ont généralement un emploi pronominal : *certains sont venus*, *deux sont repartis*. La grammaire traditionnelle considère généralement qu'il s'agit d'un cas de polycatégoricité et que l'élément *certains* ou *deux* employé seul n'est plus un déterminant, mais un pronom, avec pour principal argument que pour certains déterminants comme *quelques* ou *chaque*, l'emploi pronominal est marqué (*quelques-uns sont restés*, *chacun est venu*) ; elle en déduit alors que le déterminant ne peut s'utiliser sans nom. Mais qui dit polycatégoricité, dit phénomène lexical : or on n'observe aucun glissement de sens entre les emplois pronominaux et déterminatifs des déterminants. Nous pensons donc qu'il s'agit plutôt d'un phénomène syntaxique et que les déterminants peuvent s'employer sans nom, et ce d'autant plus facilement qu'ils sont moins grammaticaux (les déterminants les plus grammaticaux étant les articles).

Pour le nom, la situation est également complexe : le nom doit nécessairement avoir un déterminant lorsqu'il est sujet du verbe, ainsi que dans la plupart des positions actancielles ; mais dans certaines positions syntaxiques, le nom peut ne pas avoir de déterminant : *Pierre est professeur, avoir peur, parler politique, une chaise sans dossier, livraison gratuite* (sur un tract), etc. L'indissociabilité du nom et du déterminant n'est donc pas totalement claire et elle est au moins conditionnée par le rôle syntaxique du nom. Elle doit néanmoins être retenue pour le nom accompagné de l'article en position sujet.

La quasi-totalité des cas d'indissociabilité concerne des couples formés d'un lexème et d'un morphème flexionnel. On peut donc se demander si les déterminants n'expriment pas une sorte de marquage flexionnel du nom. Le problème est que la classe des déterminants n'est pas fermée, notamment si on y inclut les nombreux déterminants complexes : *trop de gens, un tas d'amis, une kyrielle de problèmes, deux ou trois personnes, je ne sais combien de livres*, etc. On peut pourtant remarquer que les déterminants expriment toujours l'opposition entre défini et indéfini. Donnons un exemple : en français, si on veut dire d'un chat qu'il dort, on ne peut pas juste dire « *chat dormir* ». On doit nécessairement situer l'action dans le temps et donc choisir entre *dort, dormait, dormira, ...* et on doit préciser si le chat en question est déjà identifié ou non par les interlocuteurs et donc choisir entre *le chat et un chat*, c'est-à-dire entre le défini et l'indéfini. Mais contrairement à la flexion verbale qui est un choix clairement fermé, le choix du déterminant reste assez ouvert et les articles *un* et *le* ne sont pas les seuls choix possibles. Pourtant, quel que soit le déterminant qu'il choisira, le locuteur devra exprimer la définitude, car les déterminants se répartissent entre ceux qui expriment le défini comme *le, ce et son* et ceux qui expriment l'indéfini comme *un, chaque, plusieurs, etc.* Les articles *le* et *un* ont une place particulière parmi les déterminants : ils se caractérisent par le fait qu'ils expriment uniquement les sens 'indéfini' et 'défini' et rien d'autre (à part le nombre et le genre avec lesquels ils se combinent) (Kahane 2007). Ce sont les déterminants par défaut. Il existe par ailleurs une corrélation, dans les langues, entre l'existence d'articles, c'est-à-dire d'éléments exprimant la seule définitude et l'obligation d'avoir un déterminant : dans la plupart des langues (russe, chinois, japonais, etc.), le déterminant n'est jamais obligatoire et il n'existe pas non plus de mots exprimant la seule définitude.

Nous allons introduire une propriété équivalente à l'indissociabilité.

