

L'identification des unités lexicales

Bernard Fradin

1. Objet

La question du choix des unités devant faire partie de la nomenclature se pose à tout dictionnaire. La réponse qu'y apportent les dictionnaires de langue est fonction à la fois de la tradition lexicographique et des visées pragmatiques propres à ce type d'ouvrage. Mais cette question se pose également à toute théorie linguistique qui se donne un composant lexical articulé au reste de la grammaire. Très souvent cependant les théories linguistiques n'indiquent pas explicitement comment elles obtiennent les unités qu'elles font figurer dans leur lexique. Elles se contentent de reprendre, pour l'essentiel, les choix que leur lèguent les dictionnaires existants. Quant aux approches qui se placent dans l'optique de l'ingénierie linguistique (e. g. les dictionnaires électroniques), la visée pratique qui est la leur les amène à être plus exhaustives, plus cohérentes et plus systématiques que les dictionnaires classiques dans l'établissement de leur nomenclature (plus grand nombre d'expressions phraséologisées¹, dégroupement plus fréquent en cas de polysémie, etc.). Celle-ci cependant se fonde sur celle des dictionnaires existants (ce qui est normal), et il est extrêmement rare que ces approches engagent une réflexion d'envergure sur les principes linguistiques qui justifient le choix des unités constituant leur nomenclature (peut-être simplement parce que leurs préoccupations se situent bien en aval) et encore moins sur les catégories linguistiques qu'elles emploient².

L'objet de cet article est de reprendre à la lumière de quelques développements récents de la morphologie des questions classiques, qui abordent de front le problème précédemment évoqué, telles que : comment délimiter les unités lexicales ? Quelles unités introduire dans un dictionnaire ? Comment les distinguer ? Comment les noter ?

L'actualité de ces questions déborde largement l'intérêt qu'on peut porter à la question du lexique, puisque, en droit, quiconque utilise les termes de *morphème*, *mot*, *affixe*, *lexème* peut être sommé de les définir de façon cohérente et empiriquement justifiée.

¹Dans cet article, expression dénote toute portion d'énoncé (cf. § 2. pour ce terme). Lorsqu'il s'agira d'expression idiomatique ou d'expression phraséologisée, ce sont ces derniers termes qui seront employés.

²Exemples entre mille, Courtois [1990] ou Namer [1994] utilisent les notions de mot, suffixe, thème ("stem") sans les définir et Ritchie et al. [1992] donnent une définition de l'affixe interne à leur représentation. Il est vrai qu'en ingénierie des langues, cette absence de définition est compensée en pratique par des définitions en extension (liste de tous les suffixes, etc.). Ce qui la rend d'autant plus insupportable en linguistique classique où un tel biais, même si l'on y recourt parfois, n'est pas de mise.

2. Position du problème

2. 1. La délimitation des unités. La question des unités lexicales peut tout d'abord être envisagée du point de vue de la délimitation des unités. Etant donné un corpus de textes ou d'énoncés, quelles sont les unités linguistiques qui composent ces textes ou ces énoncés ? Cette démarche correspond à la phase d'analyse que pratique toute science empirique. Parmi les unités d'analyse qu'on obtient, les unités minimales figurent en bonne place même si elles n'en constituent pas l'unique objet. Cette démarche demande qu'on éclaircisse d'emblée deux points : celui de l'ensemble de départ qu'on se donne à analyser d'une part, ce qu'on entend par unité minimale d'autre part.

L'ensemble à analyser peut être appréhendé de façon purement externe, sans faire intervenir aucun critère permettant de le caractériser formellement. Il sera composé de toutes les productions langagières susceptibles d'être enregistrées ou retranscrites. C'est ce que fait Mel'čuk lorsqu'il caractérise le matériau premier (de la théorie sens ↔ texte), à savoir le texte :

“(...) nous appliquons le terme *texte* au côté extérieur, physique de TOUTE manifestation langagière. Ainsi nous considérons comme des textes les signifiants des morphes, des mots-formes, des phrases, des alinéas, etc. de même que des nouvelles ou des romans entiers” [Mel'čuk, 1993, p. 42 (c'est l'auteur qui souligne)].

Cette conception du texte s'apparente à celle de corpus. Rastier [1996], qui récuse les trois pôles extrinsèques à partir desquels toute une tradition a pensé le texte (rapport à l'énonciateur, au réel et au lecteur), donne une caractérisation du texte dont le premier point intègre la précédente :

“Un texte est une suite linguistique empirique attestée, produite dans une pratique sociale déterminée, et fixée sur un support quelconque” [Rastier, 1996, p. 19].

En général cependant, l'ensemble de départ est conçu comme formé d'unités moins vastes, qui sont les unités maximales à soumettre à l'analyse. Ces unités sont les énoncés. On se souvient que les structuralistes américains s'en tenaient à une vision purement extensionnelle de l'énoncé. Harris [1951] le définit ainsi : “toute partie de discours, tenue par une seule personne, avant et après laquelle il y a un silence de la part de cette personne”. Cette définition introduit un critère formel permettant, en théorie, de délimiter les énoncés du corpus. On a souvent souligné son caractère insatisfaisant, mais force est de reconnaître que peu d'alternatives ont été proposées. Il est significatif que Mel'čuk [1993, p. 69] la reprenne presque terme pour terme : “un énoncé est tout fragment de parole capable d'apparaître, dans les circonstances normales,

entre deux silences du locuteur". Dans la mesure où, comme le rappelle Feuillet [1988, p. 42], "il s'agit d'une description préscientifique des données du linguiste", les caractérisations qu'on vient de rappeler suffisent à discriminer à la fois l'ensemble et les unités de départ. Ducrot a été amené à définir en compréhension l'énoncé :

"J'appelle «énoncé» un segment de discours qui fait l'objet d'un choix global — en ce double sens qu'aucune de ses parties n'est présentée comme choisie indépendamment du choix de l'ensemble et qu'il est lui-même présenté comme choisi indépendamment de ce qui suit dans le discours" [Ducrot, 1983, p. 171].

L'idée était de faire le lien entre l'énoncé et la visée illocutoire de l'énonciateur (laquelle pouvait être complexe). Mais cette définition, outre qu'elle introduit la notion difficilement maniable de "choix global", n'a que peu de pertinence pour le problème qui nous intéresse car l'énoncé ainsi défini recouvre aussi bien de simples mots que des textes entiers (e. g. pièce de théâtre ; cf. Moeschler et Reboul [1994])³.

J'introduis ici quelques précisions terminologiques qui me semblent utiles. L'énoncé a souvent été opposé à la phrase comme étant du côté de l'actuel (il ne se répète pas), du donné pragmatique (c'est ce qu'échangent les locuteurs), alors que celle-ci est un objet virtuel, construit par le linguiste (avec beaucoup de variantes que je ne peux aborder ici (Pottier, Culioli, Benveniste, etc.) et Feuillet [1988] pour une discussion pertinente). La phrase fait intervenir la dimension grammaticale, celle d'un tout organisé par les règles de la langue. La définition qu'en donnait Meillet⁴ convient encore :

"La phrase peut être [ainsi] définie : un ensemble d'articulations liées entre elles par des rapports grammaticaux, et qui, ne dépendant grammaticalement d'aucun autre ensemble, se suffisent à elles-mêmes" [Meillet, 1937, p. 355].

Il reste que cette conception implique que toute répétition d'un énoncé ne peut être décrite (ou subsumée) qu'en termes de phrase (cf. Martin [1983, p. 249]). Or on a besoin de dire que des énoncés peuvent être répétés, non pas matériellement, mais du point de vue de l'effet pragmatique qu'ils encodent de manière conventionnelle. Je songe à ce que Fónagy [1982] nomme les énoncés liés (ou récurrents e. g. *Mais, c'est pas vrai !, A qui le dites-vous !, Tu parles !*) dont l'emploi est associé à un effet illocutoire (et perlocutoire) codé et impose un cadre interlocutif déterminé :

"Le sens des énoncés «libres», considérés comme non récurrents, est entièrement déterminé par le contexte et la situation à partir de la signification de la phrase sous-jacente. Ce n'est pas vrai des énoncés récurrents, liés à des situations typiques. Leur sens est prédéterminé ; c'est la langue elle-même qui prévoit ce sens en fonction de telle ou telle situation où l'énoncé est censé apparaître. L'acte d'énonciation aura donc un champ de manœuvre limité. Elle

³Derrière l'expression "choix global", point la notion de "totalité du sens" (le choix est global parce qu'il correspond à un sens complet), héritée de la tradition grammaticale antique, cf. Rastier, [1996, p. 26 notamment].

⁴Cité par Feuillet [1988, p. 44-45] qui cite également une définition très proche de Bloomfield. Le parallélisme entre cette définition de la phrase et celle que donne Ducrot de l'énoncé mérite d'être souligné.

⁵*Le rappel des notions d'énoncé (dans les deux versions) et de phrase ne veut pas dire qu'on ignore les problèmes qu'elles soulèvent et encore moins qu'on les endosse complètement. Les énoncés liés concernent notre propos dans la mesure où la question de leur intégration dans la nomenclature du dictionnaire se pose (au même titre que celle des expressions idiomatiques).*

⁶*La distinction entre substrat et abstrat ne se superpose pas aux autres distinctions qu'on a pu proposer à propos du signe. Elle vise à prendre acte du fait basique et constitutif des systèmes symboliques, à savoir que des agencements d'entités matérielles (sons, acides aminés dans le cas du code génétique) doivent être envisagés autrement que dans leur seule dimension matérielle (qui ne disparaît cependant nullement) parce qu'ils expriment des régularités pouvant être condensées sous forme d'un symbole, cf. Varela [1989, p. 179-185]. Relève du substrat tout l'aspect matériel du signe (ce qui peut être mesuré, enregistré, perçu) ainsi que son organisation. Relève de l'abstrat ce qui concerne l'interprétation mais aussi les informations catégoriques renvoyant au fonctionnement du système (les informations*

concrétise non pas une signification, mais un sens usuel acquis au cours d'actes de concrétisations précédentes" [Fónagy, 1982, p. 48].

La description de ce phénomène suppose minimalement qu'on dispose d'un moyen de noter la répétition d'énoncé. C'est à cette fin qu'avait été introduite la distinction entre énoncé-type et énoncé-occurrence : "l'énoncé est ce qui subsiste de l'énoncé-occurrence lorsqu'on fait abstraction de l'aspect événementiel de ce dernier" [Anscombre, Ducrot, 1976, p. 18]. Dans ce système à trois termes (énoncé-occurrence, énoncé-type et phrase), l'énoncé-occurrence correspond à l'énoncé du système précédent. La phrase y est un être grammatical dont la valeur illocutoire n'est pas déterminée (puisqu'elle est déconnectée de l'énonciation). L'énoncé-type en revanche permet de subsumer la répétition d'énoncés-occurrences ayant même valeur illocutoire (par exemple la répétition de "J'ai oublié ce que tu m'as dit" dans le même contexte et la même situation à quelques moments d'intervalle)⁵. Il reste à voir comment la description des énoncés liés pourrait tirer parti de l'ouverture que présente un système à trois termes.

S'il est vrai qu'un énoncé peut être "le produit de l'énonciation d'une phrase" [Moeschler, Reboul, 1994], il peut se faire aussi qu'il soit constitué d'expressions qui ne forment pas des phrases (cf. celles qui servent à répondre, les interjections, les salutations et autres phrasillons). Ce qui confirme que l'énoncé s'inscrit plus du côté de l'acte illocutoire ancré dans une situation de parole. Dans cet article, je considérerai la phrase comme l'unité syntagmatique maximale à laquelle ait affaire la syntaxe.

En ce qui concerne les unités minimales, celles-ci se conçoivent relativement à un plan d'analyse et de représentation. Les plans d'analyse pertinents relèvent de la première articulation, c'est-à-dire mettent en jeu le rapport entre face matérielle (phonique sauf dans le langage des sourds-muets), que j'appelle substrat, et face abstraite du signe, que j'appelle par symétrie abstrat⁶. Ce qui exclut la phonologie et ne laisse en présence que la syntaxe et la morphologie. Dans la mesure où l'on admet, au moins comme hypothèse de travail, que la morphologie se distingue de la syntaxe, les unités minimales en question seront de l'ordre de l'atome syntaxique et de l'atome morphologique. Pour des raisons liées à la nature même du signe linguistique (cf. Fradin [1993, p. 41-51 ; 1996b]), il est nécessaire d'admettre que les unités minimales se conforment à quelque chose de l'ordre du Principe de réalisation :

(1) **Principe de réalisation**

Toute entité (constituant, trait, contenu sémantique) introduite dans les représentations linguistiques doit présenter une dimension relevant du substrat (i. e. phonique ou gestuelle).

Pour ce qui regarde la morphologie, ce principe correspond à l'Hypothèse nulle qu'ont exprimée Bach [1983] ou, dans un cadre différent, Carstairs [1987] :

“Je considérerai en général comme hypothèse nulle que seules les distinctions morphologiquement (c'est-à-dire en définitive, phonologiquement) réalisées seront retenues” [Bach, 1983, p. 71].

“Si les propriétés morphosyntaxiques sont ce que les marques flexionnelles réalisent, alors une distinction entre deux propriétés qui ne serait jamais manifestée par aucune distinction entre des formes lexicales [word-forms] est impossible” [Carstairs, 1987, p. 3].

Ce principe fait fonction de rasoir d'Occam et élimine des analyses les entités segmentales abstraites, en l'occurrence sans contenu phonique, qui ont tendance à proliférer dans certaines approches. Il n'implique pas cependant que l'on fasse sienne l'hypothèse de base de la Morphologie classique selon laquelle les unités minimales présentent une correspondance biunivoque entre face phonique et face interprétative⁷. Il exige simplement que ce qu'on postule au niveau abstrait (celui de l'interprétation) soit corrélé au substrat phonique, fût-ce de manière multivoque. Corrélativement, un tel point de vue permet de faire l'hypothèse que minimal n'implique pas isolable au plan segmental, ce qui permet d'éviter certaines apories du distributionnalisme (cf. § 3. 3. *infra*).

De manière synthétique, on peut dire que les unités d'analyse sont obtenues suivant deux modes. Un mode externe, qui vise en propre à délimiter dans la suite syntagmatique quelles sont les unités atomiques pour la syntaxe (cf. U en (1a)). Un mode interne, qui analyse chaque U (unité lexicale) et établit les unités morphologiques M dont cette unité se compose éventuellement (cf. (3)) :

(2) X U Y X, Y suites contextuelles pouvant être nulles

(3) [U M₁, ..., M_n]

L'intérêt de (3) réside dans le fait qu'il indique, contrairement à (2), comment sont construites les unités lexicales, ce qui est très utile pour établir une typologie de ces dernières (cf. § 5. 1.).

