

CHAPITRE III

LES MODÈLES LINGUISTIQUES ET LA QUANTIFICATION

1. OBJECTIFS

Ce chapitre a pour objet de présenter, en les replaçant dans un contexte de recherche actuel, les 6 modèles linguistiques et pragmatique tels qu'ils ont été définis ou redéfinis pour notre expérimentation. Devant la masse considérable de travaux publiés, tant en France qu'à l'étranger, sur les domaines linguistiques fondamentaux que représentent l'étude de la signification et du sens, il est clair que nous ne pouvons que poser certains jalons : l'objet de cet ouvrage étant focalisé sur les stratégies prosodiques, nous nous limiterons de ce fait aux travaux qui ont alimenté la conception de nos modèles, ou aux travaux qui gravitent dans la même sphère.

Notre projet global dans cette étude est de tenter de dresser un bilan dans le cadre de nos énoncés, entre d'une part les composantes linguistiques (syntaxe, sémantique) et pragmatique, et d'autre part la prosodie.

Plus précisément, il s'agit de voir si à *l'émergence linguistique* (Grunig, 1991), par nature non quantifiable, répond une autre émergence de type prosodique, en particulier mélodique. Selon Grunig, cette émergence linguistique du fait que les unités de la chaîne n'ont pas toutes "le même poids, le même relief", "constitue des hiérarchies", "instaure des ordres", mais "en général ne se chiffre pas". Cette émergence linguistique est de divers ordres et peut se repérer en termes d'indices de nature très diverse.

Dans notre perspective, la prosodie a précisément pour fonction de restaurer au discours cette quantification qui existe certainement de manière sous-jacente dans les structures linguistiques, puisque le locuteur en élaborant son discours, peut avoir le sentiment que telle unité est prioritaire ou plus centrale qu'une autre. Dans ces conditions on peut penser que l'usage qu'il fait de sa langue dans l'acte concret de parole, lui permet d'exprimer ses priorités. Chaque modèle que nous avons défini a précisément pour objectif de décrire un type d'émergence particulier, d'en capter les indices, et en fonction d'une échelle de valeurs, de convertir cette émergence en éléments quantitatifs, qu'il est alors possible de comparer avec une quantification prosodique : le modèle a alors la propriété de prédire ces variations prosodiques.

Partant de l'idée que le locuteur dès sa prise de parole, dans le même temps où il construit une forme phonétique et prosodique en direction d'un auditeur, communique un sens par l'intermédiaire d'unités linguistiques, nous avons pensé que les structures qui régissent ces deux aspects de la signification devaient être corrélées au coeur même du processus d'encodage de la parole. Les structures linguistiques étant de nature syntaxique, sémantique et pragmatique, il y a lieu de penser que les stratégies prosodiques qui accompagnent le discours, empruntent ces mêmes voies. Il s'agit en fait de voir comment la prosodie peut apporter sa contribution dans l'élaboration du sens en s'appuyant sur certains indices de la signification à la manière d'un *pré-traitement formel en direction de l'auditeur*.

On peut considérer que pour la lecture, ces indices, structurés ensuite en modèles, fonctionnent à la production (encodage de la parole), comme à la réception (décodage), comme des instructions cognitives à l'élaboration du sens¹. C'est un point qui nous semble capital.

¹ Il est probable que la situation se limitant à la représentation que le locuteur se fait des contraintes explicites de lecture, n'interviendra à l'encodage qu'en tant que simple catalyseur des processus d'extraction et de constitution du sens.

Ces indices significatifs nous ont semblé pouvoir être empruntés à plusieurs espaces, comme celui de la phrase, ou le contexte local. S'agissant de la phrase, les indices porteront sur des structures, et nous aurons alors à faire aux modèles de *description globale ou holistique*². S'agissant d'un contexte local ou discursif, les indices pris en charge par les modèles seront de type lexical, et de ce fait la visée des modèles sera de type *analytique*. Dans ces différents cas, la nature des indices sera syntaxique, sémantique, pragmatique. Pour que cette étude soit possible sur le plan concret, il faut également que ces indices, pour tous les modèles, concernent non les processus périphériques de la langue, mais les processus fondamentaux, et soient suffisamment prégnants pour que chacun des lexèmes de l'énoncé, en référence soit à une structure de phrase soit à un contexte local, puisse être directement impliqué dans ce mécanisme.

En clair cette expérimentation doit nous permettre d'identifier les indices utilisés par les locuteurs pour reconstruire un sens (naturel, intelligible), à destination d'un tiers. Ces indices sont repérables dans des espaces différents (phrase, contexte local, réseau isotopique ...), sur plusieurs dimensions (pragmatique, sémantique, syntaxique), et constituent selon leur nature, les différents modèles. Dans tous les cas leur granularité est faible (le lexème).

Mais se pose alors le problème de la quantification, car d'un côté nous avons une représentation symbolique de la langue, donc qualitative, et de l'autre une représentation quantitative.

Dans la réflexion que nous avons menée sur la nature des modèles les plus adéquats, syntaxiques, sémantiques, pragmatiques, il nous a semblé que ceux qui évoquent une échelle de grandeur pouvaient répondre à cette exigence de quantification. Les notions de *hiérarchie*, de *complexité*, de *distance* qui traitent de relations essentielles en langue, nous ont semblé alors se prêter particulièrement bien à nos besoins.

Dans le domaine syntaxique, le concept de *hiérarchie* des constituants, reposant sur un principe de description holistique, était un candidat tout désigné, dans la mesure où ces niveaux pouvaient être mis en correspondance avec les niveaux de F0.

Concernant la description analytique, plusieurs possibilités sont offertes, telles qu'une description en parties du discours (nom, adjectif, verbe ...), en fonctions (sujet, objet, complément de lieu ...), ou encore en termes de distance des unités linguistiques dans la structure de phrase (dépendances). Des travaux antérieurs (Caelen-Haumont, 1978, 1981), nous avaient montré qu'une description en parties du discours et en fonctions était de toute manière insuffisante pour la prosodie du français, car fonctions syntaxiques et formes intonatives ne sont pas en relation biunivoque. Ce type d'analyse n'a donc pas été retenu.

Inversement, la notion de place s'était révélée essentielle dans l'analyse des structures prosodiques que nous avons menée. Dans ces conditions, un modèle traitant des degrés "d'attraction"³ (autrement dit, de *distance*) entre les unités linguistiques selon leur place dans le syntagme et leur position par rapport au noyau de celui-ci, était un argument convaincant pour retenir ce phénomène comme principe d'un modèle. En outre,

² Un modèle est appelé holistique dans la mesure où dans la description qu'il propose, il tient compte du cadre de l'ensemble de la phrase, les unités lexicales étant évaluées en fonction de la structure totale, c'est-à-dire en fonction de leur place dans la hiérarchie constituante. A l'inverse les modèles analytiques ne retiennent pas le cadre de la phrase comme principe d'analyse, et envisagent les relations locales ou à plus long terme dans l'ensemble du texte.

³ Comme nous le verrons plus loin, en dehors des lexies qui forment une classe particulière, *l'attraction* par exemple d'un adjectif antéposé vis-à-vis d'un substantif qui le suit, est plus forte que celle d'un substantif antéposé à un adjectif.

on pouvait faire l'hypothèse a priori que ces modes de relation entre les unités pourraient rendre compte des phénomènes de continuité et de rupture souvent observés dans la courbe de F0.

Le domaine sémantique, lieu de la constitution de la signification, de l'émergence de l'information et donc d'un sens, est un espace privilégié de la description des principaux mécanismes linguistiques. Dans une situation d'expérience où une des contraintes qui pèse sur le lecteur est de produire un discours de plus en plus intelligible, nous ont paru alors essentiels en ce domaine, les processus relevant de l'analyse de la *complexité* des unités lexicales, et ceux qui relèvent du statut de l'information (thème / rhème). De plus le modèle de la distribution de l'information répond à une description holistique, le modèle de la complexité des unités lexicales en contexte, à une description analytique.