Un signe est dit **autonome** s'il peut former un énoncé autonome, tel que « *Attendez !* » ou « *Un chat !* ». On considère également que la réponse à une question forme un énoncé autonome, même si à proprement parler celle-ci ne forme une unité de discours qu'avec la question. On aura ainsi comme énoncés autonomes : *Combien sont-ils ? deux ; Comment est-elle ? gentille ; Quel est son métier ? professeur.*

Comme on le voit, de nombreux mots peuvent former un énoncé autonome et sont donc autonomes. Inversement, il est impossible qu'un signe plus petit qu'un mot soit autonome, puisque des éléments indissociables l'un de l'autre ne peuvent apparaître l'un sans l'autre et donc constituer seuls un énoncé.

De nombreux linguistes ont donc cherché à caractériser le mot comme la plus petite unité autonome (Mel'čuk 1993). Cette caractérisation est néanmoins problématique. Ainsi en français, en dehors de l'impératif, un verbe requiert obligatoirement un sujet et la plupart des formes verbales ne sont donc pas autonomes. De même, les noms ne le sont pas non plus réellement. Le nom *professeur* est un nom un peu particulier qui rentre dans une construction attributive du type *Pierre est professeur*, et il paraît plus difficile de répondre juste *éléphant* à une question. Le nom *éléphant* peut éventuellement être déclaré comme autonome, car il est possible d'utiliser un nom seul dans certaines circonstances : par exemple il n'y aurait rien d'anormal à écrire juste *ELEPHANT* sur une boîte contenant un déguisement d'éléphant (mais il ne s'agit pas à proprement parler d'un énoncé). Pour les formes verbales, on peut éventuellement les considérer comme autonomes par analogie avec les formes verbales impératives, mais l'analogie ne suffirait pas à déclarer autonomes des mots outils tels que prépositions et conjonctions.

On peut étendre l'autonomie dans un sens convenable. Nous nous inspirons ici d'une définition donnée par Mel'čuk (1993 : 171).

Le signe A est dit **faiblement autonome** si A est autonome ou s'il existe un signe B faiblement autonome tel que le signe A+B ou B+A soit un syntagme faiblement autonome.

Reprenons notre énoncé *Le chat avançait*. Le signe *le chat* est autonome, donc *avançait* est faiblement autonome (en appliquant la définition avec A = *avançait* et avec B = *le chat*). Si *chat* est autonome, alors l'article *le* est faiblement autonome, car *le chat* est un syntagme autonome. On a ainsi reconnu comme signes faiblement autonomes les trois mots de l'énoncé.

Attention : il est essentiel de préciser dans la définition de l'autonomie que A+B est un **syntagme**. Dans *Pierre découpe la viande, dé* n'est pas faiblement autonome car, même si *coupe* est autonome, *dé* et *coupe* ne se combinent pas librement et leur combinaison n'est analogue à aucune combinaison libre et n'est donc pas un syntagme.

Il existe, au moins en français, suffisamment de signes autonomes (au sens fort) pour que autonomie faible et indissociabilité soient des propriétés équivalentes au sens suivant : **les signes faiblement autonomes minimaux sont exactement les signes indissociables maximaux**. Autrement dit, si un segment ne contient aucun sous-segment faiblement autonome, alors ses composantes sont indissociables et inversement si un segment ne fait partie d'aucun segment indissociable plus large, alors il est faiblement autonome. Il nous semble préférable de retenir l'indissociabilité comme propriété définitoire des mots, dans la mesure où l'autonomie faible est quand même très largement dépendante de propriétés syntaxiques de chaque langue concernant les énoncés autonomes.

L'inséparabilité est la deuxième propriété caractéristique du mot. Nous avons vu que les syntaxèmes sont inséparables. Nous allons maintenant nous intéresser aux syntagmes inséparables. Par exemple, verbe et morphèmes d'accord en personne-nombre du français, comme *chant-* + *-ons*, sont inséparables : il est seulement possible d'insérer entre les deux des morphèmes de temps tels que *-er-* et *-i-* qui appartiennent à des classes fermées<sup>13</sup>.

L'inséparabilité et l'indissociabilité sont deux propriétés distinctes, bien que souvent associées l'une à l'autre, dans le sens où deux éléments indissociables tendent à être inséparables et vice versa. Nous sommes maintenant en mesure de définir le mot.