2. 2. Le stockage des unités. La seconde manière d'aborder la question des unités lexicales est de se demander quelles unités atomiques doivent figurer dans le lexique, étant entendu que la discrimination de celles-ci a déjà été opérée au moins pour la syntaxe. Cette seconde manière suppose qu'on s'accorde préalablement sur ce qu'on entend par lexique (et lexical). Deux conceptions de ce dernier ont cours habituellement⁸. Dans la première le lexique est le répertoire des unités

morphologiques —
classe de déclinaison,
genre, etc. —
et syntaxiques —
catégorie, type de
rection, etc.), ce que
Mel'čuk [1993] appelle
le syntactique.

⁷Cf. la définition du
morphème comme la
plus petite unité
phonique corrélée à
une unité de sens. Pour
une présentation et une
critique de la
Morphologie classique,
cf. Anderson [1992],
Fradin [1993, 1995a].

⁸Pour une discussion
récente de ce point, cf.
Aronoff [1994,
p. 16-22]. Cette
discussion interfère
avec la distinction entre
lexique et grammaire,
mais je ne peux
développer ce point ici.

d'une langue qui présentent une ou des propriétés qui ne peuvent être prédites par les règles de la langue. Lexical équivaut peu ou prou à idiosyncrasique. Dans la seconde, lexical s'oppose à grammatical. Relève du grammatical tout ce qui concerne le fonctionnement de la langue, donc les mots fonctionnels (encore improprement appelés mots-vides), qui constituent des classes fermées. Relèvent du lexique tous les items dont l'apport sémantique est substantiel plutôt que centré sur le système lui-même. Par opposition, ces items forment très généralement des classes ouvertes et peuvent être têtes de constituant syntaxique (cf. Emond [1985, p. 159-160]). Ces deux visions du lexique sont conceptuellement distinctes. Il est sûr qu'on a besoin du lexique appréhendé comme le répertoire des idiosyncrasies (ce qui n'exclut pas que le lexique ainsi conçu soit structuré, notamment au plan sémantique), car il faut bien noter quelque part les informations linguistiques irréductibles à d'autres. Il est probable que l'opposition liste ouverte vs liste fermée renvoie aussi à des propriétés qui ne sont pas nulles⁹. Mais cette opposition ne recouvre pas celle existant entre les éléments qui peuvent être tête de constituant syntaxique et ceux qui ne le peuvent pas (cf. les prépositions). Ce qui indique que la seconde conception du lexique se fonde sur des propriétés moins incontestables que la première. D'autre part, même si ces propriétés peuvent jouer çà et là un rôle dans la description linguistique, cela n'implique pas qu'elles suffisent à discriminer un concept intéressant de lexique : quel intérêt peut avoir la liste de tous les noms (verbes, etc.) d'une langue, y compris ceux qui sont complexes et construits de manière régulière ?¹⁰. Dans ce qui suit, j'entendrai le terme lexique selon la première conception.

⁹Les éléments des listes fermées présentent souvent des propriétés qui ne se rencontrent pas ailleurs. Par exemple, en anglais le son [ð] ne se trouve à l'initiale qu'avec les déictiques (démonstratifs et article).

¹⁰La réponse diffère vraisemblablement suivant qu'on se place du point de vue de l'ingénierie linguistique, de la psycholinguistique, ou qu'on s'assigne pour but de décrire la compétence des sujets parlants.

Le concept de lexique étant fixé, il va de soi que la question du choix des unités à inclure dans ce dernier ne se pose que si l'on rejette l'idée d'y faire entrer toutes les unités préalablement discriminées. Ce point de vue étant admis, il nous faut répondre à deux types de question, à savoir : selon quels critères va-t-on inclure les unités U dans le lexique ? quelle représentation va-t-on donner aux unités ainsi sélectionnées ? Cette dernière question fera l'objet du § 4.2.

En ce qui concerne les critères, plusieurs types ont été proposés. Le premier fait jouer la distinction grammatical vs non grammatical sous la forme suivante : les segments dont le fonctionnement (tant au plan phonétique que sémantique) met en jeu des mécanismes spécifiques de la grammaire n'ont pas à entrer dans le lexique. A titre d'exemple, pour le français cela inclut les interrogatifs (*qui, que, quoi, combien, etc.*) et les divers clitiques pronominaux (*je, tu, il, elle, les, etc.*) ou non (*ne, si, ci*). Dans cette optique, ne sont retenus dans le lexique que les unités dont le comportement peut être décrit (suivant les divers plans de description) en fonction de leurs seules propriétés et des règles générales de la grammaire. Ce critère se fonde sur l'idée que le lexique ne dispose pas de l'outillage

nécessaire pour assurer la description des phénomènes dont les éléments en question sont la manifestation. Conséquemment, ces éléments n'ont pas à entrer dans le lexique¹¹. Le second, plus largement admis, fait intervenir l'opposition basique imprédictible vs non basique prédictible. Il englobe tout ce qui relève de la flexion : ne figurent dans le lexique que les formes lemmatisées des unités U à l'exclusion de leur version fléchie (cf. § 4. 1. pour un développement). Le troisième type de critère participe de la même idée. Sous sa forme idéale, il vise à distinguer entre un lexique restreint, qui ne comporte que les unités simples, et un lexique étendu, qui englobe toutes les unités complexes construites à partir des unités simples. Le lexique étendu, qui comporte toutes les unités U atomiques pour la syntaxe est en effet fortement redondant. L'idée est d'exprimer cette redondance sous forme de règles toutes les fois que cela est possible et de réduire d'autant le nombre d'unités stockées dans le lexique (qui se trouve donc restreint). Le rapport entre lexique restreint et lexique étendu peut s'envisager selon deux options (qui ne sont pas équivalentes). Dans l'option générative (au sens mathématique), on ne dispose que du lexique restreint et d'un ensemble de règles qui s'appliquent à ce dernier permettant ainsi d'engendrer le lexique étendu. Dans l'option statique en revanche, on dispose d'emblée des deux types de lexique (le lexique restreint est inclus dans l'étendu). Les règles se bornent à établir si une unité de surface U correspond à une suite d'unités sous-jacentes (i. e. appartenant au lexique restreint) ou, à l'inverse, si une suite d'unités sous-jacentes peut être corrélée à une unité U attestée¹².

Indépendamment des critères ci-dessus, la seconde manière d'aborder la question des unités lexicales pose la question de la lexicalisation des expressions U qu'on a discriminées dans les énoncés et, corrélativement, du rapport entre construction des lexèmes et lexicalisation. Cette question sera abordée au § 5. 1.

3. La délimitation syntagmatique des énoncés

3. 1. Méthode. La question de la délimitation syntagmatique des unités a fait l'objet de beaucoup de travaux structuralistes dans les années 40 et 50 mais semble relativement passée de mode depuis. A l'heure actuelle, la plupart des théories ne s'attardent pas sur cette question. Peut-être considère-t-on cette affaire comme réglée ou, plus simplement, est-on rebuté par l'aspect fastidieux des manipulations qu'elle demande d'entreprendre et la difficulté qu'il y a à les mettre au point. Il n'en demeure pas moins qu'une discipline qui affiche des prétentions scientifiques devrait être en mesure d'établir clairement quelles sont les unités minimales qu'elle manipule. On aurait tort toutefois de s'étonner

¹¹Certains dictionnaires de langue (e. g. DFC, GLLF), plus conscients du problème évoqué, se sentent obligés d'introduire des encarts grammaticaux pour expliquer le comportement des items en question.

¹²La distinction entre lexique étendu et lexique restreint d'une part et les deux options mentionnées de l'autre ont des incidences importantes en morphologie (cf. la Morphologie à deux niveaux et les approches déclaratives). Pour une discussion, cf. Zwicky [1992], Fradin [1994, 1995a] et les références qu'ils citent.

¹³Il aurait été instructif de présenter aussi Cruse [1986], qui se sert de procédures basées sur les proportions alors que Mel'čuk et Fradin utilisent surtout des procédures basées sur les manipulations (cf. *infra*).

¹⁴Pour une présentation critique de celles-ci, cf. Tooby [1965] et aussi Fradin [1993, ch 2].

¹⁵Ces procédures sont examinées en détail dans Fradin [1993, p. 21-32]. Elles sont basées sur ce que Cruse [1986, p. 16] appelle les "relations contextuelles". Contrairement aux deux autres, l'inférence n'intervient pas sur la suite phonique de l'énoncé et ne sert qu'à trouver les données sémantiques. Pour cette raison, elle peut être utilisée en combinaison avec les deux autres types. Les procédures basées sur l'inférence restent largement inconnues des structuralistes qui se privaient ainsi du principal moyen objectif de mettre au jour les données sémantiques.

des embarras que fait naître cette question, car ils ne font que refléter la complexité des signes linguistiques. Dans les lignes qui suivent, je présenterai brièvement deux tentatives récentes de reprendre cette question : Mel'čuk [1993], Fradin [1993]¹³.

Le but, je le rappelle, est d'obtenir les unités, minimales et non-minimales pour la syntaxe et la morphologie à partir de l'analyse des énoncés. Ces analyses mettent en jeu des procédures réglées dont l'archétype sont les procédures élaborées par les structuralistes (notamment américains)¹⁴. Ces procédures sont de trois types : (a) manipulations opérées sur la chaîne syntagmatique, (b) construction de proportions, (c) inférence¹⁵. Le type (b) consiste à construire des paradigmes de formes dans lesquels on s'assure que le rapport entre numérateur et dénominateur est constant (e. g. *laver/lavable* = *chanter/chantable*, etc. (cf. Saussure [1916, ch 6], Cruse [1986])). Le protocole général des manipulations (a) est le suivant (L désigne une langue donnée. Entre chevrons, le jugement des locuteurs) :

(4) **Protocole général des manipulations linguistiques**

I. Prendre un énoncé E_i tel que $E_i \in L$ (i. e. <oui, E_i >).

II. Opérer une manipulation sur E_i pour obtenir E_k .

III. Evaluer E_k par rapport

(a) à L (i. e. <oui, E_k > vs <non, E_k >).

(b) à E_i .

L'étape (I) signifie que les données produites doivent toujours être comparables, en dernière analyse, avec des énoncés attestables, sinon attestés. Les manipulations évoquées en (II) sont de divers types : substitution, suppression, déplacement, ajout de matériel linguistique (cf. Fradin [1993, p. 28-31] pour une présentation plus soignée de ce protocole). L'étape (III) est celle où le candidat-énoncé nouvellement produit est soumis au jugement du locuteur. En (III)(a) soit $E_k \in L$. On dit que E_k est un énoncé acceptable (grammatical) dans langue en question. Soit $E_k \notin L$ (l'énoncé n'est pas acceptable dans la langue en question). L'étape (III)(b) est envisageable que le résultat de la manipulation soit un énoncé acceptable ou non. La confrontation de E_i , l'énoncé de départ, à E_k , le résultat de la manipulation, est à la base de l'élaboration des faits en linguistique. Il serait probablement sans grand intérêt de donner une typologie détaillée de tous les cas de figure que cette comparaison autorise *a priori*. Il importe de retenir qu'elle vise à établir trois choses : (i) si E_i et E_k peuvent s'analyser de manière identique ou non pour n plans de description ; (ii) si E_i et E_k ont les mêmes interprétations pour un contexte donné ; et enfin (iii) si E_i implique E_k ou *vice versa*. Établir ces points conduit à réappliquer récursivement les manipulations mentionnées en (II) ainsi que celle faisant usage de l'inférence évoquée plus haut.

3. 2. Mel'čuk [1993]. Cet ouvrage est la première partie d'un énorme travail de fond qui vise à fournir "une présentation FORMELLE de toutes les connaissances morphologiques. (...) Il s'agit en fait de la construction d'un métalangage artificiel, adapté à la description du mot et de son comportement en langue" [Mel'čuk, 1993, p. 1]. Je me contenterai ici de faire saisir, dans ses grandes lignes, la manière dont Mel'čuk s'y prend pour discriminer les unités atomiques pour la syntaxe (le U de (2)), qu'il appelle mots-formes¹⁶.

Le premier objectif est de parvenir à isoler les expressions libres de la langue. A cette fin, le critère d'autonomie forte est défini :

(5) Autonomie forte :

un signe segmental X est appelé *autonome au sens fort* dans une langue L si et seulement si il existe en L un énoncé complet qui contient X et qui ne contient aucun autre signe segmental [Mel'čuk, 1993, p. 170].

Seront autonomes au sens fort, les phrases entières, les titres, les réponses dans un dialogue (qui peuvent être des phrases, des syntagmes (cf. *infra*) ou des expressions simples e. g. *Que fait Jules ? — Rôtisseur. Où allez-vous ? — A la gare. Vous reviendrez ? — Pas question*). Le pas suivant consiste à élargir l'emploi de l'autonomie forte à des contextes autres que ceux qui satisfont la définition de celle-ci :

(6) Autonomie forte généralisée :

un signe segmental X est appelé *autonome au sens fort généralisé* si et seulement si il appartient à la même classe distributionnelle syntaxique qu'un signe X' autonome au sens fort (*ibid.*).

Dans la mesure où *rôtisseur* est autonome au sens fort, tout nom commun partageant sa distribution sera *ipso facto* fortement autonome en vertu de (6). La notion d'autonomie forte généralisée a, semble-t-il, pour fonction de faire passer d'une autonomie forte valable pour une instance à une autonomie forte valable pour un type (le type auquel appartient l'instance, l'appartenance étant fondée sur la distribution)¹⁷. Mel'čuk ne précise pas comment sont établies les classes distributionnelles. On supposera qu'il a en tête les tests classiques (du type de (4)), mis en place par les structuralistes américains [Harris, 1951 ; Wells, 1947] et aussi Gleason [1955]. Il reste maintenant à traiter les cas où l'on n'a pas d'autonomie forte.

(7) Autonomie faible :

un signe segmental X est appelé *autonome au sens faible* dans une langue L si et seulement si les deux conditions suivantes sont vérifiées simultanément :

(a) L possède un signe ou un complexe de signes autonomes au sens fort de la forme $X\Psi$, ΨX ou $\Psi_1 X \Psi_2$ [$\Psi_1 + \Psi_2 = \Psi$]

¹⁶Ici, comme dans Mel'čuk [1993] ou Fradin [1993, 1996a, 1996b], mais contrairement d'autres [Gruaz, 1987 ; Cruse, 1986 ; Corbin, 1987], le terme mot est exclu de la terminologie qu'on cherche à établir. Chaque fois qu'il sera dorénavant employé, il gardera le sens qu'on lui attribue dans la langue générale, laquelle ne se préoccupe pas de description linguistique.

¹⁷Sinon on ne voit pas très bien son utilité dans la mesure où le contexte servant à établir qu'une expression est fortement autonome fait nécessairement partie des contextes constitutifs de la classe distributionnelle.