Un autre type d'analyse, complémentaire et très proche du domaine sémantique, s'est alors présenté. Ce type de description est en quelque sorte transversal aux deux descriptions sémantiques précédentes, dans la mesure où il ne fait plus seulement référence à la signification, mais au processus d'interprétation lui-même, c'est-à-dire aux traitements qu'un locuteur est susceptible d'opérer lors de la production d'un énoncé à destination d'un tiers. Le modèle ainsi ne traite plus de la complexité lexicale, mais de la *complexité* dans le traitement cognitif des informations que le locuteur opère pour faciliter la tâche de compréhension chez l'auditeur. Ces traitements désignent par exemple les opérations d'inférence appliquées aux unités lexicales dont le sens est nouveau ou déjà mémorisé. Ce type d'analyse a mené à un modèle analytique, et dans la mesure où il traite des processus psychocognitifs du locuteur en situation de parole, ce modèle est de nature pragmatique.

Par la suite 6 modèles ont donc été définis ou repris à la tradition, deux dans le domaine de la syntaxe (modèle de la hiérarchie syntaxique ou HR, de la dépendance des unités lexicales en contexte ou DP), trois dans le domaine de la sémantique (modèle de la hiérarchie énonciative ou EN, de la hiérarchie énonciative à dominante rhématisation ou ER, de la complexité lexicale en contexte ou CM), et un dans le domaine de la pragmatique (modèle des informations supposées attendues ou inattendues, ou encore appelé modèle de la connaissance partagée, ou CP). Comme nous le verrons par la suite, tous ces modèles ont par ailleurs la particularité de n'évaluer et de ne quantifier que les mots lexicaux. Ils sont tous présentés à l'annexe de ce chapitre.

De manière générale, l'hypothèse qui sous-tend la relation entre lexique et mélodie est la suivante : *plus le mot lexical est complexe, haut dans la structure hiérarchique, distant sur le plan syntaxique, inattendu, ou moins simple sur le plan du traitement cognitif, plus son poids est important, et plus le niveau de F0 est supposé haut dans le registre, ou l'écart de F0 plus grand dans le mot lexical.*

En fait cette hypothèse est fondée sur les résultats en psycholinguistique. De nombreux travaux ont établi depuis longtemps (parmi de nombreuses autres publications, Williams & Stevens, 1972) que l'expression instantanée des émotions (joie, peur ...) passe par l'utilisation de registres de la voix les plus aigus. De la même façon, au niveau de la communication des informations, il semble justifié de penser que ce qui relève de l'interprétation la plus intime pour un locuteur et de l'expression de sa subjectivité, s'exprime également par le registre vocal le plus aigu.

2. LES MODELES SYNTAXIQUES

2.1. MODELE DE LA HIERARCHIE SYNTAXIQUE (HR)

Ce modèle de la hiérarchie syntaxique HR est le seul à avoir été repris de la tradition linguistique américaine, celle du structuralisme puis du générativisme (modèle standard, Chomsky, 1965), et ce, sans beaucoup de modifications. Comme ce modèle est bien connu, je n'insisterai pas davantage sur ses catégories.

Le postulat des grammaires universelles se fonde comme le souligne Rastier (1991), sur le principe d'homogénéité de la langue. En élaborant la "grammaire d'une langue écrite standardisée", le linguiste évacue ainsi tout type de variations relatives par exemple aux niveaux de langue, à l'oral, aux dialectes ou aux parlers locaux. Avec le deuxième principe d'autonomie de la langue, il va de soi que "le postulat de l'identité à soi comme celui de l'autonomie concourent à désocialiser les langues." On pourrait ajouter encore que ces postulats 'désubjectivent' le discours, dans la mesure où le recours à ce modèle par un locuteur traduirait le primat d'une structuration linguistique sur d'autres types de motivations linguistiques, psychologiques ou pragmatiques.

Bien entendu ce modèle, comme ceux de la hiérarchie sémantique se développent dans le cadre de la phrase : à ce titre s'ils peuvent être plus facilement sollicités dans les conditions de lecture du fait de la structure textuelle, ils auraient beaucoup de mal à l'être dans le cadre de l'oral spontané pour lequel la notion de phrase n'existe pas. Sur ce sujet, on peut se référer à Blanche-Benvenite (1990).

Comme ce modèle passe pour un modèle psychocognitif ancré dans notre patrimoine génétique, il peut être intéressant précisément d'évaluer la performance de ce modèle dans les conditions qui lui semblent les plus favorables, et de tester ainsi l'hypothèse qui ferait du primat du linguistique un modèle psychocognitif comme l'autorise notre méthodologie. En outre ce modèle reposant sur une hiérarchie de constituants par nature quantitative, se prête assez bien à cette confrontation avec les données prosodiques concrètes.

Comme on le sait néanmoins, certaines solutions avancées par la grammaire générative et transformationnelle nécessitent des transformations qui ne sont pas les plus naturelles chez un locuteur "standard", c'est-à-dire non linguiste.

Notre texte renferme un exemple de ce type dans la première phrase. En effet le groupe *d'éminents biologistes et d'éminents zoologistes américains* qui en soit se présente comme un groupe assez banal, a la particularité de posséder un adjectif *éminents* qui de manière explicite qualifie chacun des deux noms, et un autre adjectif *américains* qui du fait de la coordination, qualifie les deux groupes nominaux comme s'il était "mis en facteur commun".

En tout état de cause la théorie prévoyait⁴, pour régler ce problème, une transformation par effacement de l'adjectif *américains* après *biologistes*, ce qui peut se représenter comme suit :

(1) d'éminents biologistes [américains] et d'éminents zoologistes américains,

Ø

ce qui aboutit à la structure arborescente (graphique 3-1) présenté à l'annexe du chapitre.

Dans le modèle, chaque mot lexical a pour coefficient la valeur numérique qui correspond à son niveau dans la structure constituante. Il s'agit en fait d'attribuer à chaque item lexical le poids qui dans l'arbre, correspond au niveau du noeud qu'il referme.

⁴ Dans le nouveau développement de la grammaire chomskyenne, l'effacement n'existe plus. La seule transformation qui est admise est le déplacement.

Par exemple, si on a un arbre tel que celui proposé dans le graphique 3-1, on dira que le noeud "refermé" par l'item lexical *éminents* est seulement le noeud ADJ, qui est de niveau 2 (les niveaux sont comptés à partir des noeuds pré-terminaux), donc l'item *éminents* reçoit le poids 2, alors que l'item *biologistes* qui referme non seulement sa catégorie pré-terminale N mais aussi le syntagme GN de niveau 3, reçoit le poids 3. Dans le modèle HR, la distribution des coefficients est donc censée indiquer la *constituance syntaxique*.

La quantification est donc ici d'ordre iconique : plus le constituant est élevé dans la structure hiérarchique de l'arbre, plus le poids est important, et plus ce constituant est supposé déterminer une élévation de F0 (ou augmenter l'écart de F0) dans le mot.

La seule modification qui a été apportée concerne le niveau du dernier mot lexical : ce niveau correspond non pas à celui du niveau phrase comme on aurait pu s'y attendre, mais à celui du niveau du syntagme qui le domine immédiatement dans la structure, car dans notre interprétation des faits, l'unité lexicale terminale de phrase, avant de représenter de manière purement symbolique le niveau phrase, est insérée avant tout concrètement dans le tissu des relations intra- et intersyntagmatiques de cette dernière.

Ceci nous semblant légitime quelle que soit la structure hiérarchique considérée, nous adopterons le même principe pour quantifier l'unité lexicale terminale de phrase dans les modèles de l'énonciation EN et ER. Ce traitement exceptionnel des unités lexicales de phrase est souligné dans les graphiques et leur quantification par l'astérisque * à la droite du poids concerné. Le graphique n° 3-1, ci-dessous à l'annexe, présente la première phrase avec sa quantification.

2.2. MODELE DES RELATIONS DE DEPENDANCE (DP)

Ce type d'analyse a fait école au niveau international et a été l'objet de nombreuses formalisations, par exemple aux Etats-Unis (Hays, 1964), en Russie (Kulagina & Mel'cuk, 1967), en Allemagne (Kunze, 1975; Hellwig, 1980, 1986), en Tchécoslovaquie (Hajicova, 1988), en Finlande (Jäppinen et al. 1988; Valkonen et al., 1987) et bien sûr en France (Veillon 1970; Vauquois, 1975; Courtin, 1977; Bailly, 1983; Genthial, 1991).