Un **mot** est un segment maximal formé de syntaxèmes indissociables et inséparables. Autrement dit, un mot est un signe formé d'un syntaxème seul ou d'une combinaison de syntaxèmes indissociables et inséparables, tel qu'aucun autre syntaxème ne peut être à la fois indissociable et inséparable des syntaxèmes de ce signe.

Le syntagme *avançait* est un mot, mais pas les syntagmes *je mange*, *le chien* ou *fruit de mer*, même s'ils sont particulièrement cohésifs et se rapprochent fortement des mots.

Il n'existe pas à notre connaissance de couples de classes ouvertes qui soient à la fois indissociables et inséparables. Si l'on exclut le cas par conséquent improbable qu'un mot puisse contenir deux lexèmes (à l'exception des quasi-lexèmes des syntaxèmes syntagmatiques), on peut donc considérer (au moins en français) qu'un mot est soit un lexème invariable (non fléchissable), soit la combinaison d'un lexème et de sa flexion.

## 6 Les unités de la syntaxe

La syntaxe est assez souvent définie comme l'étude de l'organisation des mots dans la phrase. Nous préférions la définition suivante (qui ne nécessite pas de définir la phrase, ni même de considérer l'existence d'un domaine maximal de la syntaxe).

La **syntaxe** est l'étude des syntagmes, c'est-à-dire l'étude des combinaisons libres ou analogues.

Notre définition induit que **l'unité (minimale) de la syntaxe est le syntaxème** et non le mot. Ceci nous semble parfaitement justifié car les combinaisons libres possèdent des propriétés bien différentes des combinaisons liées. Parmi ces différences, on notera la suivante, déjà notée par d'autres linguistes. En

français, il existe plusieurs contrastes de prononciation entre lexèmes et formes fléchies. Par exemple, le lexème nominal *percheron* et la forme verbale *percherons* : le lexème est nécessairement trisyllabique, tandis que le syntagme peut être bisyllabique [pers.rɔ̃] (Mel'čuk 1993). Autrement dit, le segment *perch-* de la forme verbale *percherons* peut être prononcé comme si le [ʃ] était la consonne finale d'un mot, comme dans *une perche ronde* par exemple. La limite entre le radical et la flexion reste ainsi un point de rupture possible et la forme fléchie peut être prononcée comme le lexème *percheron* ou comme la combinaison de lexèmes *perche ronde*. Ce point de rupture particulier entre le lexème verbal et sa flexion se retrouve avec d'autres formes verbales. On observe ainsi qu'une forme comme *mourrai* peut être prononcée [mur.rɛ] avec un doublement de la consonne [r] totalement impossible à l'intérieur d'un lexème. De même, une forme comme *cèderont* est prononcée avec un [ɛ] ouvert (à contraster avec *nous cédons* ou *Cédric* [cedrik]), marquant que la syllabe finale du radical verbal est une syllabe fermée et que donc le /d/ de [ced.rɔ̃] appartient au lexème verbal. On a encore des paires minimales du type (*un*) *bouclier*, nécessairement trisyllabique [bu.cli.je], et (*vous*) *boucliez*, possiblement bisyllabique [bucl.je] (Morin 2007). A l'opposé, on présente souvent les phénomènes d'allomorphie du radical verbal dans la conjugaison française comme un phénomène typique de la combinaison à l'intérieur du mot, justifiant le fait de séparer les combinaisons à l'intérieur du mot des combinaisons entre mots et faire ainsi du mot la frontière entre syntaxe et morphologie. Mais ces phénomènes d'allomorphie ne sont pas limités à la combinaison de syntaxèmes à l'intérieur du mot. En voici quelques exemples : **ma copine** vs **mon amie**, **un fol espoir** vs **un espoir fou**, **le sucre** vs **du sucre** vs **au sucre**, **six chats** [sisa] vs **six amis** [sizami] vs **six ou sept chats** [sisusetja].