(b) X satisfait au moins quelques-uns des critères d'autonomie spécifiques pour L [*op. cit.*, p. 171].

Une expression est autonome au sens faible quand elle n'apparaît qu'en combinaison avec des expressions autonomes au sens fort sans être soi-même telle (c'est ce que dit (7a)). Ainsi, pour reprendre l'exemple de Mel'čuk, l'expression *Le professeur lit un journal* est un signe autonome au sens fort, de même que *le professeur* et *un journal*. *Lit*, qui n'est pas autonome au sens fort, sera un signe autonome au sens faible "comme résidu d'un énoncé complet après soustraction de tous les signes autonomes au sens fort" [Mel'čuk, 1993, p. 171]. La nécessité d'introduire les critères spécifiques (7b) montre que tous les cas d'autonomie faible ne peuvent être saisis d'un seul mouvement. Ces critères font intervenir la séparabilité (i. e. la possibilité d'intercaler des éléments entre deux signes), la variabilité distributionnelle (i. e. la capacité d'un signe à se combiner à n classes distributionnelles, $n \geq 1$) et la transmutabilité (i. e. la capacité pour un signe de permuter avec un autre ou d'occuper une autre position dans la chaîne). Dans les trois cas, la relation sémantique entre les signes et leur contenu sémantique ne doit pas être affectée lors des manipulations mises en jeu par ces critères. Ces derniers ne sont pas catégoriques mais graduels : chacun peut être satisfait à un degré variable. D'autre part, ils sont hiérarchisés¹⁸ :

(8) séparabilité > variabilité distributionnelle > transmutabilité.

Ces critères s'avèrent indispensables pour différencier de manière fine des signes comme les clitiques, qui sont toujours faiblement autonomes mais présentent de grandes variations y compris à l'intérieur d'une même langue (cf. Mel'čuk [1993], Klavans [1982, 1985], Nevis [1986], Zwicky [1985]). L'autonomie faible généralisée est définie en recourant à la classe distributionnelle sur le modèle de (6) [Mel'čuk, 1993, p. 172]. Ces distinctions étant opérées, il devient possible de définir le concept de mot-forme :

(9) **Mot-forme de la langue L :**

un signe segmental X de L est appelé *mot-forme de la langue de L* si et seulement si une des deux conditions suivantes est vérifiée :

1. soit X est suffisamment autonome en L et X n'est pas représentable en termes d'au moins deux mots-formes de la langue de L
2. soit X est représentable en termes d'un mot-forme de la langue w de L et des alternances de L, tel que X est toujours remplaçable par w sans affecter la grammaticalité ou le sens de la phrase [Mel'čuk, 1993, p. 188].

Parmi les expressions, cette définition permet d'opérer les tris suivants :

¹⁸Mel'čuk ne dit pas si la hiérarchie joue en faveur de l'autonomie faible (i. e. la séparabilité serait un argument plus fort pour celle-ci que les autres) ou de l'autonomie forte (l'inverse).

(i) sont des mots-formes : *livre, vite, tard, loup, recommandable* (autonomie forte) ; *de, à, chatouillera, chiennes* (autonomie faible).

(ii) ne sont pas des mots-formes les éléments en gras suivants : *déconnecté, balayeur, tigresse* (ce sont des affixes).

(iii) ne sont pas des mots-formes (mais des syntagmes phraséologisés ou phrasèmes) : *chemin de fer, manger la grenouille, à façon, c'est-à-dire, chambre froide, mener en bateau.*

(iv) ne sont pas des mots-formes : ang. *could've, would've* (la condition 1 n'est pas respectée puisque ces expressions sont subdivisibles : *could* est mot-forme et *'ve* est un clitique simple. (cf. Zwicky [1985], Fradin [1996b] sur ce concept).

(v) ne sont pas des mots-formes : rus. *texsovet* "conseiller technique", *speckomissija* "commission spéciale" (condition 2 non respectée : *tex* ou *spec* sont reliés à d'autres mots-formes de la langue — les adjectifs *texničeskij* "technique" et *special'nyj* "spécial" — mais leur comportement grammatical n'est pas identique : *tex, spec* ne se déclinent pas). Ces expressions s'apparentent aux mots-valises (cf. § 5. 2.).

Mel'čuk insiste beaucoup sur le caractère relatif du mot-forme et sur la nécessité de prendre en compte dans sa caractérisation des paramètres propres à chaque langue (ou famille de langues). Dans la mesure où, dans une langue, les mots-formes typiques présentent en général des propriétés phonologiques, morphologiques ou syntaxiques assez identifiables, il faudra toujours comparer les expressions dont on cherche à savoir si elles ont un statut de mot-forme à ces mots-formes typiques suivant les trois plans de description mentionnés¹⁹ (cf. Mel'čuk [1993, p. 190-222] et § 4. *infra* pour une discussion et des exemples).

La définition (9) s'attache à distinguer les expressions qui sont des mots-formes stabilisés dans tous les contextes. Mais il peut se faire qu'une expression acquière le statut de mot-forme de manière éphémère, le temps d'un contexte, sans qu'on doive la considérer comme un mot-forme en langue. Pour cette raison, Mel'čuk prend soin de définir une notion de mot-forme de la parole de L (que je ne reproduis pas ici, cf. [*op. cit.*, p. 187]).

En résumé, le concept de mot-forme apparaît comme une tentative d'appréhender la notion d'atome syntaxique d'une manière rigoureuse mais qui laisse néanmoins une marge pour traiter les cas moins typiques.

3. 3. Fradin [1993]. Cette tentative reprend à nouveaux frais les procédures élaborées par les structuralistes américains mais sans présupposer que les segments minimaux qu'on mettra au jour seront nécessairement des morphèmes au sens classique. La démarche procède

¹⁹On songe en particulier aux "particules" séparables de l'allemand ou du hongrois, aux éléments entrant dans certains composés, aux postpositions du hongrois, au noyau verbal fléchi avec clitiques en français et dans les autres langues romanes, aux formes du futur avec clitique incorporé du portugais, aux temps composés.

en deux étapes (un peu comme le fait Mel'čuk). Dans la première, il s'agit de proposer des procédures (du type de (4)) permettant de tester à l'aveugle la cohésion des segments contigus afin de faire un tri entre ceux qui sont pleinement dépendants, ceux qui le sont moyennement et ceux qui ne le sont pas. Dans la seconde, des tests supplémentaires, tirant cette fois parti des propriétés propres à chaque langue, sont mis en place afin de classer les segments obtenus. Ces tests sont graduels car ils reposent aussi sur les notions de séparabilité, de variabilité distributionnelle et de transmutabilité (même si ces termes ne sont pas explicitement employés). Ils visent à établir sur des bases objectives les classes d'unités morphologiques que sont les mots-types, les affixes et les clitiques, du moins dans leur version prototypique car — c'est une donnée des langues — des cas qui dérogent au type se rencontrent toujours.

Dans la mesure où les procédures figurant dans Fradin [1993] s'intéressaient surtout aux unités morphologiques, la manière de caractériser les expressions libres n'était pas spécifiée directement. Pour ce faire, on peut reprendre la notion d'autonomie forte de Mel'čuk (cf. aussi Zwicky [1992]). Les procédures en question sont applicables à n'importe quelle (portion d') expression libre (i. e. énoncé complet). Elles sont basées sur la possibilité de substitution de segments. Préalablement, j'introduis les notions de groupe distributionnel et de construction franche :

(10) **Groupe distributionnel :**

si dans une suite σ ABC la sous-suite BC peut être remplacée par D et qu'on ait toujours un énoncé recevable, alors BC forme un groupe distributionnel (D est distributionnellement équivalent à BC dans σ).

Une construction franche est un groupe distributionnel débarrassé de ses éléments optionnels (*optionnel* est défini dans Fradin [1993, p. 166]), dont le rapport sémantique avec ceux-ci est compositionnel. (Dans ce qui suit, $[[X]]$ désigne la signification de X, \oplus note l'opération qui combine les significations des éléments, ε désigne la suite vide et I' l'élément neutre pour l'interprétation) :

(11) **Construction franche :**

XAY est une construction franche si

- (a) on a XABY (ou XBAY) et que B est optionnel
- (b) $[[XABY]] = [[XAY]] \oplus [[B]]$

A partir de l'expression libre *le petit renard dort*, on voit que *le petit renard* forme un groupe distributionnel et *le renard* une construction franche²⁰. En revanche, *le petit* n'est pas une construction franche car cette expression peut désigner un enfant (la condition (ii) n'est pas remplie). L'autonomie d'un segment est testée relativement à ses segments voisins :

²⁰La présentation que je viens de donner est un peu inexacte car les clauses (a) et (b) concernent toujours l'aspect phonique des expressions. Normalement, la transcription devrait être donnée en API. Une construction franche est une expression autonome (au sens de [Mel'čuk, 1993]) qui ne contient pas de segments optionnels.

(12) Indépendance pleine :

dans une suite XABY, A est pleinement indépendant de B si

- (a) XAY est possible
- (b) $X' A \in$ avec $X' = X$ ou $X' \neq X$
- (c) $[[XABY]] = [[XBY]] \oplus [[A]] \wedge [[A]] = [[A]] \oplus I'$

Intuitivement, A est pleinement indépendant de l'expression contiguë B si l'on peut avoir la même suite non seulement sans B, mais aussi sans rien à la place de B. La clause (c) garantit que A conserve son interprétation qu'il figure seul ou combiné avec B (compositionnalité). Ne pouvant illustrer le fonctionnement de ces tests de façon détaillée, des exemples sont présentés sous (16). Les expressions autonomes (dites aussi libres), sont celles pour lesquelles on a $\varepsilon A \varepsilon$ en (12b). Les degrés moindres d'autonomie sont éprouvés par les trois tests qui suivent :

(13) Indépendance par parentèle :

dans une suite XABY, où X et Y peuvent être nuls, A est dépendant de B par parentèle (et inversement pour B par rapport à A) si

- (a) *XAY (et donc *XAε)
- (b) XAB'Y où $B' \neq B$
- (c) $[[XABY]] = [[XBY]] \oplus [[A]] \wedge [[XAB'Y]] = [[XB'Y]] \oplus [[A]]$

Intuitivement, A est dépendant par parentèle d'un élément B dans XABY si B peut être remplacé par n'importe quel élément B' de la même classe que B (B' partage les caractéristiques de B, dont la distribution), si B' ne peut être supprimé et si A garde une interprétation stable qu'il soit composé avec B ou avec B' (cf. la notion de contraste sémantique identique récurrent chez Cruse [1986, ch 2]). Par exemple, dans *pendable* /pãd/ et /abl/ sont tous deux dépendant par parentèle (parce qu'on a *pendaison, pendre, pendu, pendant, etc.* d'une part, *lavable, louable, changeable, etc.* de l'autre). Les éléments non autonomes sont dits pleinement dépendants :

(14) Dépendance pleine :

dans une suite XABY, où X et Y peuvent être nuls, A dépend pleinement de B (et inversement pour B par rapport à A) si.

- (a) *XAY (donc *XAε)
- (b) *XAB'Y quel que soit $B' \neq B$
- (c) $[[XABY]] = [[XBY]] \oplus [[A]] \wedge [[XB'Y]] \oplus [[A]] = \perp$

Un élément A dépend pleinement de B quand il n'est pas même dans une relation de dépendance par parentèle avec ce dernier et qu'il ne présente aucune interprétation stable, quel que soit l'élément B avec lequel on le compose. Dans *érable*, et /εr/ et /abl/ sont tous deux pleinement dépendants (autrement dit, *érable* n'est pas une expression complexe).

Dans *royaume* /om/ sera pleinement dépendant et /rwaj/ sera dépendant par parentèle (cf. *royal*. Il faut établir ensuite qu'il s'agit d'un allomorphe de /rwa/ qui, lui, est pleinement indépendant).

Le concept de dépendance structurale est introduit comme une variante de la dépendance pleine pour traiter les séries telles que *percevoir*, *décevoir*, *recevoir*, *concevoir*, etc., *conduire*, *déduire*, *produire*, etc. ou encore *permettre*, *démètre*, *commettre*, etc.

(15) **Dépendance structurale :**

dans une suite XABY, où X et Y peuvent être nuls, A dépend structurellement de B (et inversement pour B par rapport à A) si

- (a) *XAY (et donc *XAε)
- (b) XAB'Y où B' ≠ B
- (c) $[[XABY]] = [[XBY]] \oplus [[A]] \wedge [[XB'Y]] \oplus [[A]] = \perp$

Cette notion diffère de la dépendance pleine uniquement par le fait que la clause (b) y est possible (comme dans la dépendance par parentèle). Elle indique que *percevoir* est composé des deux segments structurellement dépendants /pɛr/ et /səvwar/ et, à ce titre, complexe.

Les types d'éléments segmentaux mis au jour sont illustrés dans le tableau (16)²¹ :

²¹Dans Fradin [1993, p. 170-173] sont aussi discutés les cas des "berry-words" de l'anglais e. g. huckleberry, loganberry, raspberry, etc. et des postpositions locatives du hongrois.

| (16) | Indépendant | Dépendant par parentèle | Dépendant plein | Dépendant structural |
|-----------------------|--------------|-------------------------|-----------------|----------------------|
| <i>le chat</i> | /ʃa/ | /lə/ | | |
| <i>le petit chat</i> | /pti/ /ʃa/ | /lə/ | | |
| <i>son expédition</i> | /ɛkspedisjõ/ | /sõn/ | | |
| <i>expédition</i> | | /ɛkspedi/ | | |
| | | /sjõ/ | | |
| <i>de la gare</i> | /gar/ | /də/ /la/ | | |
| <i>tropical</i> | /trõpik/ | /al/ | | |
| <i>pendable</i> | | /pãd/ /abl/ | | |
| <i>démontrable</i> | | /demõtr/ | | |
| | | /abl/ | | |
| <i>érable</i> | | | /ɛr/ /abl/ | |
| <i>royaume</i> | (/rwa/) | /rwaj/ | /om/ | |
| <i>percevoir</i> | | | | /pɛr/ /səvwar/ |
| <i>permettre</i> | /mɛtr/ | | | /pɛr/ |

Une expression indépendante pour laquelle la clause (12b) vaut aussi bien à gauche qu'à droite sera un mot-forme : c'est le cas pour *chat* /ʃa/, mais pas pour *pend-* /pãd/. Alors que pour Mel'čuk [1993, p. 188], les prépositions et les articles du français sont des mots-formes (dépendance faible), le traitement proposé en fait des expressions dépendantes par parentèle (ce qui semble plus conforme à leur statut de clitique).