Cette perspective est féconde sur le plan linguistique, mais par rapport aux processus prosodiques, elle nous semble légèrement inadéquate car parfois trop simplificatrice des processus syntaxiques, parfois au contraire trop riche. Au niveau syntaxique lui-même, nous ne pensons pas par exemple que deux adjectifs qui dépendent d'un même gouverneur et qui encadrent un substantif se situent dans le même rapport, la relation premier adjectif / nom étant certainement plus forte que la relation nom / deuxième adjectif, et ceci bien que d'un point de vue sémantique, la relation des deux adjectifs au nom soit entièrement symétrique. En fait il nous semble qu'entre tout adjectif et nom, il y a non pas une seule relation mais deux sous-jacentes : celle de l'adjectif au nom et celle du nom à l'adjectif. Lorsque l'adjectif est antéposé, la relation dominante est celle de l'adjectif au nom et c'est une relation forte. Lorsque l'adjectif est postposé, celle qui domine est la relation nom / adjectif et elle est plus faible parce que le nom ne dépend pas syntaxiquement de l'adjectif qui le suit dans la mesure où il constitue le noyau nécessaire et suffisant du syntagme.

En outre les processus purement linguistiques (syntaxiques mais aussi sémantiques) et les processus prosodiques ne sont pas dans le même rapport vis-à-vis de la linéarité du discours. A notre avis les relations sémantiques ou syntaxiques s'appuient sur des relations gauche-droite et droite-gauche, alors que la prosodie de lecture où le discours est planifié (Guaïtella, 1999) privilégie le sens gauche-droite, comme le montre à l'évidence dans le discours lu, le contour à la finale des mots lexicaux.

Nous avons donc défini un nouveau modèle qui intègre ces considérations. Il a la propriété de ne traiter que les unités lexicales qui se succèdent, mais de deux manières : les unités lexicales se suivent en étant en relation directe ou indirecte sans autre élément intermédiaire, soit leur relation s'opère à distance. Par exemple un verbe se construisant de manière directe peut être séparé de son COD par un complément circonstanciel : dans ce cas, l'attribut "direct" est neutralisé.

Ce modèle (graphique n° 3-2 à l'annexe) des relations de dépendance (ou DP) propose une échelle des relations entre le degré le plus grand de dépendance jusqu'à celui de l'indépendance le plus élevé. Pour respecter l'axe linéaire de lecture, le modèle ne traite que les relations de dépendance gauche-droite. Pour la quantification, c'est le degré d'autonomie qui est évalué : ainsi plus la relation syntaxique entre une première unité lexicale et la suivante est faible, plus le degré d'autonomie est fort, et plus le poids est élevé. L'hypothèse sous-jacente est que plus l'élément est autonome vis-à-vis d'un suivant, et plus l'excursion du contour mélodique final est importante. Voici les relations retenues :

1° relations internes au groupe syntaxique

* dépendance locale ou DL : la dépendance locale concerne directement les adjectifs et les adverbes lexicaux qui sont par nature très fortement liés respectivement au substantif, à l'adjectif (etc.) qui les suit (ex : *un nouveau phylum...*, *fortement secoué ...*). Le niveau d'indépendance est le plus faible.

* indépendance locale ou IL : cette catégorie correspond au substantif ou à l'adjectif qui se situe ailleurs qu'à la frontière de droite du groupe syntaxique, car ce dernier cas relève de l'indépendance finale (traitée ci-dessous). L'indépendance locale correspond par exemple au cas du substantif qui ne dépend pas syntaxiquement de l'adjectif qui le suit (ex : *vers géants ...*), ou encore à celui d'un premier adjectif vis-à-vis du deuxième lorsque deux adjectifs sont postposés au substantif (ex : *zones sous-marines profondes*). Le niveau d'indépendance est un peu plus élevé.

2° relations de dépendance à la borne de droite du groupe :

* indépendance finale ou IF : c'est le cas de l'item lexical en fin de phrase. D'un point de vue syntaxique, c'est sans doute le degré d'indépendance maximal. En fait le modèle tente d'évaluer les relations entre deux mots lexicaux qui se suivent et dans ces conditions, un mot final de phrase n'a pas la possibilité de se démarquer. Le niveau d'indépendance est jugé du même ordre que la relation IL.

* subordination directe ou SD : cette catégorie n'est pas différente de la relation syntaxique habituelle et concerne donc la construction directe d'un verbe lorsqu'elle est effectivement réalisée (ex : *maintiennent une température ...*). Le niveau d'indépendance est jugé équivalent à celui des catégories précédentes.

* subordination indirecte ou SI : ceci vise non seulement les constructions indirectes des verbes (objet indirect), mais aussi toutes les autres constructions verbales par préposition (attribution, complément circonstanciel, agent ...), et les constructions substantivales (complément de nom ...), adjectivales ... etc. A cette série s'ajoutent comme on l'a précisé supra, les verbes de construction directe qui par permutation des unités se trouvent au contact d'un groupe prépositionnel (ex : *créé pour des vers géants ...*). La distance entre constituants est plus grande que pour les catégories précédentes.

* transsubordination ou T : nouvelle par rapport à la syntaxe traditionnelle, cette catégorie recouvre les cas où la construction d'un verbe, directe ou indirecte, transite par un ou plusieurs groupes intermédiaires jusqu'au complément voulu. Ceci ne concerne pas la relation entre le verbe et son complément différé, mais, comme déjà précisé, les mots lexicaux à la jointure, soit entre la dernière unité lexicale du groupe intermédiaire (qui reçoit l'étiquette T et du poids correspondant) et la suivante du groupe ultérieur, le mécanisme se renouvelant pour tous les groupes intermédiaires (ex : *pour des vers géants un nouveau phylum dans l'actuelle classification ...*). L'indépendance des constituants est jugée du même ordre que précédemment.

* coordination et juxtaposition ou C : les éléments coordonnés ou juxtaposés sont considérés comme plus indépendants que les catégories antérieures (ex : *d'éminents biologistes et d'éminents zoologistes américains ...*).

* indépendance ou I : ce cas traite de tous les mots lexicaux placés en fin absolue de syntagme (sauf en fin de phrase, cf ci-dessus), quelle que soit leur nature. Ils correspondent en fait aux éléments terminaux des constituants majeurs de la phrase (ex : *d'éminents zoologistes américains ont créé...*). C'est évidemment le degré le plus grand d'indépendance entre mots lexicaux.

Pour ce modèle, la quantification est également d'ordre iconique : plus la distance syntaxique entre un mot lexical et l'unité qui suit est grande, plus le poids est important, et plus le niveau de F0 est censé être élevé (ou l'écart de F0 important). Nous envisageons maintenant les modèles sémantiques.

3. MODELES SEMANTIQUES

3.1. MODELES DE LA HIERARCHIE ENONCIATIVE (EN, ER)

On sait que les notions fondamentales de "thème " et de "rhème" ont été introduites à l'origine par les linguistes de l'École de Prague (Danes, 1968 ; Firbas, 1974), et qu'elles ont été ensuite largement reprises au niveau international.

Parmi les publications, on peut citer⁵ par exemple en France (Pottier, 1967; Slakta, 1975; Combettes, 1977; Hagège, 1978; Perrot et al., 1974; Perrot, 1978; Touratier, 1993), et en relation avec la prosodie, (Rossi, 1973; Caelen-Haumont, 1978, 1981, 1991; Rossi et al., 1981, 1985, Danon-Boileau et Morel, 1994, 1995).

Ces structures ont posé dès leur origine, et toujours encore, des problèmes d'interprétation, car d'une part leur terminologie n'est pas encore fixée, d'autre part les concepts centraux ou voisins qu'elles véhiculent tels que sujet-prédicat, donné-nouveau, topicalisation, focalisation, information périmée - information nouvelle, thème-rhème du propos, thème-rhème de l'assertion, support-apport, présupposé-posé ... ne se laissent pas facilement distinguer, et enfin parce que les différentes interprétations de ces concepts hésitent entre une explication sémantique, syntaxique et prosodique ... On peut trouver dans Galmiche (1992) un état des problèmes posés (voir également le chapitre I de cet ouvrage).

Ainsi les notions sujet-prédicat (héritées d'Aristote), appelées encore thème-rhème de l'assertion, sont syntaxiques, contrairement à l'expression thème-rhème (du propos), qui sont sémantiques (et pragmatiques, car un rhème peut s'appliquer à un élément de la situation). En outre, le couple donné-nouveau est vu souvent comme un équivalent de thème-rhème (sémantique-pragmatique), alors qu'un thème peut comporter une

⁵ Voir à ce sujet le chapitre 1, "Enjeux de la prosodie et bibliographie".

information nouvelle et être considéré en même temps à juste titre comme le donné. Qui plus est, le donné-nouveau fait référence à des données lexicales, le thème-rhème concerne la phrase. Le terme d'information peut aussi prêter à confusion, car une information véhiculée par un mot lexical (ou grammatical), peut ne pas être nouvelle.