Le mot est donc simplement un syntagme possédant des propriétés particulières, dont les composants sont à la fois indissociables et inséparables, ce qui en fait un syntagme très cohésif. Plus cohésif encore est le sous-syntagme formé par les morphèmes flexionnels d'un mot, que nous appelons l'**amas flexionnel** (cf. le découpage phonologique de *mourrai* en *mour+rai*) : ce syntagme se caractérise en particulier par des amalgames fréquents, qui rendent quasi indiscernables les uns des autres les différents syntaxèmes qui le composent. En fait, le mot se situe sur une échelle de cohésion syntaxique allant du syntaxème seul jusqu'au discours complet. Il existe des couples de syntaxèmes, comme le nom et le déterminant, qui, bien que ne formant pas un mot, ont une grande cohésion en ayant à la fois des propriétés d'indissociabilité et d'inséparabilité, qui sans être aussi strictes que celles des syntaxèmes d'un mot, sont néanmoins remarquables. D'une part, l'article et le nom sont indissociables dans les positions actancielles où le nom seul est impossible (et aucun des deux n'est réellement autonome). D'autre part, les éléments qui peuvent séparer l'article et le nom sont très contraints : leur ordre est rigide (la séquence *le seul autre joli petit chien* n'admet aucun autre ordre) et aucun vrai syntagme ne peut s'insérer. Des adjectifs éventuellement modifiés par des adverbes comme *très* ou *plus* peuvent venir s'insérer, mais en aucun cas d'autres types d'éléments : on dit **un haut mur**, **un très haut mur** ou **un mur haut de trois mètres**, mais pas **\*un haut de trois mètres mur**. De même, on peut dire **le mur le plus beau du monde** ou **le plus beau mur du monde**, même si *du monde* dépend syntaxiquement de *plus beau*, mais pas **\*le plus beau du monde mur**.

## 7 Conclusion

Nous avons montré la nécessité de distinguer trois types d'unités de la langue : morphème, syntaxème et unité significative. La distinction de ces différentes unités résulte de l'évolution des langues. De nouvelles unités significatives sont créées régulièrement et, à part les emprunts à d'autres langues, les onomatopées et certains noms de marque, les nouvelles unités sont des polymones construits en combinant des morphèmes déjà existants.

Le propre des unités significatives est de pouvoir se combiner librement, en dehors du phénomène tout à fait remarquable des collocations. La notion de combinaison libre joue ainsi un rôle central dans la caractérisation des unités de la langue. On peut s'étonner que cette notion relativement simple à définir reste si peu mentionnée<sup>14</sup> dans les travaux en syntaxe. Nous pensons qu'elle marque la frontière entre la syntaxe et la morphologie et que l'unité de la syntaxe est le syntaxème, que nous avons grossièrement défini comme la plus petite unité à pouvoir commuter librement.

Outre la combinaison libre, nous avons introduit deux autres notions qui ont servi à caractériser les différents types d'unités de la langue : l'analogie structurelle et la notion de choix. Seule l'analogie permet de distinguer parmi les combinaisons liées celles qui sont des syntaxèmes ou des syntagmes. Enfin la caractérisation des unités significatives est indépendante de celles des syntaxèmes et syntagmes et doit donc faire appel à une notion particulière : c'est la notion de choix, dégagée par Martinet, et trop peu souvent invoquée dans les travaux de linguistique.

Nous avons enfin proposé une caractérisation du mot, en rappelant que le mot n'est qu'un niveau de cohésion syntaxique parmi d'autres. Il est important de souligner que cette notion complexe doit être nécessairement définie en deux étapes : un mot est 1) une combinaison indissociable et linéairement inséparable de signes 2) analogues à aucune combinaison libre. Ces derniers signes sont les syntaxèmes, que nous avons donc définis avant et indépendamment des mots. Nous avons par ailleurs montré l'équivalence des notions, pourtant très différentes, d'indissociabilité et d'autonomie faible.

Sur le plan descriptif, nous avons considéré plusieurs constructions du français qui présentent des cas limite d'unités d'un type ou d'un autre, notamment les constructions avec un adjectif invariable régi par un verbe (*Marie pèse lourd*), les constructions NN libres et liées (*une moules frites, un oiseau-lyre*) ou la syllabification remarquable des formes verbales (*mour.rai, forge.rons, boucl.iez, cède.rons*).