Les tests présentés discriminent trois classes de segments en fonction de leur degré de dépendance : indépendants, dépendants par parentèle,

dépendants structuraux. Si on combine ce paramètre et celui de la minimalité, nous obtenons les possibilités suivantes :

| | | Dépendant | |
|---------|---|-----------|-----|
| | | + | - |
| Minimal | + | (a) | (b) |
| | - | (c) | (d) |

Les segments du tableau (16) se répartissent comme suit :

- (18) (a) /lə/, /la/, /də/, /sɔn/, /al/, /sjɔ̃/, /pɑ̃d/, /ɛkspedi/, /abl/, /rwaj/, /om/, /pɛr/, /səvwar/
 (b) /rwa/, /erabl/, /pti/, /ʃa/, /trɔpik/, /mɛtr/, /gar/
 (c) /demɔ̃tr/
 (d) /ɛkspedisjɔ̃/, /pɛrsəvwar/, /rwajom/, /ləʃa/

A priori, tous les éléments minimaux sont des candidats-morphèmes (c-morphèmes). Cependant, ne seront pas considérés comme morphèmes ceux qui sont dépendants pleins (e. g. /om/) et ceux qui sont dépendant structuraux (e. g. /pɛr/, /səvwar/). A la suite de Tournier [1985], je dénommerai ces derniers paléomorphèmes²². Cette dénomination me paraît tout à fait adaptée dans la mesure où *percevoir*, *permettre*, *conduire*, etc. gardent encore visible la trace d'une découpe morphématique qui valait en latin mais ne vaut plus en français. Inversement, tous les éléments pleinement indépendants à la fois à gauche et à droite par le critère (12b) — en l'occurrence (b) et (d) — sont des candidats-mots-formes. Ceux qui sont minimaux (i. e. (d)) seront de fait des mots-formes. Parmi ceux qui ne sont pas minimaux, seuls seront des mots-formes ceux qui ne sont pas une construction franche. Cela élimine à juste titre /ləʃa/. Etant construction franche, /ləʃa/ peut être augmenté d'éléments optionnels qui sont des mots-formes (cf. /ləptiʃa/), ce qui n'est normalement pas le cas des mots-formes : cela montre que /ləʃa/ garde le statut de construction minimale. Ne restent que les éléments où intervient au moins un élément dépendant (quel qu'il soit). Le tri que nous poursuivons, où les tests font office de trémies successives, donne ceci :

- (19) (a) **Mots-formes** : /rwa/, /erabl/, /pti/, /ʃa/, /trɔpik/, /mɛtr/, /gar/, /ɛkspedisjɔ̃/, /pɛrsəvwar/, /rwajom/.
 (b) **C-morphèmes** : /rwa/, /erabl/, /pti/, /ʃa/, /trɔpik/, /mɛtr/, /gar/, /lə/, /sɔn/, /al/, /sjɔ̃/, /pɑ̃d/, /ɛkspedi/, /abl/, /rwaj/.
 (c) **Formants résiduels (segments)** : /om/, /pɛr/, /səvwar/²³.
 (d) **Non pris en compte (dans ce classement)** : /demɔ̃tr/.

²²Gruaz [1987, p. 30], qui a le mérite de bien distinguer ces éléments, les appelle morphons. Cette appellation me paraît moins évocatrice que celle de Tournier.

²³J'appelle formant toute manifestation phonique (segmentale ou suprasegmentale) pouvant être identifiée au moyen de manipulations linguistiques. Cette définition est plus descriptive que celle de Matthews [1972, p. 185].

Comme le rappellent les sept premiers exemples de (19a), un mot-forme peut parfaitement se réduire à un c-morphème. La question de savoir si tous les c-morphèmes seront des morphèmes dépend de la définition et de la représentation qu'on donnera de cette notion. Tout dépend des options théoriques qu'on est amené à prendre concernant la morphologie. Je ne ferai qu'effleurer ces questions au § 6.

Les classes que nous avons obtenues jusqu'à présent, et qui sont illustrées en (16) et (19), concernent uniquement des segments. Ce qui est normal, compte tenu que les tests ayant permis ces tris sont des tests de bas niveau, c'est-à-dire faisant appel à des propriétés peu spécifiques des expressions concernées. L'étape suivante consiste justement à caractériser plus précisément ces segments en ayant recours à des propriétés linguistiques plus variées et plus précises. Ces propriétés sont les suivantes (elles sont des variations sur (8), comme je l'ai dit) : affectation catégorielle, dépendance catégorielle, directionnalité de combinaison, ordre fixe, coordonnabilité, dépendance prosodique (on peut en ajouter d'autres). Je ne chercherai ni à les expliciter, ni à les illustrer ici (cf. Fradin [1993]), et me bornerai à indiquer qu'elles permettent d'identifier provisoirement trois catégories morphologiques, à savoir : mots-formes, affixes et clitiques. Comme souvent, ces catégories présentent des caractères centraux et d'autres plus périphériques. Il en va de même des segments qu'auront mis au jour les procédures présentées ici. Ceci explique que l'assignation d'un segment à une catégorie morphologique ait parfois un caractère graduel. En ce qui concerne les segments minimaux figurant en (16), (19), le résultat est donné en (20) (cf. (40) pour une vue définitive) :

| (20) Mot-forme | | Non défini à ce stade | Affixe | Clitique |
|-------------------|---------------|---------------------------|---------------------|------------------|
| <i>chat</i> | /ʃa/ | <i>pend-</i> /pād/ | <i>-al</i> /a/ | <i>le</i> /lə/ |
| <i>petit</i> | /pti/ | <i>roy-</i> /rwaj/ | <i>-able</i> /abl/ | <i>son</i> /sɔn/ |
| <i>expédition</i> | /ɛkspedisjɔ̃/ | | <i>-tion</i> /sjɔ̃/ | <i>de</i> /də/ |
| <i>tropique</i> | /trɔpik/ | | | <i>la</i> /la/ |
| <i>roi</i> | /rwa/ | <i>expédi-</i> /ɛkspedi/ | | |
| <i>mettre</i> | /mɛtr/ | <i>démontr-</i> /demɔ̃tr/ | | |

Les éléments sans catégorie morphologique dans le tableau (20) n'en recevront une qu'après l'introduction de la notion de lexème (cf. § 4. 1.). Sur la base de critères faisant intervenir de nouvelles propriétés, on pourra déterminer s'ils sont des radicaux et ou des variantes allomorphiques (radicales ou affixales). Notons que des éléments comme *à* ou *de*, faiblement autonomes chez Mel'čuk et dépendants par parentèle ici, sont

des clitiques suivant le classement (20) alors qu'ils sont des mots-formes pour Mel'čuk (ce qui cadre mal avec leur comportement morphologique).

3. 4. Le point sur la délimitation syntagmatique. Contrairement à ce que laissent accroire nombres d'ouvrages sur la question, la délimitation syntagmatique des unités lexicales n'est pas chose aisée. La démarche qui a été suivie par les deux approches présentées est assez semblable. Chacune procède par tris successifs et s'efforce d'identifier le type d'entité obtenu à chaque étape. Bien que concernées en premier lieu par l'analyse sur le mode externe (cf. (2)), ces approches, pas plus que d'autres, n'ont pu éviter d'aborder l'analyse interne des unités.

Je voudrai brièvement mettre en perspective les catégories obtenues avec celles qu'on trouve dans des classifications proposées par ailleurs, qui reprennent aussi de façon critique une partie de l'héritage structuraliste en la matière (cf. Zwicky [1992] pour un rappel). Cela nous permettra de surcroît d'appréhender les problèmes qu'il reste à étudier.

Les signes autonomes au sens fort correspondent à ce que les structuralistes (américains surtout) ont appelé les expressions libres de la langue. Le mot-forme correspond aux expressions libres minimales, ce que Zwicky [1992] note W_{MIN} (voir plus loin). Quel rapport entretient le mot-forme avec le concept d'atome syntaxique ? Cruse [1986, p. 35-36] reconnaît deux propriétés caractéristiques à l'atome syntaxique typique :

- (21) (a) Il est le plus petit segment manifestant une mobilité positionnelle dans un contexte donné.
- (b) Il est le plus petit segment ne tolérant pas l'insertion de matériel linguistique.

En fait ces caractéristiques sont celles qui ont été attribuées de longue date au mot-forme typique (cf. [Wells [1947] et Nida [1949] qui le désignent par *mot* ; Mel'čuk [1993])²⁴. Cela voudrait donc dire que le mot-forme équivaut à l'atome syntaxique. Bien que cela soit vrai dans beaucoup de cas, cette équivalence demeure fautive au niveau des principes. Comme l'a souligné Zwicky [1992], des exemples du genre (22) montrent que les expressions qui fonctionnent comme atomes syntaxiques ne sont pas forcément minimales :

- (22) (a) tire-au-flanc, va-et-vient, quant-à-soi, sot-l'y-laisse, m'as-tu-vu, monte-en-l'air, tourne-à-gauche, décrochez-moi-ça, laissez-passer ; us et coutumes
- (b) chasse-neige, protège-cahier, cure-dents, rabat-joie, passe-droit, ouvre-boîte
- (c) bar-tabac, café-concert, jupe-culotte
- (d) blanc-bec, grand-route, gras-double, moyen-âge ; peau-rouge,

²⁴Elles distinguent du même coup celui-ci des affixes et des syntagmes.

tête-brûlée, messe-basse, haut-fourneau, poids-lourd, société anonyme, chambre froide, trou noir

- (e) hors-jeu, en-avant, plus-value, arrière-plan, base arrière
- (f) hélice à pas variables, pile à maillet, heure de pointe, avion à réaction
- (g) plus grand commun diviseur, association loi 1901, format demi-commercial, arbre recouvrant euclidien minimal ; missile mer-sol, tricoteuse rectiligne doublefonture simple chute à mécanisme jacquard avec report²⁵.

²⁵Ce dernier exemple, tiré de Perdrizet [1983, p. 82], montre que dans des domaines technologiques ou conceptuels fortement structurés les atomes syntaxiques peuvent être très longs.

Chaque exemple de (22) est un atome syntaxique qui satisfait les propriétés énoncées en (21). Il a aussi la même distribution (aux restrictions de sélection près) que des mots-formes — de catégorie nominale — comme *chaise, arbre, chat, garagiste*. Ces exemples ne sont pourtant pas des atomes syntaxiques minimaux puisqu'ils mettent en jeu plusieurs mots-types (outre, parfois, des éléments qui ne sont pas des mots-types). Suivant Zwicky [1992], je les noterai W_{MAX} et réserverai W_{MIN} aux atomes syntaxiques minimaux (les mots-types). Les propriétés des atomes syntaxiques non-minimaux sont résumés en (23) :

(23) L'atome syntaxique non minimal (W_{MAX})

1. Expression constituée d'éléments qui sont des mots-formes ou des syntagmes.
2. Qui a la distribution d'une classe de mot-formes (pour une catégorie syntaxique donnée).
3. Qui satisfait les critères (21).

A partir de la notion d'expression libre — c'est-à-dire autonome au sens fort pour Mel'čuk [1993], pleinement autonome à droite comme à gauche pour Fradin [1993] —, le syntagme peut se définir, suivant la tradition, comme une expression libre qui n'est pas minimale et qui, de surcroît, est autorisée par les règles de la syntaxe. En bref, le syntagme (catégorisé G, groupe) englobe les sous-expressions libres qui ne sont ni des phrases ni des W_{MAX} ²⁶ tout en étant un assemblage de plusieurs unités atomiques (W_{MIN} ou W_{MAX}). Par souci de complétude, je rappelle que la phrase (catégorisée S) est l'unité syntagmatique maximale à laquelle a affaire la syntaxe.

Alors que le mot-forme vient d'être caractériser de manière assez précise, il n'est pas possible de faire de même pour les affixes et les clitiques, les deux autres catégories effectives recensées en (20). La raison en est que la définition et la représentation de celles-ci ne peut se faire qu'après avoir résolu un certain nombre de problèmes (d'analyse et de méthode) concernant en particulier la flexion (par rapport à la dérivation), la nature des objets auxquels s'appliquent les règles morphologiques, le type d'opérations qu'elles peuvent exécuter, etc. Il est hors de question d'aborder ces problèmes ici (cf. Zwicky [1985], Dressler [1989], Anderson [1992], Fradin [1993], Miller *et al.* [1992], Rainer [1995] pour quelques éclairages).

²⁶Cependant les W_{MAX} peuvent être des syntagmes. Sur cette asymétrie cf. § 5. 1. De notre point de vue, il n'y a aucun sens à parler de syntagme pour décrire la constitution interne des mots-types (comme le fait [Gruaz, 1987, p. 30]). Le syntagme reste une unité syntaxique supérieure au mot-type.

Arrivé à ce point, nous en avons fini avec la question de la délimitation syntagmatique des unités. Les questions qui se posent concernent maintenant le stockage des unités (cf. § 2. 2.). La première à résoudre est : sous quelle forme les mots-formes doivent-ils être enregistrés dans le lexique ? La seconde découle de l'existence même des unités atomiques non minimales : doit-on les faire figurer dans le lexique et sous quelle forme ? Cette interrogation pose la question de la lexicalisation, laquelle, par ricochet, amène celle de la construction des unités lexicales et de l'instance grammaticale qui s'en charge. En un mot, s'il est vrai qu'un des buts de la morphologie est de construire les mots-types, cela implique-t-il que toutes les unités figurant dans le lexique seront construites par la morphologie ? C'est à ces questions que les sections suivantes se proposent d'apporter quelques éléments de réponse.

4. La délimitation paradigmatique des énoncés

4. 1. **Le lexème.** Dans l'énoncé (24) la typographie dénombre seize éléments segmentaux (qui correspondent aux mots de la tradition grammaticale scolaire), mais seulement treize segments-types, à savoir : *l', amie, de, sa, mère, porte, la, coiffe, que, portaient, les, amies, grand :*

- (24) L'amie de sa mère porte la coiffe que portaient les amies de sa grand-mère.

Pour cet énoncé, les procédures présentées dans la section précédente retiendraient les sept mots-formes suivants : *amie, mère, porte, coiffe, portaient, amies, grand-mère*. Or il est clair que *amie* et *amies*, *porte* et *portaient* ne sont que des variantes flexionnelles de la même unité lexicale. Le concept de lexème sert justement à capter cette identité. Il est une entité abstraite obtenue à partir des mots-formes que dégage l'analyse. Le lexème a été défini comme l'élément lexical auquel appartiennent les unités formant paradigme (cf. Bally [1965, p. 287nt], Trnka [1949, p. 28], Matthews [1972, p. 160 ; 1974, p. 25], cf. aussi Lyons [1977]). Pour la notion de paradigme, je reprendrai la définition qu'en donnent Fabri et Wunderlich²⁷ :

“Un paradigme est un espace à **n-dimensions** dont les dimensions sont les attributs (ou traits) utilisés pour classer les formes lexicales (*word forms*). Pour constituer un attribut, une dimension doit avoir au moins **deux valeurs**” [Wunderlich, Fabri, 1993 (ce sont les auteurs qui soulignent)].