Selon la théorie de la "perspective fonctionnelle de la phrase" (ou FSP), le "thème" désigne la partie du texte ou du discours sur laquelle quelque chose est dit et plus exactement selon Touratier (1993) "ce à *propos de quoi* (et non au sujet de quoi⁶) le locuteur dit quelque chose", ce "quelque chose qui est dit" représentant le "rhème". Classiquement, l'information est supposée reposer exclusivement sur le rhème, le thème véhiculant l'ancienne information maintenant périmée, et/ou ce qui relève de la présupposition et qui est contenu dans la question sous-jacente. Dans cette perspective, la suite des thèmes assure au texte sa cohérence, alors que celle des rhèmes permet la progression de l'information.

Cependant dans la théorie fonctionnelle, ces notions de thème et de rhème par définition sont considérées dans une perspective linéaire, et assignées à une place fixe dans l'énoncé, ce qui a pour conséquence de projeter et de démultiplier les notions sur l'axe linéaire pour rendre compte de la complexité naturelle du texte (cf par exemple la notion de "transition", d'éléments "propres" et "secondaires" ...).

Cette théorie⁷ qui mérite réflexion, peut cependant difficilement résoudre autrement que par la qualification certains problèmes tels que la prise en compte de deux thèmes dans une même phrase qui ne possèdent pas pourtant le même niveau de signification, ou encore celle de thèmes (ou de rhèmes) qu'il faut bien qualifier de "composites" puisque des éléments de rang inférieur répondent dans les limites de leurs frontières, à la définition de la notion de rhème (ou de thème). Pour nous aussi, la place ne suffit pas à elle seule à assigner une fonction.

Il a donc paru nécessaire de développer la théorie dans un autre cadre, et ce particulièrement dans des applications de type prosodique, où les unités semblent se référer à une organisation non linéaire.

Ce problème de la répartition des thèmes et des rhèmes en couches successives a d'ailleurs été très tôt pressentie (Hazaël-Massieux, 1974). Pour l'auteur, contexte et situation sont en relation étroite, et plus exactement en relation de dépendance stricte :

Si le contexte s'accroît, la part d'implicite (laissée à la situation) diminue. Si le contexte diminue, le rôle de la situation augmente pour la bonne compréhension du message. Nous dirons donc que le contexte est la situation explicitée. Hazaël-Massieux (1974).

C'est dans ce cadre précis que fonctionnent l'information et les unités qui en dépendent, dénommées à la suite de Pottier (1967) *support* et *apport*. Ces notions, précise l'auteur, ne sont pas exactement superposables à celles de *thème* et de *rhème*, ou de *topic* et *comment*. Ces unités sont susceptibles d'être découpées en constituants superordonnés, ce faisant, elles peuvent changer de statut par permutation des identités :

⁶ ce qui distingue ce concept de celui du sujet syntaxique.

⁷ Bien entendu le recensement des divers types de thème et de rhème aurait pu déboucher sur une quantification, mais d'une part l'analyse en niveaux aurait été plus subjective et vraisemblablement moins riche qu'une analyse hiérarchique, car dans ce dernier cas, tout mot lexical est pris en compte.

syntactiquement il est dissocié du rhème et entre dans l'ensemble présentatif-relative (*y'a mon frère qui ...*). Le sujet logique se différencie du thème, dans la mesure où la dissociation du rhème se fait par l'intermédiaire d'un simple pronom de rappel (*mon frère, il ...*). Le thème introduit généralement une information nouvelle, alors que le sujet logique n'est souvent que la reprise d'un thème précédent. Le sujet syntaxique pour sa part se situe en dehors du préambule, et n'est pas dissocié du rhème. Enfin apparaissent dans le post-rhème (autre notion générique), plusieurs éléments dont la valeur est proche de celle du préambule, et en particulier, thème et cadre. Parmi d'autres moyens, la *focalisation* construit pour sa part un rhème, soit prosodiquement, soit de manière syntaxique, par le présentatif.

Finalement cette reconstruction se fait sur des critères syntaxiques : il s'agit d'une syntaxe énonciative qui a le mérite entre autres de la clarté. Cette analyse est appliquée à l'analyse de l'intonation.

Antérieurement, et sur des bases différentes, un nouveau développement a vu le jour et a été ensuite appliqué à l'analyse prosodique. Cette conception systématise en fait certains développements antérieurs et opère sur une analyse en constituants immédiats (mais énonciatifs), comme en syntaxe générative. Bien entendu ses concepts demeurent spécifiques, ce qui aboutit à une structuration différente en essence. Cette méthode procède par distinction hiérarchique et successive des thèmes et des rhèmes, l'analyse progressant des thèmes et rhèmes à l'extension la plus large à ceux à l'extension la plus réduite (Caelen-Haumont, 1990, 1991).

Ce type d'analyse est maintenant couramment utilisé dans le contexte international. Pour l'anglais, Prevost et Steedman (1993) se contentent de 2 niveaux de structuration, le premier découpant l'énoncé en thème / rhème, le second niveau isolant dans le thème et le rhème, l'élément focalisé (ou focus). Pour le français, on peut citer par exemple les travaux de Touratier (1993), Kim (1995). En particulier Touratier (1993) précise quels sont les moyens pour identifier le support sur le plan syntaxique :

Si l'énoncé a une ou plusieurs extrapositions antéposées, la première ou la seule extraposition, c'est-à-dire le C.I. [constituant immédiat] de la P endocentrique⁸ qui n'est pas lui-même une P, contient le support informatif. Sinon, si la phrase est exocentrique, le sujet, c'est-à-dire, en général, le C.I. de P qui n'est pas un SV, est le support de l'énoncé. Sinon, si la phrase n'a pas de sujet, mais est formée d'un SV, le support correspond à l'individu qui est désigné par le morphème personnel du verbe. Dans les autres cas, le support est ce à quoi peut sémantiquement s'appliquer, en fonction des données énonciatives, le contenu de l'énoncé à un seul constituant de base. Touratier (1993).

Pour revenir au modèle proposé dans cet ouvrage, il faut insister sur le fait que cette structuration étant par nature différente de la structuration syntaxique, les éléments thème et rhème, n'entretiennent pas nécessairement entre eux de relation de type complémentaire ou actancielle, ces notions recouvrant parfois des unités linguistiques de niveaux syntaxiques différents ou encore sans relation syntaxique.

De la même façon, la nature des constituants syntaxiques immédiats et celle des constituants énonciatifs étant distincte, la coïncidence des frontières entre ces constituants syntaxiques et énonciatifs, n'est pas à rechercher⁹.

⁸ La terminologie et les concepts remontent à Bloomfield (1970). Selon François (1974), dans la construction endocentrique, "l'ensemble est grammaticalement identique à un des termes", par exemple $P \rightarrow SP + P$. Selon Touratier (1994), une construction exocentrique est "une construction qui ne remplit pas le même rôle syntaxique que l'un de ses constituants immédiats", par exemple $P \rightarrow SN + SV$.

⁹ Ceci induit en particulier l'absence de "symétrie" syntaxique entre les constituants énonciatifs, ces derniers n'ayant pas nécessairement le même nombre de syntagmes, et pouvant être issus de constituants

Plus exactement si les modèles de l'énonciation s'appuient bien sur les constituants syntaxiques en ce sens que le principe de génération ne remet pas en cause la structuration en constituants (les constituants générés restent évidemment des constituants syntaxiques), la différence repose sur la relation entre ces constituants, qui est sémantique pour les modèles de l'énonciation par la bi-partition qui est faite à tous les niveaux entre thème et rhème.

Cette structuration énonciative hiérarchique réorganise en fait les groupes minimaux en fonction des lois qui lui sont spécifiques : le facteur fondamental qui permet de générer les structures de rang inférieur réside dans la reconnaissance de l'élément qui de manière relative est porteur d'information de plus haut niveau.