## Remerciements

Le contenu de cet article a été développé dans le cadre de l'écriture d'un livre sur la syntaxe écrit avec Kim Gerdes. Je remercie les différentes personnes qui ont relu et commenté le manuscrit de cet article : Kim Gerdes bien sûr, mais aussi Julie Glikman, Sara-Anne Leblanc, Pierre Magistry, Alain Polguère, Marie-Pierre Sales, et surtout Igor Mel'čuk, qui a beaucoup souffert devant cette tentative de redéfinir en quelques pages les notions qu'il a développées tout au long des 2000 pages de son Cours de Morphologie Générale et dont il a longuement discuté avec moi à maintes reprises.

## Références

Anscombe J.-C. (1990), Pourquoi un moulin à vent n'est pas un ventilateur, *Langue Française*, 86, 103-125.

Aronoff M. (1976), *Word Formation in Generative Grammar*. Cambridge : MIT Press.

Bloomfield L. (1926), A set of postulates for the science of language, *Language*, 2, 153-164.

Corbin D. (1992), Hypothèse sur les frontières de la composition nominale, *Cahiers de grammaire*, 25-55.

Costaouec D., Guérin F. (2007), *Syntaxe fonctionnelle – Théorie et exercices*, Presses Universitaires de Rennes.

Ducrot O. (1995), Unités significatives, in Ducrot O. & Schaeffer J.-M., *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris : Seuil, 358-365.

Fradin B. (à paraître), Compounding in French, in Lieber R. & P. Atekauer (eds), *Oxford Handbook on Compounding*, Oxford: Oxford University Press.

Gerdes K. & Kahane S. (2006), L'amas verbal au cœur d'une modélisation topologique de l'ordre des mots, in K. Gerdes & C. Müller (éds), *Ordre des mots et topologie de la phrase française*, *Linguisticae Investigationes*, 29:1, 75-89.

Goldberg A. (1995), *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.

Harris Z. S. (1951), *Structural linguistics*, Chicago & London : The University of Chicago Press.

Kahane S. (2007), La distribution des articles du français, in M. Charolles, N. Fournier, C. Fuchs & F. Lefevre (éds.), *Parcours de la phrase - Mélanges offerts à Pierre Le Goffic*, Ophrys, Paris, 159-174.

Kerleroux F. (2004), Sur quels objets portent les opérations morphologiques de construction ?, *Lexique*, 16, *La formation des mots : horizons actuels*, 85-123.

Le Goffic P. (1993), *Grammaire de la phrase française*, Paris : Hachette.

Martinet A. (1960), *Éléments de linguistique générale*, Paris : Armand Colin.

Mel'čuk I. (1993), *Cours de morphologie générale, vol 1 : Le mot*, Paris : CNRS/Presses de l'Université de Montréal.

Mel'čuk I., Clas A. & Polguère A. (1995), *Initiation à la lexicologie explicative et combinatoire*, Louvain-la-Neuve : Duculot.

Morin Y.-C. (2007), Les yods des terminaisons -ions et -iez de l'imparfait de l'indicatif et du présent du subjonctif des verbes français, in Delais-Roussarie É. et Labrune L. (éds.), *Des sons et des sens: Données et modèles en phonologie et en morphologie*, 125–140. Paris : Hermès.

Saussure F. (de) (1916), *Cours de linguistique générale*, Paris.

Touratier C. (2003), *Morphologie et Morphématique. Analyse en morphèmes*, Presses de l'Université de Provence.

Villoing F. (2002), *Les mots composés [VN]<sub>N/A</sub> du français : réflexions épistémologiques et propositions d'analyse*, thèse de l'Université Paris X.

<sup>0</sup> A part la correction de coquilles et l'ajout des notes 6bis et 7bis, cette version comporte une seule modification par rapport à la version publiée dans les actes du CMLF, mais celle-ci est importante : nous avons utilisé le terme *syntaxème* en lieu et place du terme *morphème* et le terme *morphème* comme variante dominante du terme *monème*.