Ici “dimension” correspond à ce que Mel'čuk [1993, p. 261] appelle *catégorie*, à savoir un ensemble maximum de significations qui s'excluent mutuellement dans une même position (sémantique ou logique). Les

²⁷Cette définition correspond à celle de paradigme₁ de Carstairs-McCarthy [1994, p. 739]. Les listes de formes qui constituent les paradigmes concrets d'une langue sont appelées paradigme₂ par cet auteur.

paradigmes typiques étant les paradigmes flexionnels (cf. la définition de paradigme par Carstairs [1987, p. 48-49]), les dimensions en question seront des dimensions flexionnelles. En français par exemple, pour les noms, les dimensions pertinentes sont celle du genre (notée GER ou ger), du nombre (NB, nb) et de la personne (PER, per). En allemand ou en russe, il faudrait y ajouter celle du CAS. Ces dimensions peuvent concerner des unités qui ne sont pas des lexèmes, c'est-à-dire qui ne se réalisent pas comme mots-formes dans les énoncés. Par exemple, (25) illustre quelques-unes des variations flexionnelles affectant l'article en français, qui est un clitique pour la morphologie et un grammème pour la grammaire (cf. Pottier [1974], Mel'čuk [1994] sur ce concept) :

- (25) LE [ger:mas, nb:sg] = le
 LE [ger:fém, nb:sg] = la
 LE [ger:X, nb:plu] = les

En (25), GER, NB sont des traits flexionnels (ou sèmes génériques dans les approches structuralistes) exprimant les dimensions pertinentes, et FÉM, SG, PLU etc. sont des valeurs de ces traits (les sèmes spécifiques)²⁸. Je note en petites capitales les lexèmes (et les grammèmes). Les corrélats linguistiques mis en jeu par les lexèmes demandent qu'on élabore une terminologie appropriée, nécessairement lourde parce que précise. Je distinguerai donc (cf. Zwicky [1992] et surtout Fradin [1994] pour une discussion argumentée) :

²⁸A propos de ces dimensions, je n'aborderai pas ici les questions suivantes (voir [Mel'čuk, 1994]) : comment établir ces dimensions/catégories ? Comment en établir les valeurs ? Comment déterminer le domaine de validité de ces dimensions ? (lexical vs grammatical ; situations où le trait n'est pas pertinent).

- **Occurrence/lexème** : l'occurrence d'un constituant et le lexème dont elle relève. Exemples : "porte", "portaient" ∈ PORTER ; "l", "la", "les" ∈ LE.

- **Forme-phonol** : la notation sous forme phonologique de la face phonique d'une expression. Exemples : /pɔʁte/, /ami/.

- **Traits-flx** (traits flexionnels) : les traits enregistrant le marquage flexionnel : GER (genre), NB (nombre), PER (personne), CAS (cas), FVT (flexion verbale temporelle).

- **Trait-valué** : un trait auquel est assignée une valeur. Exemple [GER:MAS].

- **Ensemble-valué** : un sous-ensemble cohérent de traits munis de leur valeur. Exemple : [GER:FEM, NB:SG, PER:3].

- **Forme-graph** : la forme graphémique d'une expression (sa notation orthographique). Exemple : 'portera' (généralement en italiques lorsqu'on mentionne l'expression : *portera*).

- **Forme-morph** (forme morphologique) : le lexème en tant qu'il est pourvu d'un ensemble de traits valués. Exemple : 'PORTER [fvt:fut, nb:sg,

per:3]' = le verbe *porter* fléchi à la troisième personne du singulier du futur.

• **Forme-valorée** : toute unité munie d'un ensemble de traits valorés. Ex. /pɔrtərə/ associé à [FVT:FUT, NB:SG, PER:3] et *portera* associé au même ensemble-valoré.

Deux remarques importantes. La forme-morph est associée à la forme-phonol /pɔrtərə/ aussi bien qu'à la forme-graphique 'portera' dans la mesure où, quand on dit [pɔrtərə] ou qu'on écrira *portera*, ce à quoi il est fait allusion est souvent ni la forme phonologique ni la forme-graphématique mais l'une ou l'autre de celles-ci en tant qu'elle est associée à l'ensemble-valoré [fvt:fut, nb:sg, per:3]. Il n'est pas toujours possible d'associer une forme-morph à une (ou des) forme-valorée. Par exemple, 'FRIRE [fvt:pst, nb:plu, per:1]' ne correspond pas à aucune forme-phonol ni forme-graph en français (le verbe est déficient pour cet ensemble-valoré).

L'ensemble des variations qu'on observe entre phonétique (forme-phonol), graphie (forme-graph) et valeurs assignées aux traits flexionnels est résumé en (26) (≠ : indique qu'il y a une variation pour l'aspect mentionné ; id : qu'il n'y en a pas) :

| (26) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------------|---|----|----|----|----|----|----|----|
| ens-valoré | ≠ | ≠ | ≠ | id | ≠ | id | id | id |
| f-phonol | ≠ | id | ≠ | ≠ | id | ≠ | id | id |
| f-graph | ≠ | ≠ | id | ≠ | id | id | ≠ | id |

Les huit cas de figure recensés se rencontrent avec plus ou moins de fréquence dans les langues. Le premier illustre la variation maximale e. g. it. *albero*, /albero/, [ger:mas; nb:sg, per:3] vs *alberi*, /alberi/, [ger:mas, nb:plu, per:3], et le dernier l'absence de variation e. g. *arrhes*, /ar/, [ger:fem, nb:plu, per:3] (cf. Fradin [1993, p. 10-11 pour le détail]).

Jusqu'à présent, j'ai éclairci les rapports qu'entretient la notion de lexème vis-à-vis de la flexion. La compréhension de ces rapports est cruciale pour définir le lexème. Cependant, la définition de ce dernier fait intervenir d'autres caractéristiques, qui relèvent des plans phonologique (cf. (27c)) ou sémantique (cf. (27b)) :

(27) Le lexème

- (a) est un mot-forme débarrassé de ses variations de forme (i. e. flexionnelles)
- (b) il est doté d'une signification stable et unique ou d'un réseau de relations stables et uniques
- (c) il est doté d'une représentation phonologique et est prosodiquement autonome
- (d) il est listé dans le lexique = il appartient à une liste ouverte et/ou sert d'unité de base pour la construction d'autres lexèmes²⁹.

²⁹Cette dernière caractéristique n'exclut pas que des lexèmes soient construits à partir de grammèmes, cf. russe populaire *ixnij* < *ix* 'eux.GEN.PLU' + *nij* 'sfx adjectivant' = "d'eux" e. g. *s ixnim* synom "avec leur fils".

(27b) renvoie au fait qu'au lexème est associée une signification sémantique (ou plusieurs corrélées dans le cas de lexèmes polysémiques). (27c) interdit que des clitiques soient des lexèmes, puisqu'ils sont dépourvus d'autonomie prosodique par définition. Quant à (27d), il signifie que si la notion de lexique a un sens, ce dernier contient au moins les lexèmes. Cette définition du lexème est en accord pour l'essentiel avec celles qui ont cours généralement (cf. Dell [1979], Mel'čuk [1993])³⁰.

³⁰Mais pas Cruse [1986, p. 77] pour qui le lexème est "une famille d'unités lexicales".

4. 2. Le format des lexèmes. Le format des lexèmes définit la manière dont sont stockées les données linguistiques figurant dans le lexique. Il appartient au dispositif de la grammaire dans la mesure où des règles morphologiques opèrent à la fois à partir de lui et sur lui. Les lexèmes simples serviront de point de départ à sa présentation³¹.

³¹Cette section reprend en partie Fradin [1995b].

La plupart des linguistes s'accordent à reconnaître que les informations contenues dans un lexème (ce qu'on appelle *entrée lexicale* dans la tradition de la Grammaire générative) se rangent sous trois rubriques, correspondant à des champs conceptuels différents. Une rubrique phonologique qui fournit une représentation de la mélodie intonatoire conforme au système phonotactique de la langue et inclut, le cas échéant, un schème accentuel ou tonal. Une rubrique sémantique qui donne la signification de l'unité lexicale, c'est-à-dire une interprétation stabilisée, et reconnue comme telle, et différenciée dans un réseau de relations sémantiques. Dans l'idéal, cette rubrique doit fournir les informations permettant aux mécanismes sémantiques de la grammaire de tirer les bonnes inférences des énoncés où figure le mot-forme correspondant au lexème en question. Enfin une rubrique donnant la catégorie syntaxique, c'est-à-dire une information ayant trait aux propriétés combinatoires de l'unité. Il existe une corrélation entre l'information sémantique et l'identification catégorielle. Les catégories syntaxiques majeures correspondent de façon privilégiée à certains typages sémantiques : prédicat pour les verbes et les adjectifs, argument pour les noms, foncteur pour les prépositions (cf. Chierchia [1984])³². De plus, tout changement catégoriel entraîne normalement un changement sémantique, cf. *pris*, *prise* participe passé et *prise* nom, *devoir* verbe, *devoir* nom, etc. Cette corrélation suggère que la catégorie syntaxique pourrait être vue comme un codage qui tient compte des propriétés sémantiques mais joue non plus au plan sémantique mais au plan de la combinaison et de l'arrangement des unités de la chaîne syntagmatique.

³²Cette correspondance est explicite et formalisée dans les systèmes du type de la Grammaire catégorielle.

Si l'on admet que les signes linguistiques sont constitués avant tout par une association son/sens (la catégorisation venant en sus), les rubriques mentionnées ci-dessus constituent la base du signe linguistique, les éléments sans quoi ce dernier cesse d'exister. Cependant l'entrée lexicale doit fournir également des informations conceptuellement moins centrales

mais pourtant indispensables parce qu'idiosyncrasiques et irrégulières. Par exemple celles concernant les classes de déclinaison ou d'accord, les possibilités d'allomorphie, les formes supplétives etc. (cf. Aronoff [1994] et Fradin [1993] pour des exemples). Dans les langues dotées d'une écriture, une rubrique graphémique notant la transcription orthographique de l'unité s'avère en général nécessaire³³. En résumé, le format d'un lexème doit distinguer les cinq rubriques suivantes : (G) graphémique, (F) phonologique, (SX) syntaxique, (M) morphologique et (S) sémantique. Il s'agit de rubriques conceptuellement distinctes, ce qui n'interdit pas de les subdiviser ou éventuellement de les regrouper pour des raisons pratiques ou formelles (cf. Briscoe *et al.* [1993], Mathieu-Colas [1994, annexe]). Cela dépend en grande partie des buts qu'on s'assigne et du système de représentation qu'on se donne (système de traits uniquement ou système hybride). Pour le lexème *mer*, cela nous donne schématiquement (28) :

| | |
|------|------------------|
| (28) | |
| (G) | mer [#] |
| (F) | (mɛr) |
| (SX) | catsx:n, ger:fém |
| (M) | allomorp :RA30 |
| (S) | mer' |

Le dièse note la forme citationnelle lemmatisée selon les conventions en usage ; catsx = catégorie syntaxique, ger = genre ; (M) indique que l'unité est sujette à la règle allomorphique /ɛ/ → /a/ (numérotée arbitrairement RA30) dont l'effet est de produire l'allomorphe (mar) utilisé dans les dérivés e. g. *marin*, *maritime* (cf. Fradin [1993, ch 7 pour le détail] ; mer' note sous forme d'un prédicat en logique intensionnelle la signification du lexème³⁴ sans toutefois l'épuiser (des informations pragmatiques et le typage en font aussi partie). Chaque rubrique constitue une strate indépendante.

Arrivé à ce stade, il nous est possible de définir les éléments (20) qui ne l'étaient pas comme des lexèmes radicaux, c'est-à-dire des lexèmes dont la forme-phonol (le contenu de la rubrique (F)) n'est jamais celle d'un mot-forme attesté. La représentation qu'on pourrait donner du radical de *démontrable*, *démontrer* correspond à la première colonne de (29) :

| | | |
|------|------------------------|------------------------|
| (29) | I | II |
| (G) | démontrer [#] | °démonstr [#] |
| (F) | (demɔ̃tr) | (demɔ̃str) |
| (SX) | catsx:v | |
| (M) | gconj:l | catsx:¬v, rés:sav |
| (S) | démontrer' | |

(gconj = groupe de conjugaison ; cette information peut être assignée par défaut.) Suivant Plénat [1981], je considère qu'il n'y a que deux groupes

³³Elle est indispensable pour les langues où l'écriture est (partiellement) idéographique (chinois, japonais, coréen) car elles jouent sur les deux systèmes, cf. Tamba [1986]. D'autres arguments en faveur de cette rubrique sont donnés dans Fradin [1993, p. 251-252]. Rien ne sera dit ici du rapport entre les rubriques (F) et (G).

³⁴En fait, cette formulation est abusive puisqu'elle reste muette sur les éléments qui constituent cette signification (sèmes ou prédicats, etc.), sur son organisation, ainsi que sur la manière dont cette signification se différencie dans le réseau sémantique auquel, le cas échéant, elle appartient. Cette présentation sommaire ne préjuge pas du type – ou des types – de sémantique dont on a besoin.

de conjugaison en français. Par défaut, les rubriques de II non spécifiées ont un contenu identique à celles de I. La seconde colonne donne la variante supplétive du lexème employée pour construire les dérivés savants à partir du verbe, cf. *démonstration, démonstrateur, démonstratif*. Le trait rés:savant = "réservé (au dérivés) savants", garantira que les lexèmes dérivés savants construits sur le verbe prendront le radical de la colonne II (sur la gestion de ce type de trait et la supplétion, cf. Fradin [1993, 1996a]). Les paires du genre *roi /rwa/, roy- /rwaj/* recevraient un traitement similaire.

Rien n'a été dit jusqu'à présent du sort réservé aux unités indépendantes formées d'éléments dépendants structuraux (i. e. de formants résiduels tels que /pɛr/, /səvwar/, cf. (19)). En tant qu'il constitue un mot-forme, *percevoir* correspond à un lexème dont la représentation sera (30) (sur la notation des traits, cf. Fradin [1993]) :

| (30) | I | II | III |
|------|------------------------|------------------------|--------------------|
| (G) | percevoir [#] | °perçœiv- | °percep- |
| (F) | (pɛrsœv) | (pɛrswav) | (pɛrsɛp) |
| (SX) | catsx:v | per:3, nb:plu, fvt:pst | catsx:→ v, rés:sav |
| (M) | gconj:2 | | |
| (S) | percevoir' | | |

Le lexème PERCEVOIR présente deux radicaux verbaux, le second étant utilisé à la 3^e personne du pluriel du présent (cf. rubrique (SX) colonne II). La colonne III donne le radical supplétif employé avec les dérivés savants (*perception, perceptible*). Ce que ne dit pas la représentation (30), c'est qu'on peut distinguer dans le lexème en question plusieurs parties, dont certaines se retrouvent dans d'autres lexèmes (*recevoir, décevoir, concevoir, apercevoir; remettre, démettre, commettre, admettre; conduire, déduire*). En bref, deux questions se posent : doit-on noter la structure des lexèmes où figurent des paléomorphèmes ? Si oui, comment le faire ?