Comme les dénominations de "thème" et de "rhème" appartiennent en priorité à la théorie fonctionnaliste de l'énonciation, il nous est apparu nécessaire d'adopter une autre terminologie. Dans ces conditions, notre analyse étant plus proche de celle de Pottier (1967) et d'Hazaël-Massieux (1974), Perrot (1978), nous substituons à "thème" le terme de "support" (symbole "S"), et à "rhème", celui de "apport" (symbole "A"). Ces termes ont été choisis parce qu'ils expriment bien le mécanisme énonciatif fondamental ("*dire quelque chose à propos de*"), qui fonctionne par vagues successives (l'apport) à partir d'une unité de référence qui la précède (support) sur la chaîne syntagmatique et qui, comme l'a bien vu encore De Cornulier (id., 1979), la conditionne.

Pour nous, il est bien clair que l'apport d'information tel que l'auteur d'un énoncé le formule (et que cette information paraisse à l'auditeur, réelle, feinte, supposée, nouvelle ou pas, vraie ou fausse ...) constitue le principe de base de l'analyse en constituants inférieurs. Mais cette méthode d'analyse ne nous empêche pas d'analyser le processus plus à fond et de reconnaître que ce mécanisme d'introduction de l'information est chronologiquement second.

De toute évidence dans l'énonciation, on est bien en présence de deux phénomènes distincts, l'un qui repose sur la relation syntagmatique des constituants de l'énonciation, le constituant "support" servant de base au constituant "apport" et le conditionnant, l'autre, sur le contenu de ces constituants, contenu en relation avec la notion d'information, le "donné" et le "nouveau". Et du fait que généralement, un locuteur prend la parole pour apporter un élément d'information, il existe dans la majeure partie des énoncés, une coïncidence entre cette structure syntagmatique et l'expression de ce contenu. Cependant, il n'en va pas toujours ainsi : lorsque par exemple un discours commence, ou que les interlocuteurs changent de sujet, il arrive que thème et rhème soient tous deux de nature informative, mais les conditions de l'énonciation exigent toutefois qu'entre les deux s'établisse une relation d'ordre, une construction, un "dire à propos de". Ces remarques mettent en relief le mécanisme central de l'énonciation comme structure : dans le cas *d'énoncé neutre*, cette notion peut rejoindre le processus syntaxique de la prédication, avec ses unités, topique et commentaire, ce qui n'est pas nécessairement le cas lorsque le locuteur s'investit dans le propos.

Comme le souligne Combettes (1988), contrairement au texte, l'oral spontané répugne à introduire un référent nouveau comme thème :

Dans le domaine de l'oral spontané, tout se passe comme si une règle "pragmatique" se mettait en place, qui interdit d'introduire un référent nouveau [...] en position de sujet, devant le verbe [...]. Il n'en va pas de même à l'écrit [...] où aucun indice syntaxique particulier ne signale la "nouveau" d'un référent [...] : double tâche, donc, pour le lecteur :

de niveaux hiérarchiques différents. La structure énonciative en constituants immédiats obéit en fait à ses propres lois et les principes de symétrie et d'homogénéité sont à rechercher au sein de son propre espace structural.

travail d'identification et de "stockage", dans la connaissance partagée d'un référent nouveau, et interprétation de l'information qui est donnée à propos de ce référent. Combettes (1988).

A rapprocher également des propos précédents, le point de vue complémentaire de Hazaël-Massieux (1974) portant sur le critère de "nouveau" que l'on attache trop souvent au rhème :

L'analyse en support / apport s'oppose à l'analyse en donné / nouveau (Halliday, 1967), en ce sens que l'apport n'est pas nécessairement "nouveau". Ce peut être au contraire quelque chose de donné qui est simplement "redonné" [...]. On pourrait dire [...] que l'apport est un nouvel élément dans le discours, mais n'est pas nécessairement un élément nouveau. Hazaël-Massieux (1974).

En tout état de cause cependant, ces remarques montrent que le plan de l'information est en quelque sorte surimposé au plan du support et de l'apport. Il y aurait ainsi un processus exclusivement sémantique (introduction de l'information) qui se développerait sur un autre, de nature plus syntaxique dans lequel la classe morphologique (unités lexicales) ou l'identité des constituants (groupes syntaxiques), et leur position dans l'énoncé de phrase seraient déterminants. Toutefois, il semble que ce processus envisagé au niveau syntaxique, ne puisse même pas exclure totalement le recours à la sémantique, notamment lorsque deux unités de même classe morphologique sont en présence : ainsi la préposition *de* (ou *du*) peut introduire un *support* comme dans *il apprécie le talent de l'artiste*, mais peut aussi introduire un *apport* comme dans *il aime le vent du sud*.

Pour revenir à notre texte et illustrer notre méthode d'analyse, nous nous reportons au graphique n° 3-3. Dans cet exemple, le groupe minimal "pour des vers géants" constitue par rapport à l'unité immédiatement supérieure dont il est issu (un support de rang 1), l'apport d'information relatif de plus haut niveau (rang 2), ce qui précède "d'éminents biologistes ... créé" lui fournissant le support formel (rang 2) nécessaire et suffisant à son expression. Le mécanisme d'affectation des poids est le même que celui décrit à propos du modèle en constituants syntaxiques HR.

En conclusion, le modèle de la hiérarchie énonciative EN représentait en 1991 une création originale des relations thème - rhème, et pouvait dans sa simplicité s'adapter à l'analyse d'un texte lu. Un modèle très proche de ce dernier, privilégie l'élément "apport de l'information" dans la structure hiérarchique en augmentant la pondération (+1), et ceci aboutit au modèle ER (graphique n° 3-3 à l'annexe du chapitre).

La quantification des modèles EN et ER est comme pour les précédents, iconique : plus un constituant énonciatif est haut dans la hiérarchie, plus son poids est grand, et plus il est censé déterminer sur le plan mélodique, un F0 élevé ou un écart de F0 important dans le mot lexical.

3.2. MODELE DE LA COMPLEXITE SEMANTIQUE (CM)

Ce modèle tente d'analyser les indices qui concourent à complexifier la signification d'un mot, ce qui augmente le traitement cognitif. Cette analyse s'appuie sur les travaux qui sont opérés dans le domaine psycholinguistique. Ainsi dans une revue de travaux portant sur l'identification des mots pendant la lecture et la durée des mouvements oculaires, Hazaël-Massieux (1974) précisent en effet que

la durée de fixation est fonction de la difficulté de traitement d'un mot (cf. entre autres, Rayner and Duffy, 1987). Ainsi par exemple, on observe des durées de fixation très longues sur les termes rares, situés dans des contextes peu contraignants, alors que les mots

fortement prédictibles en fonction du contexte, tout comme les termes fréquents, sont fixés moins longtemps. Hazaël-Massieux (1974).

Cette distinction entre mots courants, rares, ou plus accessibles, attestée dans de nombreuses études sur l'accès lexical en lecture, permet donc de définir un principe d'analyse. Une hypothèse vraisemblable est de poser qu'il n'y a pas lieu de différencier les processus cognitifs qui interviennent en phase de compréhension ou en phase de production, puisque les contraintes explicites données au locuteur pour sa lecture mettent l'accent sur le caractère intelligible du discours : processus de production et processus de compréhension doivent ainsi converger. Ceci concerne à la fois le modèle de la complexité lexicale CM et le modèle de la connaissance partagée CP que nous présenterons par la suite.

Plus que les modèles de l'énonciation, le modèle de la complexité sémantique, qui repose plus précisément sur la complexité intrinsèque et contextuelle, laisse une part plus importante à la notion de *sèmes*. Selon Le Ny et Denis (1980) en effet,

la construction par un sujet compreneur (et éventuellement par un compreneur artificiel) de la signification d'une fraction de la phrase, inclut l'amorçage non, comme dans la théorie lexicale, de lexèmes figurant dans le lexique du sujet, avec leurs relations préétablies, mais de traits sémantiques sélectionnés dans les signifiés, ceux des traits qui sont pertinents pour la phrase considérée. Le Ny et Denis (1980).

Cette notion de *sème* étant fondamentale, il faut l'analyser plus à fond. Les sémanticiens distinguent généralement les *sèmes inhérents* et les *sèmes afférents*. Selon Rastier (1987), "les sèmes inhérents relèvent du système fonctionnel de la langue, et les sèmes afférents, d'autres types de codifications : normes socialisées, voire idiolectales."

Cette différence de nature entre les deux types de sèmes a suscité des prises de position opposées parmi les linguistes : ainsi Greimas et Courtès (1979) ne prennent pas les sèmes afférents en considération, Kerbrat (1979) les traite à part, tandis que d'autres (Kiefer, 1974; Groupe Mu, 1976; Martin, 1983) leur refusent le statut de sèmes.