<sup>1</sup> Nous utilisons le terme *segment* pour désigner toute portion continue d'un énoncé.

<sup>2</sup> Nous utilisons le terme **collection** pour désigner une réunion d'éléments du même type. Ainsi une collection de signes est un signe (généralisé) plutôt qu'un ensemble de signes.

<sup>3</sup> La notion remonte à Bloomfield qui utilise le terme *morpheme* : « A minimum form is a morpheme; its meaning is a sememe. Thus a morpheme is a recurrent (meaningful) form which cannot in turn be analyzed into smaller recurrent (meaningful) forms. Hence, any unanalyzable word or formative is a morpheme. » (Bloomfield 1926 : 155, cité d'après Villoing 2002).

<sup>4</sup> L'allomorphie complique quelque peu la situation. Mais même une combinaison comme *all+ons* est libre dans le sens donné ici, puisqu'ils existent d'autres morphes verbaux, comme *finiss-*, qui se combinent avec quasiment les mêmes flexions (-*ons*, -*ez*, -*ais*, ...) que *all-*. Par ailleurs, une fois établi que *all-* est en distribution complémentaire avec *v*-, *i*- et *aill-* et que donc ces quatre morphes forment un unique morphème, on voit que la distribution de ce morphème est la même que celle des autres verbes.

<sup>5</sup> On ne peut pas considérer que deux éléments se combinent ou forment un segment simplement parce qu'ils sont linéairement adjacents. Ainsi considérer que *chat avanç-* est un segment qui commute librement avec son environnement, cela presuppose que l'on considère que *chat* et *avanç-* se combinent bien l'un à l'autre. Nous pensons qu'un lexème verbal se combine effectivement avec le nom tête de son sujet, mais le fait que le verbe est indissociable de sa flexion et le nom quasi indissociable de son déterminant, nous amène à penser que la combinaison du lexème verbal et du nom sujet met obligatoirement en jeu l'ensemble de ces éléments et que donc *chat avanç-* n'est pas un segment valide.

<sup>6</sup> Nous utilisons les majuscules pour noter les syntaxèmes lexicaux. Pour les verbes, nous utilisons alors la forme infinitive qui est la forme conventionnelle de citation en lexicographie française. Nous utilisons les minuscules italiques pour désigner tous les types de signes par leur signifiant. On ne confondra donc pas le lexème CHANTER (équivalent à *chant-*) avec la forme fléchie *chanter*.

<sup>6bis</sup> Cela signifie que nous ne considérons pas comme de la dérivation la combinaison libre de certains affixes, comme cela est courant dans les langues agglutinantes, du hongrois à l'inuktitut. Des préfixes comme ANTI-, MINI- ou SUPER- seront également considérés comme des syntaxèmes et non comme des morphèmes dérivationnels.

<sup>7</sup> La notion de faisceau de signes a été introduite par Mel'čuk (1993 : 278) dans la description des morphèmes flexionnels, sans être exploitée réellement.

---

<sup>7bis</sup> Comme me l'a fait remarquer Sara-Anne Leblanc après la publication de l'article, cette définition est fausse : il est possible dans certaines locutions de commuter un lexème par des lexèmes quasi-synonymes ('REMUER/RETOURNER/ENFONCER LE COUTEAU DANS LA PLAIE<sup>1</sup> ou 'CHIER/FAIRE DANS SON FROC/PANTALON').

<sup>8</sup> Que la distinction soit réelle, n'empêche pas la frontière entre les deux notions d'être perméable. Ainsi, l'expression *poser un lapin* 'ne pas venir à un rdv', qui est à l'origine une locution (métaphore figée), est devenue aujourd'hui une collocation pour certains locuteurs, puisque des énoncés comme *c'est le deuxième lapin qu'il me pose en une semaine* sont attestés (le clivage et la modification sur LAPIN indique clairement que LAPIN est traité comme une unité significative à part entière et POSER comme un verbe support).