En admettant que la première question reçoive une réponse positive, plusieurs solutions sont envisageables pour la seconde. Soit cette structure est indiquée directement dans la représentation du lexème. On peut alors imaginer qu'une rubrique supplémentaire enregistre ce fait en donnant les formants du lexème. Mais dans ce cas, pourquoi ne pas le faire pour tous les lexèmes complexes ? (cf. la rubrique Archive dans Fradin [1993]). On peut aussi recourir à des patrons lexématiques dont l'objet spécifique est d'indiquer ce type de structuration. L'intérêt de cette seconde démarche tient au fait que ces patrons pourraient être employés aussi pour les phonesthèmes. A titre d'illustration, la série de verbes à laquelle appartient PERCEVOIR mettrait en jeu le patron (31) joint à la spécification (32a) :

| (31) | I | II | III |
|------|-------------------------|-------------------------|-----------------------|
| (G) | (X') ^o cev-# | (X') ^o coiv | (X') ^o cep |
| (F) | (X)(səv) | (X)(swav) | (X)(sep) |
| (SX) | catsx:v | per:3, nb:plu, fvt :pst | catsx:¬ v, rés:sav |
| (M) | gconj:2 | | |
| (S) | | | |

(32) Classes de paléomorphèmes préfixés (non exhaustives)

- (a) X = {(k̄s), (de), (rə), (apɛr)}
- (b) Y = {(k̄s), (de), (re), (se), (pro), (tra)}

(En (31), X' note le correspondant graphémique de X). La spécification (32b) serait employée pour les lexèmes de la série CONDUIRE, dont je ne donne pas le patron lexématique ici. Ce modèle peut être élargi aux autres séries de verbes de ce type (*permettre*, etc.).

Les phonesthèmes sont des éléments idéophoniques apparaissant dans plusieurs lexèmes, dont le substrat est un fragment de la mélodie intonatoire de ces lexèmes et dont l'abstrat est une signification qui s'ajoute à la signification propre de chacun³⁵. Selon Tournier [1985, p. 155], les éléments idéophoniques se distinguent des éléments onomatopéiques par le fait (i) qu'ils ne relèvent pas du domaine de perception acoustique, (ii) qu'ils constituent des ensembles qui ne sont plus ouverts. L'un et l'autre partagent la propriété d'avoir pour substrat un fragment de la représentation phonologique d'un lexème. Voici quelques exemples anglais (tirés de Tournier ; pour le français, cf. Guiraud [1967]) :

- (33) (a) /kl/ : "son plat souvent métallique" :
clang, clank, clap, click, clash, clink.
- (b) /gl/ : "reflet, lumière, surface lisse" :
glisten, glimmer, glare, glow.
- (c) /ʌmp/ : "masse molle ou arrondie" :
bump, hump, lump, rump, stump.

³⁵ "Un élément idéophonique est un segment submorphématique du signifiant d'un mot dont le signifié peut appartenir à divers champs notionnels, à l'exception de celui des sons ; ce segment est commun à une série de mots et associé au sens général commun à tous les mots de la série", Tournier [1985, p. 145] (c'est l'auteur qui souligne). Sur les rapports entre phonétique et signification, cf. [Fónagy, 1979].

Pour (33a), on peut proposer le patron lexématique suivant :

- (34) (G) (kl...)
- (F) (kl...)
- (SX) catsx :v
- (M)
- (S) "son plat souvent métallique"

En d'autres termes, tous les verbes de l'anglais dont la représentation phonologique commence par /kl/ auront la signification globale indiquée en (S). Cette manière de voir les choses permet de décrire les phonesthèmes ou les paléomorphèmes sans qu'on soit commis à dire

³⁶Ceci suppose une définition et une représentation du morphème. Cette question sera brièvement abordée au § 6. (cf. [Fradin, 1993, ch 4 et 5] pour une discussion). L'existence de patrons comme (32) signifie que certaines propriétés lexicales seront notées en dehors du lexique (conçu comme répertoire des unités lexicales). Ceci vaut aussi de propriétés sémantiques que partagent plusieurs lexèmes. Le lexique n'épuise donc pas le lexical.

³⁷Parmi ces procédés, il faudrait rajouter l'incorporation. Dans la mesure où il n'existe pas dans les langues romanes, je n'en dirai rien ici. Le statut de l'incorporation (morphologique vs syntaxique) reste ouvert à discussion (cf. [Mithun, 1984] pour une présentation). Les procédés mentionnés en (35) n'ont pas tous la même fréquence d'emploi. Le cas (36f) pose des problèmes particuliers, cf. David [1993].

qu'ils sont des morphèmes³⁶. Les deux sont des formants (ils passent le test (15)) mais se distinguent l'un de l'autre par deux aspects : les premiers ne sont pas des unités phonotactiquement bien formées (les seconds le sont) et ils sont associés à un potentiel de signification (cf. (34)) qui fait défaut aux seconds.

5. Typologie des unités lexicales

La section précédente a avancé quelques suggestions concernant la manière dont les lexèmes (qui sont les correspondants des mots-formes) sont enregistrés dans le lexique. Cette dernière section (basée sur Fradin [1993, 1995b]) voudrait brièvement répondre à la double question posée à la fin du § 3. 4., à savoir : qu'est-ce qui fait qu'une unité figure dans le lexique et quel est le rapport entre les unités qui y figurent et la morphologie ?

5. 1. Construction des lexèmes vs lexicalisation. Les unités stockées dans le lexique peuvent être simples ou complexes. Les lexèmes complexes peuvent être construits ou non construits. La construction des lexèmes renvoie à deux choses. D'une part aux procédés formels, auxquels on attribue en général le statut de règle, disponibles en langue pour produire de nouvelles unités lexicales. D'autre part, de manière plus statique (cf. § 2. 2.), au fait que ces lexèmes sont analysables en formants dont l'agencement et les rapports se retrouvent dans d'autres lexèmes. Cette analysabilité doit pouvoir être mise en évidence par des manipulations du type de celles décrites au § 3. Les procédés formels en question sont avant tout morphologiques (35), mais il peuvent être syntaxiques (36)³⁷ :

| (35) Type | Exemples |
|--------------------|--|
| (a) dérivation | salage, dénucléarisation ; fleurdéliné, moyenâgeux |
| (b) conversion | orange, dégainé ; double-cliquer, court-circuiter |
| (c) reduplication | guéguerre, babase, papatte, fils |
| (d) composition VN | lance-flammes, rabat-joie, ouvre-boîte |
| (e) composition NN | poisson-lune, voiture-balai, jupe-culotte, bateau-lavoir |
| (36) Type | Exemples |
| (a) synapsie | avion à réaction, heure de pointe, moulin à vent |
| (b) syntagme AN | moyen-âge, blanc-bec, grand attracteur, faux-frais |
| (c) syntagme NA | peau-rouge, messe-basse, poids-lourd, chambre froide |

- | | | |
|-----|----------------|---|
| (d) | syntagme autre | <i>tourne-à-gauche, va-et-vient, dommages et intérêts</i> |
| (e) | phrase | <i>sot-l'y-laisse, m'as-tu-vu, décrochez-moi-ça</i> |
| (f) | juxtaposition | <i>langage auteur, opération prix, case départ</i> |

Les procédés morphologiques de dérivation, de conversion et de reduplication se caractérisent par le fait qu'ils opèrent sur un seul lexème (de format (28) dans l'optique présentée ici), alors que la composition opère sur deux, le résultat étant dans les deux cas un lexème (de format (28)). La distinction entre (35) et (36) n'a évidemment de sens que si la morphologie se distingue de la syntaxe par les mécanismes qu'elle offre pour construire des unités complexes. Alors que la première construit des lexèmes, la seconde construit des syntagmes. Ce qui implique que les règles syntaxiques ne sont pas à même de construire des lexèmes de type (35d, e)³⁸. Très souvent, les unités (35d, e) et (36) sont regroupées sous l'appellation de noms composés au double motif (i) qu'elles fonctionnent comme des noms, (ii) que leur interprétation n'est pas compositionnelle mais idiosyncrasique (cf. Silberztein [1990, 1993], Gross [1990] par exemple), ce qui est vrai, même si dans bien des cas cette interprétation n'est nullement arbitraire. Il n'en reste pas moins que, du point de vue de leur structure interne (cf. (3)), ces unités ne sont pas identiques et que l'agencement des éléments qui les composent n'a pu être réalisé par le même corps de règles. En fait, les unités de (36) sont des syntagmes lexicalisés (mais voir note 38 ; Mel'čuk [1993] dit "phraséologisés" et les nomme *syntagmèmes*)³⁹. Ce qui nous amène à la question de la lexicalisation. Mais auparavant je vais dire un mot des lexèmes complexes non construits.

En contraste avec ce qui vient d'être dit, ces lexèmes seront des lexèmes formés de plusieurs parties dont la combinaison ne peut être décrite — ou pas totalement — par les règles de la langue. Parmi ceux-ci, il faut compter les lexèmes complexes non construits proprement dits (cf. [Corbin, 1987] comme *royaume* ou *jasserie*, dans lesquels on retrouve des formants résiduels (cf. (19c)). Suivant que les mécanismes qui construisent les mots-valises font partie ou non de la grammaire on y inclura ou non les mots-valises e. g. *foultitude, éternullité, photocopillage*, etc. Autres candidats à cette catégorie, des mots-formes comme *zyeuter, nono* (*zone nono* i. e. "zone non occupée", pendant la dernière guerre) et quelques autres du même acabit.

Suivant Corbin [1992], je considère que la lexicalisation est le fait qu'une expression linguistique accède au statut d'entité codée, présentant une association réglée du rapport son/sens reconnue et employée comme telle. La lexicalisation d'une expression dépend de nombreux facteurs sociologiques et s'étend sur une durée plus ou moins longue [Guilbert 1975, p. 49]. A un niveau très général, elle s'inscrit dans la relation entre fonctions sociales de la langue et fonctions linguistiques de la société (cf.

³⁸Des arguments pour ce point de vue sont donnés dans Corbin [1992] et Fradin [1993, 1995b] notamment. Mais la distinction entre syntaxe et morphologie reste au cœur des débats actuels (ou bien n'est pas faite). La question de savoir si les synapsies classiques de type (35a) et les juxtapositions (35f) relèvent de la syntaxe ou de la morphologie mérite aussi discussion (cf. Vivès [1990], David [1993]). Il en va de même de l'étendue de la composition morphologique. Doit-elle inclure les patrons qui permettraient de construire des lexèmes comme *malentendant, clairvoyant ou encore plus-value, contretemps, arrière-plan* ? (cf. Riegel et al., [1994, p. 548-549] pour d'autres cas). Je laisse la question ouverte. Notons que les procédés énumérés en (35), (36) ne sont pas productifs au même degré.

³⁹D'un point de vue terminologique, il est plus cohérent et plus clair de restreindre l'emploi de nom-composé aux seuls cas (35d, e) et d'appeler les exemples (36) unités polylexicales (ou mieux unités polylexématiques), ce qui est une appellation purement descriptive qui s'en tient au niveau U de (2).

⁴⁰L'idée est que la signification des éléments composant une expression (polylexématique ou pas) et la manière dont ils se combinent ouvrent certaines possibilités, et pas d'autres, qu'exploitent les schémas cognitifs (métonymie, saillance, ressemblance, typification, etc.). En bref, la signification que fixe une expression lors de sa lexicalisation n'est pas complètement arbitraire mais a un coût qu'on peut tenter d'évaluer. Par exemple, compte tenu des lexèmes en jeu, il est plus plausible que chaise longue dénote ce qu'il dénote que, par exemple, une sorte de ver à bois ou une maladie du cheval.

⁴¹La productivité morphologique a fait l'objet d'un grand nombre d'études. Du point de vue qualitatif, la productivité d'une règle morphologique est inversement proportionnelle au nombre de facteurs qui la conditionnent, lesquels peuvent être d'ordre syntagmatique (phonologiques, syntaxiques, etc.) ou paradigmatiques (cf. Booij [1977], Van Marle [1985]). Sur la mesure quantitative de la productivité, cf. Baayen [1991].

⁴²"La néologie lexicale se définit par la possibilité de création de nouvelles unités lexicales, en vertu des règles de production

Calvet [1996, p. 20-23]). Elle suppose des instances sociales (institutions savantes, académies, dictionnaires etc.) qui avalisent l'entrée de telle ou telle expression dans le lexique. Indépendamment de cette face sociale, la lexicalisation présente une face linguistique sur laquelle je ne m'étendrai pas⁴⁰. Dans l'optique développée ici, je fais l'hypothèse que la lexicalisation est consubstantielle au fait d'acquérir une forme lexicale, c'est-à-dire justement un format de stockage des données linguistiques sur le modèle de (28). Ce point de vue permet de rendre compte de deux faits : (i) que les expressions (22) fonctionnent comme unités atomiques pour la syntaxe ; (ii) que ces expressions puissent, sous certaines conditions, servir de lexèmes-bases au processus de dérivation (e. g. *fleurdelisé*, *capcorsin*, cf. Fradin [1995b]). Il permet également d'éviter les problèmes que soulève la composition dans certains cadres [Di Sciullo, Williams, 1987].

5. 2. Créativité lexicale et productivité lexicale. La distinction entre ces deux concepts a été remise en avant par Van Marle [1985]. La productivité est la capacité à créer des expressions, qui ont vocation à devenir des unités lexicales, en recourant aux moyens formels qu'offre la langue pour construire des lexèmes ou des expressions (cf. (35), (36))⁴¹. La productivité s'inscrit du côté du régulier, du systématique, du formellement non marqué. Cela signifie notamment que les expressions nouvellement produites s'intègrent en général parfaitement dans les paradigmes existants, au point qu'il est difficile de sentir leur nouveauté linguistique. Par exemple des mots-formes tels que *déjudiciarisation* ou *génocideur*, *saisonnalité* sont compris d'emblée en contexte. Pour le premier, il est difficile de dire à sa première rencontre s'il est une création récente ou s'il s'agit simplement d'un mot-forme qu'on ne connaissait pas. La créativité lexicale en revanche s'affranchit des règles servant à la construction des unités lexicales. Elle ne met pas en œuvre — ou pas uniquement — des procédés appartenant à la grammaire de la langue (analogie, verlan, etc.). Elle opère au coup par coup, de manière aléatoire, en privilégiant les moyens marqués non reproductibles et, partant, non productifs. Les expressions relevant de la créativité lexicale sont en général fortement dépendantes des circonstances énonciatives et du contexte extralinguistique (cf. les mots d'auteur). La distinction entre productivité et créativité est ancienne. Hjelmslev la faisait, Guilbert [1975] aussi, mais en nommant "créativité" ce que Van Marle dénomme "productivité"⁴². Comme beaucoup d'autres, cette distinction est graduelle même si certains seuils semblent bien marqués.