Toutefois comme le note Rastier (id.), "même si l'on admet que ces composants sont secondaires, ils n'en sont pas moins importants" car "pour comprendre un texte, la connaissance du système fonctionnel de la langue ne suffit pas, [...] il faut connaître toutes sortes d'autres codes". Pour donner un seul exemple (tiré de J. Audiberti) à l'appui de cette thèse à laquelle nous souscrivons, il est facile de montrer avec Rastier (id) que dans l'exemple suivant "l'adjectif lexicalise un sème afférent au nom" : "*Je ne vais pas raconter la pièce, boulot transpirant*", dans lequel /sueur/ est un sème afférent à *boulot*. Comme on le voit, on ne peut rendre compte de ce processus sémantique qu'en faisant référence aux sèmes afférents.

Cette réflexion amène l'auteur à distinguer *cohésion* et *cohérence*, dans ces termes :

[...] la *cohésion* d'un texte dépend de ses relations sémantiques internes, et sa *cohérence*, de ses relations avec son entour extralinguistique, défini comme l'ensemble des phénomènes sémiotiques qui lui sont associés. La cohésion est donc d'ordre intralinguistique; la cohérence, d'ordre intersémiotique. Rastier (1987).

Ces réflexions linguistiques nous mettent en mesure de mieux appréhender les catégories sémantiques du modèle que nous avons défini. Le modèle CM a fait l'objet de plusieurs communications antérieures (pour les plus anciennes, Caelen-Haumont, 1986, 1988). On peut toutefois rappeler que ce modèle tente de donner une vue d'ensemble des principaux indices qui interviennent dans la saisie de la complexité des mots lexicaux

en contexte et qui donnent prise à une quantification (graphique n° 3-4 ci-dessous à l'annexe).

Le modèle traite de l'accès à la signification et fait référence à une saisie du mot hors et en contexte. Ce schéma d'"accès au lexique" est en conformité avec les nombreux travaux menés par ailleurs sur le sujet en psycholinguistique, par exemple ceux de Le Ny (1985b) :

Au cours de la compréhension du discours, chaque audition ou lecture par le sujet d'une occurrence d'un mot (par ex. 'pomme') apparaissant au sein d'une phrase, (ou, à la rigueur, isolément) produit chez ce récepteur une chaîne d'événements mentaux ou cognitifs parmi lesquels on peut isoler les trois catégories suivantes : E1) le stimulus est perçu et identifié comme étant un exemplaire du mot (type) 'pomme'; E2) il se produit un événement sémantique appelé, selon le cas, 'accès' à la signification, ou 'production' (voire 'construction') de la signification du mot ; on pourrait ainsi parler [...] d'une identification du signifié du mot ; en d'autres termes encore, le récepteur saisit que c'est à une ou des pomme(s) que l'on fait référence ; E3) enfin, si le mot est inclus dans une phrase, sa signification particulière doit être intégrée à celle qui est construite par la phrase entière. Le Ny (1985b).

Pour notre part, le modèle fait référence à certains aspects des événements mentaux classés E2 et E3 par l'auteur.

Le modèle présente un module d'analyse intrasyntagmatique et un module d'analyse intersyntagmatique. La méthode prévoit d'évaluer successivement l'item lexical dans chacun de ces modules. Les catégories les plus élevées, et les catégories intermédiaires fonctionnent sur le mode additif (on passe en revue chaque embranchement, ex : un mot peut être à la fois d'un *registre* fondamental et désigner un *réfèrent* concret), alors que les catégories terminales (plus précisément, celles qui sont dotées d'un poids sur le graphique) fonctionnent sur le mode de l'exclusion (ex : le registre d'un mot est "fondamental", "standard" ou "spécialisé").

* Module intrasyntagmatique

Précisons tout d'abord avec Rastier que

du point de vue sémantique, chaque fois que deux sémèmes¹⁰ sont placés dans un même syntagme, ils peuvent interagir l'un sur l'autre en actualisant ou en virtualisant un ou plusieurs de leurs sèmes [...]. Alors le contenu du nom modifie celui de l'adjectif autant qu'il est modifié par lui. Rastier (1987).

Nous ajouterons qu'à notre sens, cette remarque ne vaut pas seulement pour la relation nom / adjectif, mais aussi pour d'autres comme verbe / adverbe, ou encore adverbe / adjectif. C'est dans cette perspective qu'a été menée l'analyse qui suit.

Au sein du syntagme le lexème est analysé en fonction de 5 faisceaux d'indices (soit 5 étapes) :

¹⁰ Par ailleurs, un *sémème* (ou contenu d'un morphème, soit d'un lexème ou d'un grammème) selon Rastier (id.), se structure d'une part en *classème* et d'autre part en *sémantème*, l'un et l'autre regroupant les sèmes inhérents et afférents, mais le premier des sèmes *génériques*, et le second, des sèmes *spécifiques*. "On appelle sème *générique* une propriété caractéristique des éléments de l'ensemble considéré, et sème *spécifique* une propriété caractéristique d'un élément de l'ensemble considéré."

En reprenant les définitions à Rastier (id.), "le *taxème* est l'ensemble de rang inférieur. Les sèmes spécifiques sont définis à l'intérieur du taxème, ainsi que certains sèmes génériques de faible généralité." Selon Coseriu (1976), c'est une "structure paradigmatique constituée par des unités lexicales ('lexèmes') se partageant une zone commune de signification et se trouvant en opposition immédiate les unes avec les autres". Mais comme le remarque Rastier, les grammèmes peuvent se constituer également en taxèmes.

1° registres fondamental, standard / spécialisé banalisé, et spécialisé

Le registre fondamental fait référence au vocabulaire le plus simple, généralement¹¹ mentionné dans les différents dictionnaires qui analysent ce niveau, que ce soit *L'Élaboration du Français Fondamental* (Gougenheim et al., 1964), les *Listes Orthographiques de base du Français* (Catach, 1984) ou d'autres encore. Le registre spécialisé correspond aux termes spécifiques non tombés dans l'usage commun et dans le cas de l'application de ce modèle à la lecture, il est conçu comme n'appartenant pas à la spécialité du locuteur ; dans le cas contraire, le modèle renvoie au registre standard ou spécialisé banalisé. Le poids le plus fort du modèle est attribué au registre spécialisé.

2° le trait concret, concret / abstrait et abstrait ou imaginaire correspond à la reconnaissance d'un sème du niveau de généralité le plus grand, la *dimension*. Seule la catégorie concret / abstrait demande à être explicitée : elle recouvre en effet des termes qui lexicalement concrets s'appliquent cependant par l'effet de textualisation à des notions abstraites ou imaginaires, ou bien désignent des termes qui présentent un référent concret sous une catégorisation abstraite (ex: "biologiste" : être humain concret + catégorisation abstraite).

3° notion d'essence

L'analyse de ce niveau retrouve en beaucoup de points la conception de Lyons (1980). Selon l'auteur en effet la dénotation est indépendante des classes syntaxiques et morphosyntaxiques.

Aux deux catégories de substance et d'attribut (qui ne reçoivent une quantification qu'avec leurs catégories subordonnées) correspondent :

- pour la substance, 2 niveaux d'analyse

(a) Le premier distingue les sous-catégories d'état et de procès qui s'opposent soit comme les notions de structure et d'évolution, soit comme les notions d'espace et de temps (aspect statique vs. dynamique), avec la possibilité d'une combinaison de ces notions duales. Des termes comme "addition" par exemple peuvent contextuellement tour à tour désigner un état, un procès, ou la combinaison des deux (état résultant d'un procès).

Pour Lyons (id.),

noms et verbes sont dans un même rapport sémantique avec la catégorie ontologique d'état, qui correspond à ce que nous avons appelé *dénotation* [...]. L'existence de lexème tels que *paix* ou *connaître* [...] peuvent tous les deux être considérés comme dénotant des états bien qu'ils appartiennent à des classes d'expressions syntaxiquement et morphosyntaxiquement distinctes. Lyons (1980).

(b) Le deuxième envisage quatre types d'opposition. Les étiquettes discontinuité / continuité renvoient à la fois aux notions d'état et de procès.