<sup>9</sup> Il faut bien sûr exclure les infinitifs convertis en nom comme *un dîner* ou *le savoir*, où la combinaison n'est plus libre et l'infinitif fonctionne comme un suffixe dérivationnel.

<sup>10</sup> Dans la tradition de Martinet, on parle également de monème lié, libre et libérable (Costaouec & Guérin 2007 : 59). Dans ce cas, un monème libérable est défini un monème dont l'un des mones qui peut former un mot et cette définition est censé caractériser les monèmes qui contiennent un lexème. Cette définition ne nous convient pas pour deux raisons au moins : d'une part, contrairement à nous, elle n'inclut pas les morphèmes flexionnels parmi les monèmes libres ou libérables (car ils sont indissociables du radical). D'autre part, elle suppose qu'un nom singulier ou qu'un verbe à la 3<sup>ème</sup> personne du singulier du présent de l'indicatif n'ont pas de flexion (alors que nous considérons qu'il s'agit de morphèmes flexionnels zéros). Plus généralement, ces notions n'utilisent pas la notion de combinaison libre.

<sup>11</sup> Les éléments lexicaux qui ne s'utilisent plus (guère) que dans des locutions sont quand même assez nombreux : (*être*) aux *aguets*, (*avancer*) à la *queue leu leu*, à l'*instar* de, (*utiliser qqch*) à *bon/mauvais esclent*, (*poser une question*) à *brûle-pourpoint*, (*rouler*) à toute *blinde*, (*aller*) au diable *vauvert*, (*avoir qqn*) dans le *collimateur*, de *plein gré*, (*coeur*) battre la *chamade*, faire de la *charpie* (*de qqch*), de *bric* et de *broc*, de *guingois*, de *traviole*, en *catimini*, en *filigrane*, mettre la *sourdine*, en un *tournemain*, en *vrac*, et tout le *bastringue*, manger à tous les *râteliers*, mettre *qqn* sur la *sellette*, passer au *crible*, prendre la *poudre d'escampette*, sans coup *férir*, sans *encombre*, se faire du *mouron*, sans soucier comme d'une *guigne*, sonner le *glas* (*de qqch*), tailler des *croupières* (à *qqn*), tous *azimuts*, etc.

<sup>12</sup> Harris 1951 ou Martinet 1960 considèrent des séquences d'accords, comme les trois pluriels dans *les animaux paissent*, comme un seul morphème/monème séparable. Ceci découle pour nous de la non-distinction entre monème et unité significative. La séquence précédente de pluriels est l'expression d'une unique unité significative, c'est-à-dire d'un unique choix du locuteur, mais cela n'empêche pas de considérer chaque accord comme un monème/morphème séparé. Ceci est justifié par le fait qu'il n'y a pas nécessité à ce que les différents morphèmes commandés par l'unité significative en question soient toujours présents ensemble. Considérer de tels objets comme des morphèmes reviendrait non seulement à considérer des morphèmes séparables, mais qui plus est des morphèmes à géométrie variable possédant selon les contextes un, deux, trois ou davantage encore de morceaux. Les morphèmes d'accord sont des syntaxèmes qui n'ont pas de sens, mais ont une signification grammaticale, puisqu'ils marquent un lien syntaxique entre le lexème qu'ils fléchissent et un autre lexème (Mel'čuk 1993).

<sup>13</sup> Il est également possible d'insérer un suffixe dérivationnel entre *chant-* et *-ons* : *chant-onn-ons*. Mais il ne s'agit pas alors d'une insertion entre les **syntaxèmes** *chant-* et *-ons*, car dans le segment *chantonnons*, *chant-* n'est plus un syntaxème : en effet, la combinaison *chant+onn-* n'est analogue à aucune combinaison libre et *chant-* est alors un mone sous-lexical.

<sup>14</sup> Il nous semble difficilement imaginable que cette notion n'ait jamais été introduite avant, mais nous n'avons pas réussi à en trouver de définition dans les travaux que nous avons consultés. On trouve bien sûr des utilisations du terme *combinaison libre* (à commencer par Saussure 1916, cité dans notre texte), mais jamais de caractérisation formelle de la notion.