Les distinctions que je viens de rappeler entretiennent les relations suivantes.

1. Lexème construit n'entraîne pas lexème lexicalisé. Par exemple, les expressions (37) sont bien construites au regard des règles du français

mais n'appartiennent pas pour autant à son lexique (la pastille note qu'une unité possible est non attestée) :

- (37) °lance-tomate, °pèse-neurone, °saute-loup ; °requin-faucille
 °déchacaliser ("supprimer les chacals", cf. *dératiser, décafardiser*)
 °amassage ; °articlée, °verrée, °arbrée
 °tour d'âne (vs de cochon), °passage à l'action

2. Productivité implique (par définition) construction au moyen de procédés réguliers. Ces procédés sont des règles de la grammaire (morphologiques ou syntaxiques, cf. (35), (36)). Au niveau des unités, cela suppose que les lexèmes complexes sont analysables (selon plusieurs plans, cf. (28)), ce qui permet de les assigner à des séries (paradigmes) sur la base d'une similitude d'analyse plus ou moins totale, c'est-à-dire concernant tous les plans ou seulement certains.

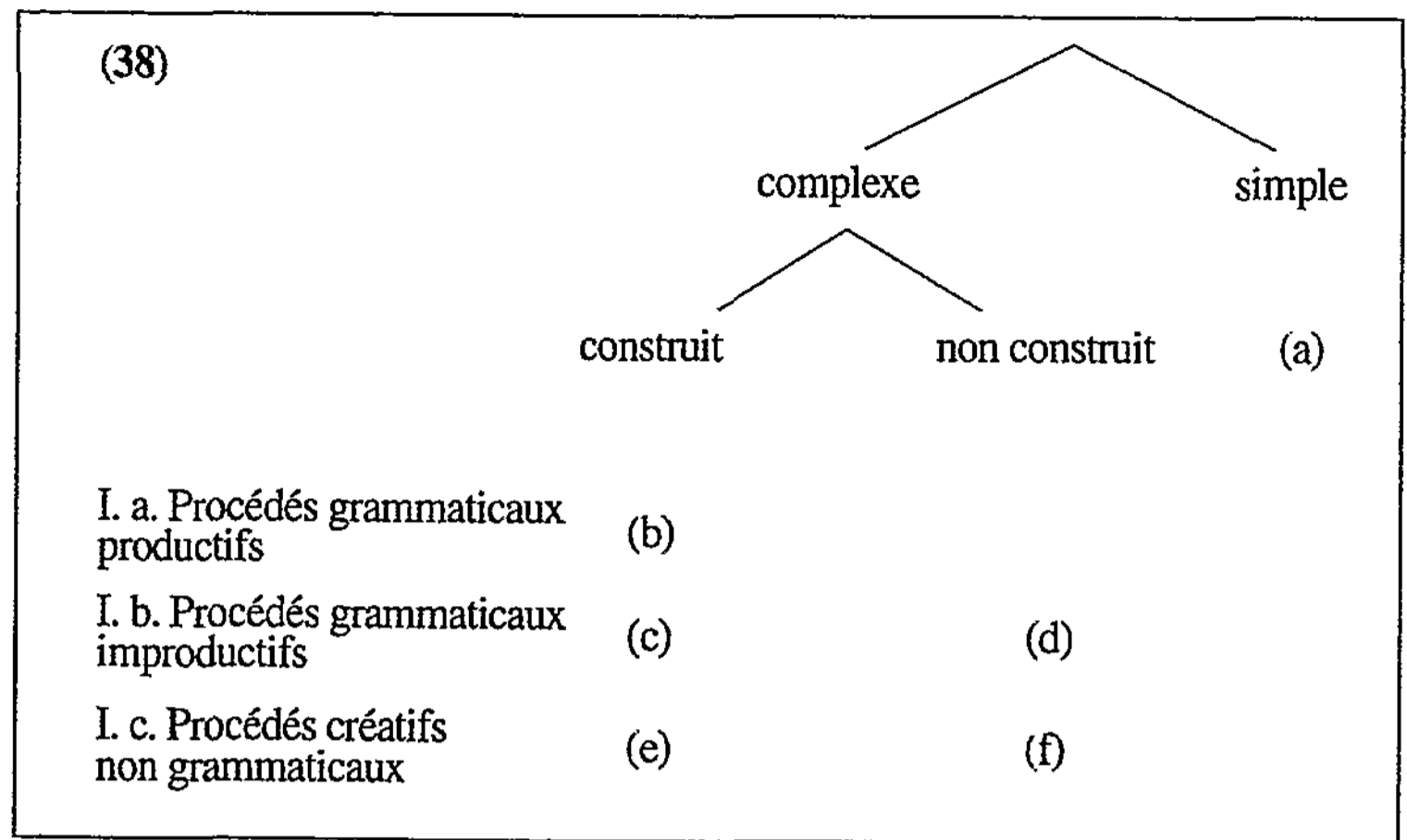
3. Lexème construit va généralement de pair avec lexème issu du système productif de la langue, dans la mesure où la productivité inclut l'usage des règles de construction des unités lexicales. Cela se vérifie notamment toutes les fois que le modèle de construction utilisé est disponible dans la grammaire de la langue.

4. Lexème lexicalisé n'entraîne pas lexème construit. Comme le remarquait Guilbert [1975, p. 43], il peut se faire qu'une expression créée en dehors du système des règles de la grammaire, de manière non systématique et visant un effet plus ou moins marqué "puisse devenir des termes du lexique de la langue". Il peut s'agir de créations analogiques comme par exemple ang. *herstory* (vs *history*)⁴³ ; de créations par amputation, très fréquentes en français : *sécu* (*Sécurité sociale*), *colon* (*colonel*), *pitaine* (*capitaine*), *dégueu* (*dégueulasse*), etc. ; par amputation et suffixation en /o/ *alcoolo*, *intello*, *rapido*, etc.⁴⁴ ; de créations par télescopage (mots-valises) : *foultitude*, *rurbain*, *photocopillage*, par abrègement : *périphérique* (*Boulevard périphérique*), *traction* (*voiture à traction avant*), (*route*) *nationale* ; de sigles ou d'acronymes, lesquels peuvent servir de base à une future dérivation e. g. *radar*, *sida*, *VTT* (*vélo tout terrain*, cf. *vététiste*) ; de créations erratiques, provenant souvent de l'argot, cf. *ripou* (<*pourri*, en parlant de la police), *meuf* (<*femme*) ou de la langue populaire, cf. *deuche* (*deux CV*), *zyeuter*, *nono*, etc. Il peut s'agir enfin d'expressions logofigées (que Benveniste nommait des conglomérats), qui sont des portions d'énoncé ayant acquis une autonomie lexicale (par délocutivité ou autrement) e. g. *sot-l'y-laisse*, *quant-à-soi*, *en-cas*, *tourne-à-gauche*, *rentre-dedans*, *faire-part*, *crève-la-faim*, *monte-en-l'air*, *dame d'onze heures*, etc. (cf. (36d, e)).

incluses dans le système lexical" [Guilbert, 1975, p. 31]. Il admet cependant que des unités créées en dehors des règles d'une langue soient lexicalisées surtout lorsqu'il s'agit de "création poétique ou littéraire" (*ibid.*).

⁴³Cf. Van Marle [1985, p. 46sq] pour des exemples néerlandais. Les expressions en -ique, -tel, -ciel participent de l'analogie mais sont plus liées à la composition, cf. Reboul [1994] et Fradin [1997].

⁴⁴Kilani-Schoch et Dressler [1992] et Kilani-Schoch [1996] distinguent les deux types d'amputation. Voir Groud et Serna [1996] pour une vue documentée sur ces procédés en français. Tout ce qui concerne ce paragraphe devrait être nuancé en distinguant procédés centraux (grammaticaux) et périphériques (extragrammaticaux).



(a) *fil, bois, nuit, érable*. (b) *expédition, pendable ; contre-indication, jupe-culotte, rabat-joie ; dégaine ; guéguerre*. (c) *recevoir, exprimer*. (d) *royaume, jasserie*. (e) *m'as-tu-vu, quant-à-soi, sot-l'y-laisse ; désormais, lorsque, enfin*. (f) *zyeuter, foulditude, ripou, deuche*.

Le diagramme (38) résume les rapports qu'entretiennent les lexèmes du français avec les distinctions qu'on vient d'opérer.

Le point important à retenir est que la lexicalisation d'une expression, c'est-à-dire son inscription dans le lexique, est indépendante du fait qu'elle soit construite ou qu'elle ait été produite par les règles de la grammaire. En d'autres termes, l'origine d'une expression n'a pas d'incidence sur sa lexicalisation. Par contre, une fois lexicalisée, l'expression acquiert le format normal d'un lexème. Il n'en demeure pas moins que les mécanismes réguliers produisant en grand nombre des expressions lexicalisables, le stock des unités lexicales, c'est-à-dire des expressions finalement lexicalisées, s'accroît essentiellement par ce biais.

6. Conclusion

La question de l'identification des unités lexicales a été abordée à partir de la découpe des expressions sur l'axe syntagmatique et de la constitution des unités selon l'axe paradigmatique. Du point de vue de la méthode, j'ai essayé de discriminer les propriétés des expressions sur la base de tests (les procédures) à plusieurs niveaux, puis de constituer les classes d'unités au moyen de définitions fondées sur les propriétés discriminées, et enfin de fournir des représentations susceptibles de rendre compte du comportement de ces unités. Du point de vue des faits, il faut

souligner, de nouveau, que les distinctions établies ont rarement un caractère catégorique et doivent être appréhendées plutôt sur le mode du plus ou moins typique.

Plutôt que de rappeler les étapes que nous avons parcourues, je vais brosser un tableau synthétique des différents termes qui ont été introduits. A un premier niveau, celui des expressions de surface, la distinction s'établit entre mot-forme et expressions dépendantes, ces dernières pouvant être subdivisées par le critère de minimalité :

| (39) | Mot-forme | Expression dépendante | |
|------|-----------|-----------------------|---------|
| | | non minimal | minimal |
| | (a) | (b) | (c) |

(a) CHAT, EXPÉDITION, ÉRABLE, RECEVRA, GARE ; (b) DÉMONTR- (c) PEND-, -TION, -ABLE, LE, SA, DE.

Le second niveau, plus abstrait, est celui des unités morphologiques où l'on trouve les lexèmes, les affixes et les clitiques :

| (40) | Lexème | | | | Affixe | Clitique |
|------|---------|----------|--------|----------|--------|----------|
| | basique | | racine | | | |
| | simple | complexe | simple | complexe | | |
| | (a) | (b) | (c) | (d) | (e) | (f) |

(a) MER, LIVRE, VALISE, FLEUR ; (b) PERSONNEL, MOYEN-AGE, COURT-CIRCUIT, BALLON ; (c) PEND-, MAR-, FLOR-, EXPÉDI- ; (d) DÉMONTR-, DÉMONSTR-, PERSONNAL- ; (e) -ATION, DÉ-, -AGE, -ISER ; (f) LE, SON, NE.

La comparaison des tableaux (20) et (40) montre le lien qui unit mot-forme et lexème. Le lexème se situe en *input* des règles morphologiques, alors que le mot-forme se situe en *output*. Ou encore, pour s'en tenir à une version statique plutôt que générative de la morphologie, le lexème appartient au lexique restreint et le mot-forme au lexique étendu, la tâche des règles étant alors de déterminer s'il existe une corrélation possible pour toute paire d'unités lexème/mot-forme. Il va de soi que dans les langues où beaucoup de lexèmes ont la même apparence phonique et graphique que les mot-formes (e. g. l'anglais), la distinction entre les deux aura plus de mal à être perçue.

Il est sûr que les propriétés sémantiques ne permettent pas de distinguer entre les lexèmes, les affixes et les clitiques⁴⁵. Par contre, au plan prosodique, les lexèmes et certains affixes sont autonomes, alors que les clitiques ne le sont pas. La typologie donnée en (40) ne règle pas la représentation qu'on doit donner des affixes et des clitiques. L'hypothèse

⁴⁵Par exemple, dans certaines langues, notamment amérindiennes (cf. Anderson [1985], Mithun [1996]), ce qui correspond à nos verbes est exprimé par des affixes ou des clitiques. Le critère de référentialité — les lexèmes auraient un "pouvoir référentiel" [Corbin, 1992, p. 33] mais pas les affixes — dépend donc des langues.

définie ici est que les lexèmes se définissent par les propriétés (27) et ont une représentation sur le modèle de (28)-(31). La question de la représentation des affixes et des clitiques peut donc se résumer à la batterie de questions suivante, les secondes étant la version forte des premières :

1. Y a-t-il des (i) affixes / (ii) des clitiques qu'on doit représenter sur le modèle de (28) ?

2. Doit-on représenter les (i) affixes / (ii) des clitiques sur le modèle de (28) ?

Une réponse positive signifie que les affixes ou les clitiques sont des unités lexicales. Même en laissant de côté les problèmes de cohérence interne, il existe des arguments montrant qu'une réponse positive à 1(ii) ne permet pas de rendre compte du comportement des clitiques (cf. Miller [1991], Fradin [1996b] notamment). En ce qui concerne les affixes, la tradition grammaticale et linguistique distingue les flexionnels des dérivationnels⁴⁶. Il existe aussi des arguments pour dire que les affixes flexionnels ne doivent pas être considérés comme des entités lexicales mais simplement comme la réalisation phonique de traits sémantiques (et/ou grammaticaux) sur les lexèmes (cf. Anderson [1992], Fradin [1993] notamment⁴⁷). En revanche la question reste plus ouverte pour ce qui concerne les affixes dérivationnels. Sur le modèle de ce qui se pratique en lexicographie, on a envisagé qu'ils soient des unités lexicales [Hoeksema, 1985 ; Fradin, 1993], munis d'une représentation semblable à (28), mais qui indique qu'ils sont des fonctions et non des arguments. Il faut savoir cependant qu'une telle approche pose des problèmes lorsqu'on se propose de décrire dans le détail n'importe quel phénomène dérivationnel un peu complexe (même pour une langue à la morphologie relativement simple comme le français).