¹¹ Comme le soulignent d'ailleurs les auteurs de ces différents ouvrages, le vocabulaire fondamental du français est une entité difficile à cerner : d'une part parce que le vocabulaire fondamental répertorié dans les textes ne recouvre pas par exemple le vocabulaire fondamental de la réalité de tous les jours et d'autre part parce que ce vocabulaire résulte d'une évaluation parfois subjective, et parfois obsolète du fait de l'évolution naturelle de la société et des réalités concrètes qu'elle recouvre ... La méthodologie que nous préconisons est de vérifier si le vocabulaire figure dans l'un des ouvrages. S'il ne figure dans aucun, le vocabulaire appartient au registre "standard". S'il ne figure que dans une partie de ceux-ci et si le doute subsiste, il sera alors également classé dans le registre "standard".

En effet il apparaît que la notion statique est réintroduite de manière secondaire dans la catégorie du procès lorsque des sèmes lexicaux de continuité (au sens d'un mécanisme récurrent) se surajoutent aux précédents, et corrélativement la notion dynamique dans la catégorie de l'état par la notion de "discontinuité". Ceci rejoint encore le point de vue de Lyons (id.) pour lequel l'aspect ne dénote pas seulement les verbes : "les noms et les adjectifs peuvent également avoir un caractère aspectuel. [...]. Le nom [...] *explosion* en anglais, par exemple, dénote un événement alors que *peace* dénote un état."

Selon qu'il s'agit donc de l'une ou l'autre catégorie, les notions de continuité et de discontinuité sont significatives (poids +1) ou redondantes (poids 0), et de ce fait les poids associés aux étiquettes laissent cette double possibilité (0/+1). Le modèle prévoit une neutralisation possible de ces catégories (poids 0) dans le cas par exemple des termes abstraits, et d'autre part inversement, une combinaison de ces notions (poids +1). Lorsque les deux catégories état et procès sont combinées, il s'opère aussi une neutralisation des notions de continuité / discontinuité (poids 0).

- pour l'attribut, un seul niveau d'analyse.

La notion d'attribut est complexe dans la mesure où elle renvoie aux qualités d'un objet considéré comme état ou procès : même si les contenus de ces deux types d'attribut ne sont pas comparables, le processus de caractérisation est dans notre perspective, analysable par les mêmes voies. En effet cette notion d'attribut peut renvoyer à l'analyse des adjectifs qualificatifs ou des adverbes, les uns et les autres précisant de manière comparable des qualités ou des modalités attachées aux noms, adjectifs, adverbes et verbes.

Cette perspective rejoint à nouveau celle de Lyons qui considère que certaines expressions syntaxiquement différentes sont liées par une relation transformationnelle :

[dans la langue] il existe fréquemment, sinon toujours, une relation qu'on peut considérer comme transformationnelle entre des expressions contenant un adjectif accompagné d'un adverbe de degré d'une part, et les noms du deuxième ordre modifiés par un adjectif, d'autre part : [la] correspondance [...] entre *étonnamment beau* et une *beauté étonnante*. [...] il existe la même relation d'ordre transformationnel entre le nom du deuxième ordre modifié par un adjectif et le verbe correspondant modifié par un adverbe (cf *articulation claire* : *articuler clairement*). Lyons (1980).

Dans cette catégorie de l'attribut, on distingue entre les attributs intrinsèques (redondance des sèmes de l'attribut par rapport à la notion désignée, ce qui renvoie au processus de détermination interne), les attributs de complémentarité (pas de redondance mais concordance des sèmes), et les attributs extrinsèques¹² (addition de sèmes), ces deux derniers attributs renvoyant aux processus de détermination externe. Cette notion d'attribut étant par ailleurs à nouveau intégrée, bien que dans une autre perspective, dans le modèle de la connaissance partagée CP, nous reviendrons sur cette analyse et préciserons les choses à l'aide d'un exemple.

* Module intersyntagmatique

Dans ce module, les relations entre lexèmes sont saisies sur un plan supérieur, à savoir par l'intermédiaire des syntagmes.

¹² Ces attributs extrinsèques associés à une dénomination générique désignent parfois une notion spécifique : dans ce cas la pertinence de l'information peut reposer essentiellement sur l'adjectif. Les conditions du glissement de l'adjectif vers la catégorie du substantif sont alors réunies (ex : l'océan atlantique → l'atlantique ; le chemin de fer métropolitain → le métro ...).

Le module intersyntagmatique prévoit trois niveaux d'analyse. Au niveau formel, une seule catégorie concernant le réemploi d'un terme de registre standard ou spécialisé est envisagée, et ce réemploi est conçu comme une diminution de la complexité dans ce registre. Ainsi un terme spécialisé réutilisé perd de sa complexité, il redevient l'équivalent d'un terme spécialisé banalisé. Dans le cas du registre fondamental, il y a neutralisation.

Ce point de vue s'appuie sur l'expérimentation en psychologie. Le Ny (1985) montre ainsi que "l'existence d'une relation associative produit une facilitation de la décision lexicale sur le second mot, manifestée par un temps de décision plus court."

Le deuxième niveau est celui de la figuration envisagée selon 3 degrés d'expressivité croissante : pas de figuration ou cliché, figuration lexicalisée (ex : "à brûle pourpoint"), figuration originale.

Le troisième niveau envisage l'appartenance d'un terme à un champ lexical et cette appartenance, du fait du traitement qui a déjà été opéré par le locuteur à propos de ce terme, tend à réduire la complexité de ce terme.

Pour ce modèle, la quantification des catégories a été déterminée en repérant d'abord les notions qui convoient plus de complexité que d'autres, en calculant de manière ascendante le poids des plus complexes en prenant en compte les catégories à la complexité la plus faible au poids minimal, et en veillant à l'équilibre respectif des poids. Par exemple, un mot lexical qui dans le module intrasyntagmatique recueillerait le maximum de la quantification, tout en n'appartenant pas au registre spécialisé, totaliserait le poids 7 : dans ces conditions, le registre spécialisé qui est certainement la catégorie la plus complexe, ne peut pas avoir un poids inférieur à 7. Autre exemple, concernant l'équilibre des poids, une répétition d'un mot (catégorie "rappel") spécialisé, baisse le total de 3, ce qui le ramène au poids d'un mot standard.

Chaque mot lexical s'enrichit donc successivement des poids affectés aux catégories qui lui correspondent, et addition faite à l'issue de l'ensemble, le poids final rend donc compte d'une évaluation approchée de sa complexité.

Comme pour les modèles précédents, la quantification est iconique dans la mesure où plus le poids est élevé, plus le mot est jugé complexe, et plus il est considéré comme susceptible de générer un F0 élevé ou un écart de F0 important.

4. MODELE PRAGMATIQUE

4.1. MODELE DE LA CONNAISSANCE PARTAGEE (CP)

Dans une perspective linguistique, pragmatique et sémantique sont des notions apparentées mais distinctes. Pour Martin (1983) "[la sémantique] rend compte du sens, [la pragmatique] de ses interprétations". Ainsi dans sa relation à la sémantique, le champ de la pragmatique prend effet dans l'interaction entre le sens d'un énoncé et l'activité du sujet. Pour notre part lorsque nous ferons allusion à l'interprétation du sujet en situation (consignes de lecture), nous parlerons de "pragmatique".

Dans cette étude nous n'envisagerons qu'un seul modèle pragmatique, à savoir celui de la connaissance partagée ou CP. Ce modèle a repris, mais en le développant amplement, le modèle original conçu par Prince (1981, 1983), commenté, appliqué et illustré ensuite dans sa version originale, à l'étude des textes français des ouvrages scolaires par Combettes et al. (1988).

Les auteurs apportent plusieurs précisions intéressantes, notamment que "la connaissance partagée est, en réalité, *supposée* (elle correspond à ce que l'émetteur s'imagine des connaissances du récepteur) et ne correspond que rarement à une connaissance commune réelle".

Par ailleurs une application de ce modèle a été faite pour l'étude de l'intonation de l'anglais (Brown, 1983) dans le contexte de courts dialogues finalisés. Les catégories fortement inspirées de Prince sont les suivantes : *inédit, nouveau, nouveau inféré, évoqué par la situation, évoqué par le texte, évoqué par le texte mais déplacé*. Il est montré que les catégories de l'inédit, du nouveau, du nouveau inférable sont associées à un F0 élevé, tandis que les catégories de l'évoqué, sont liées à un F0 bas.