La notion de morphème ne joue pas de rôle dans l'analyse que j'ai donnée des unités lexicales. On peut néanmoins s'en servir à un niveau de description métalinguistique, pour faire référence aux unités morphologiques minimales (une définition plus précise reste requise). Cette unité coifferait les lexèmes et ce que Pottier [1974] nomme grammème⁴⁸.

Quel rapport entretiennent les lexèmes à la notion d'atome syntaxique ? Traditionnellement, la sortie de la morphologie constructionnelle (celle qui construit les unités lexicalisables) produit des unités qui sont des atomes syntaxiques. Dans nos termes, cela s'exprime lexème = W_{MIN} . Dans la mesure où les W_{MAX} doivent être considérés comme des lexèmes, les exemples de (22) ou (35) montrent qu'un lexème peut être constitué de plusieurs W_{MIN} ⁴⁹. Il faut donc moduler le point de vue traditionnel et l'envisager comme le cas par défaut. Notons que l'inverse, c'est-à-dire W_{MIN} = lexème, doit également être pris comme un

⁴⁶Anderson [1992] définit la flexion comme "la morphologie qui est accessible et/ou manipulée par la syntaxe". Dressler [1989] a cependant souligné le caractère graduel des critères distinguant flexion et dérivation. De son côté, Booij retravaille la distinction d'Anderson en montrant qu'elle confond flexion contextuelle et flexion inhérente. Cf. Rainer [1995] pour une discussion.

⁴⁷Dans cette optique, les clitiques peuvent être envisagés comme une affixation sur les syntagmes (et non sur les lexèmes). Ce point de vue autorise une description plus satisfaisante du phénomène de cliticisation.

⁴⁸Dans la mesure où il subsume affixe et clitique, le grammème a un intérêt descriptif mais pas opératoire.

⁴⁹Contrairement à Zwicky [1992, p. 362] et Fradin [1994], je ne pense plus que les morphèmes portemanteaux français au, du, des soient constitués de deux W_{MIN} (respectivement à le, de le, de les). Il s'agit plutôt de la réalisation de traits sous forme d'un clitique, cf. [Miller, 1991 ; Miller et al., 1992].

cas par défaut, car dans beaucoup de langues des W_{MIN} se manifestent morphologiquement sous forme de clitiques simples (cf. le verbe ETRE en anglais (41a) et serbo-croate (41b), cf. note 45) :

- (41) (a) *John's going to help you*
John=ETRE.3SG.PST allant aider vous
"John va vous aider"
- (b) *U Moran sam*
fatigué=ETRE.1SG.PST
"Je suis fatigué"

Les exemples (22) conduiraient naturellement à s'interroger sur le statut des expressions idiomatiques — les phrasèmes — et des énoncés liés par rapport au lexique, puisque l'un et l'autre présentent des aspects idiosyncrasiques et sont des atomes syntaxiques (W_{MAX}). Cette question demandant une ample discussion, elle est laissée pour des études ultérieures.

(Courriel : bfradin@li.univ-paris13.fr)

UMR 195 : CNRS — Université Paris 13
Av. Jean-Baptiste Clément
93430 Villetaneuse)

Références

ANDERSON (S. R.)

1985, "Inflectional morphology", p. 150-201, in SHOPEN (T.), ed.

1992, *A-Morphous Morphology*, Cambridge University Press.

ANSCOMBRE (J.-C.), DUCROT (O.)

1976, "L'Argumentation dans la langue", *Langages*, n°42, p. 5-27.

ARONOFF (M.)

1994, *Morphology by Itself*, Cambridge, The MIT Press.

BAAYEN (H.)

1991, "Quantitative Aspects of Morphological Productivity", *Yearbook of Morphology 1991*, p. 109-149.

BACH (E.)

1983, "On the Relation between Word-Grammar and Phrase-Grammar", *Natural Languages and Linguistic Theory*, 1, p. 65-89.

BALLY (C.)

1965, *Linguistique générale et linguistique française*, Berne, Francke.

BOOIJ (G.)

1977, *Dutch Morphology*, Dordrecht, Foris.

BRISCOE (T.), PAIVA (V. de), COPESTAKE (A.)

1993, *Inheritance, Default and the Lexicon*, Cambridge University Press.

CALVET (J.-L.)

1996, *Les Politiques linguistiques*, Paris, PUF (Que sais-je ?, n°3075).

CARSTAIRS (A.)

1987, *Allomorphy in Inflection*, London, Croom Helm.

CARSTAIRS-McCARTHY (A.)

1994, "Inflection Classes, Gender, and the Principle of Contrast", *Language*, n°70, p. 737-788.

CHIERCHIA (G.)

1984, *Topics in the Syntax and Semantics of Infinitives and Gerunds*, PhD, Amherst, University of Massachusetts.

CORBIN (D.)

1987, *Morphologie dérivationnelle et structuration du lexique*, Tübingen, Niemeyer.

1989, "Form, Structure and Meaning of Constructed Words in an Associative and Stratified Lexical Component", *Yearbook of Morphology 1989*, p. 31-54.

1992, "Hypothèse sur les frontières de la composition nominale", *Cahiers de grammaire*, n°17, p. 25-55.

COURTOIS (B.)

1990, "Un Système de dictionnaires électroniques pour les mots simples du français", *Langue française*, n°87, p. 11-22.

CRUSE (A. D.)

1986, *Lexical Semantics*, Cambridge University Press.

DAVID (S.)

1993, *Les Unités polylexicales : éléments de description et reconnaissance automatique*, Thèse de doctorat, Paris 7.

DELL (F.)

1979, "La Morphologie dérivationnelle du français et l'organisation de la composante lexicale en grammaire générative", *Revue romane*, XIV-2, p. 185-216.

Di SCIULLO (A. M.), WILLIAMS (E.)

1987, *On the Definition of Word*, Cambridge, The MIT Press.

DRESSLER (W. U.)

1989, "Prototypical Differences between Inflection and Derivation", *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung*, 42, p. 3-10.

DUCROT (O.)

1983, "Puisque : essai de description polyphonique", *Revue romane*, n°24 (numéro spécial).

EMONDS (J. E.)

1985, *A Unified Theory of Syntactic Categories*, Dordrecht, Foris.

FEUILLET (J.)

1988, *Introduction à l'analyse morphosyntaxique*, Paris, PUF.

FÓNAGY (I.)

1979, *La Métaphore en phonétique*, Ottawa, Didier.

1982, *Situation et signification*, Amsterdam-Philadelphia, John Benjamins.

FRADIN (B.)

1993, *Organisation de l'information lexicale et interface lexique/syntaxe dans le domaine verbal*, Thèse de doctorat d'Etat, Paris VIII.

1994, "La Théorie morphologique face à ses choix", *Cahiers de lexicologie*, n°63, p. 5-42.

1995a, "L'Approche à deux niveaux en morphologie computationnelle et les développements récents de la morphologie", *TAL*, 35-2, p. 9-48.

1995b, "La Dérivation à partir d'unités polylexicales", *Atti del XXI congresso internazionale di linguistica e filologia romanza*, Palermo, Settembre 1995 (à paraître).

1996a, "On Morphological Entities and the Copy Principle", *Acta linguistica Hungarica*, 43, 1-2, p. 111-151.

1996b, "Deux théories des clitiques : enjeux pour la morphologie", in FRADIN (B.), MARANDIN (J.-M.), MAZIERE (F.), eds.

1997 (à paraître), "Combining Forms, Blends and Related Phenomena", in Doleschal U. & Thornton A. (sld), *Marginal and extragrammatical morphology*, München, Lincom Europa.

FRADIN (B.), MARANDIN (J.-M.), MAZIERE (F.), eds

1997 (à paraître), *Mot et Grammaires*, Paris, Didier-Érudition.

GLEASON (H. A.)

1955, *An Introduction to Descriptive Linguistics* (1961, éd. révisée), New York, Holt, Rinehart & Winston.

GROSS (G.)

1990, "Définition des noms composés dans un lexique-grammaire", *Langue française*, n°87, p. 84-90.

GROUD (C.), SERNA (N.)

1996, *Regard sur la troncation en français contemporain*, Paris, Didier-Érudition.

GRUAZ (C.)

1987, *Le Mot français, cet inconnu*, Publications de l'Université de Rouen.

GUILBERT (L.)

1975, *La Créativité lexicale*, Paris, Larousse.

GUIRAUD (P.)

1967, *Structures étymologiques du lexique français*, Paris, Larousse.

HARRIS (Z. S.)

1951, *Structural Linguistics*, The University of Chicago Press.

HOEKSEMA (J.)

1985, *Categorial Morphology*, New York-London, Garland Publishing Inc.

HURCH (B.), RHODES (R. A.), eds.

1996, *Natural Phonology : The State of the Art*, Berlin-New York, Mouton de Gruyter.

KILANI-SCHOCH (M.)

1996, "Syllable and Foot in French Clipping", p. 135-152, in HURCH (B.), RHODES (R. A.), eds.

KILANI-SCHOCH (M.), DRESSLER (W. U.)

1992, "Prol-o, intell-o, gauch-o et les autres : propriétés formelles de deux opérations du français parlé", *Romanistisches Jahrbuch*, 43, p. 65-86.

KLAVANS (J. L.)

1982, *Some Problems in a Theory of Clitics*, Bloomington, IULC.

1985, "The Independance of Syntax and Phonology in Cliticisation", *Language*, n°61, p. 95-120.

LEJEUNE (M.), ed.

1949, *Actes du sixième congrès international des linguistes*, Paris, Klincksieck.

LEVINE (R. D.), ed.

1992, *Formal Grammar : Theory and Implementation*, Oxford University Press.

LYONS (J.)

1977, *Semantics* (I et II), Cambridge University Press.

MARTIN (R.)

1983, *Pour une logique du sens* (2e éd. 1992), Paris, PUF.

MATHIEU-COLAS (M.)

1994, *Les Mots à traits d'union*, Paris, Didier-Érudition.

MATTHEWS (P. H.)

1972, *Inflectional Morphology*, Cambridge University Press.

MEILLET (A.)

1937, *Introduction à l'étude comparative des langues indo-européennes*, Paris, Hachette.

MEL'CUK (I.)

1993, *Cours de morphologie générale. Première partie : le mot*, Montréal, Presses de l'Université de Montréal-CNRS Éditions.

1994, *Cours de morphologie générale. Deuxième partie : significations morphologiques*, Montréal, Presses de l'Université de Montréal-CNRS Éditions.

MILLER (P. H.)

1991, *Clitics and Constituents in Phrase Structure Grammar*, Thesis, Utrecht, Rijkuniversiteit.

MILLER (P. H.), PULLUM (G. K.), ZWICKY (A. M.)

1992, "Le Principe d'inaccessibilité de la phonologie par la syntaxe : trois contre-exemples apparents en français", *Linguisticae investigationes*, 16, 2.

MITHUN (M.)

1984, "The Evolution of Noun Incorporation", *Language*, n°60, p. 847-893.

1996, "Roots and affixes", *Seventh International Morphology Meeting (16-18.2)*, Vienna.

MOESCHLER (J.), REBOUL (A.)

1994, *Dictionnaire encyclopédique de pragmatique*, Paris, Seuil.

NAMER (F.)

1994, *Workpackage WP2. Report on Morphology*, MLAP 93-07 Deliverable D2, TALANA-Université Nancy 2.

NEVIS (J. A.)

1986, *Finnish Particles Clitics and General Clitic Theory*, Columbus, The Ohio State University (Dept. of Linguistics).

NIDA (E. A.)

1949, *Morphology : The Descriptive Analysis of Words*, Ann Arbor, The University of Michigan Press.

PERDRIZET (F.)

1983, "Les Structures morphologiques du vocabulaire de la bonneterie", *Cahiers de lexicologie*, n°42, p. 72-85.

PLÉNAT (M.)

1981, "L' «Autre» conjugaison ou de la régularité des verbes irréguliers", *Cahiers de grammaire*, n°3, p. 1-304.

POTTIER (B.)

1974, *Linguistique générale*, Paris, Klincksieck.

RAINER

1995, "Inflection inside Derivation : Evidence from Spanish and Portuguese", *Yearbook of Morphology 1995*, p. 83-91.

RASTIER (F.)

1996, "Pour une sémantique des textes", p. 9-35, *in* RASTIER (F.), ed.

RASTIER (F.), ed.

1996, *Textes & Sens*, Paris, Didier-Érudition.

REBOUL (S.)

1994, *Le Vocabulaire de la télématique : du discours au lexique*, Thèse de doctorat, Université de Paris X.

RIEGEL (M.), PELLAT (J.-C.), RIOUL (R.)

1994, *Grammaire méthodique du français*, Paris, PUF.

RITCHIE (G. D.), RUSSELL (G. J.), BLACK (A. W.), PULMAN (S. G.)

1992, *Computational Morphology*, Cambridge-London, The MIT Press.

SAUSSURE (F. de)

1916, *Cours de linguistique générale*, Paris, Payot.

SHOPEN (T.), ed.

1985, *Language Typology and Syntactic Description*, vol 3, *Grammatical Categories and the Lexicon*, Cambridge University Press.

SILBERZTEIN (M.)

1990, "Le Dictionnaire électronique des mots composés", *Langue française*, n°87, p. 71-83.

1993, "Les Groupes nominaux productifs et les noms composés lexicalisés", *Linguisticae Investigationes*, XVII-2, p. 405-425.

TAMBA (I.)

1986, "Approche du «signe» et du «sens» linguistiques à travers les systèmes d'écriture japonais", *Langages*, n°82, p. 83-100.

TOGEBY (K.)

1965, *Structure immanente de la langue française*, Paris, Larousse.

TOURNIER (J.)

1985, *Introduction descriptive à la lexicogénétiq ue de l'anglais contemporain*, Paris-Genève, Champion-Slatkine.

TRNKA (B.)

1949, "Peut-on poser une définition universellement valable des domaines respectifs de la morphologie et de la syntaxe ?", p. 19-30, *in* LEJEUNE (M.), ed.

Van MARLE (J.)

1985, *On the Paradigmatic Dimension of Morphological Creativity*, Dordrecht, Foris.

VARELA (F. J.)

1989, *Autonomie et connaissance*, Paris, Seuil.

VIVÈS (R.)

1990, "Les Composés nominaux par juxtaposition", *Langue française*, n°87, p. 98-103.

WELLS (R. S.)

1947, "Immediate Constituents", *Language*, 23, p. 81-117.

WUNDERLICH (D.), FABRI (R.)

1993, "Minimalist Morphology : An Approach to Inflection", ms, Düsseldorf.

ZWICKY (A. M.)

1985, "Clitics and Particles", *Language*, 61-2, p. 283-305.

1992, "Some Choices in the Theory of Morphology", p. 327-371, in LEVINE (R. D.), ed.