Nous sommes ici en fait, et quel que soit le modèle linguistique sous-jacent, devant deux explications possibles de la prosodie : soit les traits prosodiques sont des symptômes, à valeur iconique, de traits linguistiques, soit ce sont des traits linguistiques, comme le pense Rossi (communication personnelle). Pour ma part, je pense qu'il n'y a pas de relation directe entre contenu linguistique et expression mélodique, mais médiatisée par une intention, un vouloir-dire de nature, bien évidemment psycholinguistique (et émotionnel). Cela signifie donc que si le locuteur ne perçoit pas que tel mot contient une information inattendue, la mélodie lexicale sera réalisée comme plus plate, ce qui, évidemment, n'est pas forcément la réalisation d'un autre locuteur : un même mot peut susciter des mélodies largement différentes, mais les locuteurs peuvent aussi partager, dans les mêmes conditions de lecture, les mêmes intentions. Cet ouvrage fait état d'une réalité dans sa complexité, à la fois avec des vouloir-dire identiques et différents.

Ainsi, ce modèle CP se présente comme un module d'analyse des informations attendues et inattendues du texte. L'analyse *lexicale* que nous proposons se présente comme l'aboutissement des opérations psycholinguistiques de textualisation : les mots lexicaux sont perçus en bout de chaîne comme porteurs des indices de ces opérations.

Dans le graphique n° 3-5 ci-dessous à l'annexe, les encadrés grisés correspondent à la version de Prince (1983) dans son intégralité. Pour pouvoir tirer une évaluation de chaque mot lexical en contexte, les modèles antérieurs de Prince et de Brown ont dû être fortement complétés. Le modèle tel que nous le proposons, tente d'évaluer les différents processus que les lecteurs (et les auditeurs) mettent en oeuvre dans le traitement de l'information textuelle : mais surtout ce modèle traduit la représentation que s'en font les locuteurs de manière à pouvoir la prétraiter pour autrui.

Les notions utilisées sont dans l'ensemble assez explicites par elles-mêmes : le commentaire se réduira donc de ce fait à quelques éléments.

Dans le traitement de ce qui est "nouveau" il nous a semblé utile, afin de mieux analyser le phénomène, de distinguer entre différents degrés de nouveauté, à savoir ce qui est proprement inédit et ce qui n'est que partiellement nouveau (cf "relatif"). Ce "nouveau relatif" se subdivise à son tour en deux catégories, relatives la première à la notion de temps¹³ et la deuxième à la notion d'espace textuel. Ainsi pour le temporel, la catégorie de "non utilisé" renvoie — d'après Prince — à ce qui n'a pas été utilisé dans le contexte précédent à court ou moyen terme.

Comme on le sait, la notion d'*isotopie* a été empruntée par Greimas au domaine de la chimie pour désigner la "réurrence de catégories sémiques" "qui assurent au discours-énoncé son homogénéité" (Greimas et al., 1979). En clair le terme d'isotopie renvoie aux sèmes récurrents dans le discours à qui ils donnent une cohérence sur le plan du sens. Il est intéressant de noter que Greimas attachait dès 1966 le concept d'isotopie à la notion de lecture : "Par isotopie, nous entendons un ensemble redondant de catégories sémantiques qui rend possible la lecture uniforme du récit."

¹³ Cette notion de "temps" est de toute façon réinvestie en notion d'espace textuel.

Kerbrat-Orecchioni (1979), quant à elle, en regrettant "la polysémie de ce métaterme", recense une multitude d'isotopies qui n'appartiennent pas toutes au domaine de la sémantique, à savoir les isotopies stylistiques, énonciatives, rhétoriques, présuppositionnelles, syntaxiques, narratives, dénotées, connotées, organisées et non organisées.

Pour Pottier (1974), l'isotopie est en liaison avec le principe de cohérence textuelle, et il oppose l'*isosémie* ou "redondance d'un sème à travers divers éléments d'une séquence", à l'*isotopie* ou "continuité sémantique à travers une séquence (reflétant une cohérence)", l'une et l'autre étant liées par un rapport de cause à effet, l'isotopie résultant d'une ou plusieurs isosémies.

Deux remarques s'imposent cependant. La première revient à Rastier (1987) qui constate que "ce n'est pas seulement la récurrence de sèmes donnés qui constitue l'isotopie, mais la présomption d'isotopie qui permet d'actualiser des sèmes, voire *les* sèmes."

La deuxième remarque voudrait réintroduire une notion de dynamique dans la définition : il est bien clair que dans le domaine sémantique, les sèmes récurrents (aspect statique) sont les indices de la constitution ou de l'existence d'une isotopie, mais il semble tout aussi clair que l'isotopie en vertu du principe de linéarisation du texte, doit, pour alimenter son actualisation, incorporer des sèmes nouveaux. Ces sèmes nouveaux seront introduits par l'intermédiaire de sémèmes qui comporteront également des sèmes récurrents qui assurent la cohérence, ou de sémèmes au(x) sème(s) tous nouveau(x) mais compatible(s) avec un ou plusieurs sèmes des sémèmes isotopiques. Il n'est pas exclu que, ce faisant, on passe — et pour reprendre la terminologie de Rastier — d'une isotopie moins générale à une autre plus générale, depuis l'isotopie spécifique (récurrence de sèmes spécifiques), aux isotopies micro-, méso- ou macrogénériques (récurrence de sèmes génériques de plus ou moins grande généralité).

Par conséquent dans notre perspective, les éléments informatifs, relativement à l'espace textuel, interviennent dans un contexte "d'isotopie sémantique", catégorie nouvelle par rapport au modèle de Prince, c'est-à-dire dans un contexte de structure sémantique complexe reposant sur une imbrication de champs lexicaux subordonnés — abstraits ou figuratifs — où ils apportent sur la base de sèmes récurrents, les sèmes nouveaux qui constituent les nouveaux indices du sens. Ces sèmes seuls ou avec d'autres qui font référence au "déjà connu" et qui de ce fait ne sont pas informatifs, alimentent l'isotopie sémantique. A ce titre ils doivent être acheminés vers les catégories de l'inférable, mais du fait de la nouveauté de leurs sèmes, ils bénéficient, comme on le constate sur le graphique 3-5 à l'annexe, d'une pondération supplémentaire.

Le commentaire porte maintenant sur l'explicitation des notions de "supplément de sens" et de "complément de sens". Le supplément de sens renvoie aux sèmes d'un premier sémème qui ne sont pas intrinsèquement contenus dans le deuxième sémème qui lui est associé. Ces sèmes peuvent qualifier l'item lexical de deux manières, soit en s'opposant aux sèmes de l'item, soit au contraire en les prolongeant. Nous retrouvons les notions abordées précédemment dans le modèle de la complexité lexicale (modèle CM). Le fait qu'elles soient analysées non pas dans la perspective de catégories linguistiques, mais d'opérations de type cognitif, justifie que l'on puisse les réintroduire dans un nouveau modèle.

La catégorie de supplément de sens reprenant la notion de *sèmes extrinsèques* (Caelen-Haumont, 1986), développée ultérieurement avec les notions de *contradiction*, d'*exclusion*, d'*opposition*, converge avec l'analyse de Rastier (1987) qui distingue au

sein de l'*allotopie antonymique*, les relations entre *contradictaires*, *contraires logiques*, et *contraires non logiques*. Pour illustrer prenons un exemple, telle cette phrase qui pourrait être prononcée dans un atelier de menuiserie par un artisan à son apprenti : "Prends la chaise en bois bancal qui est sans dossier".

Dans cette phrase l'élément *en bois* ajoute une précision complémentaire (cf catégorie "complément de sens") dans la mesure où le bois représente une des matières possibles de la chaise, et totalement compatible avec les sèmes intrinsèques du mot chaise. Le terme *bancal* quant à lui, constitue un supplément de sens dans la mesure où la propriété *bancal* n'entre pas dans la définition de la chaise, mais ce supplément n'est cependant pas en conflit avec les sèmes intrinsèques : dans ces conditions le mot *bancal* correspond au cas du supplément de sens par addition. Enfin le qualificatif *sans dossier* apporte lui aussi un supplément de sens mais il est contradictoire avec les sèmes intrinsèques du mot *chaise* qui, selon le dictionnaire Robert se définit comme "un siège avec dossier sans accoudoirs". Ce cas est donc traité par la catégorie "supplément de sens par contradiction".

Par ailleurs l'expression de *journal hebdomadaire* constitue un exemple de supplément de sens par opposition, et la *sombre clarté* qui tombe des étoiles" de Hugo, celui de supplément de sens par exclusion. Combettes (1984) oppose de la même façon les *contradictaires*, pour lesquels "joue la loi du tiers exclu", et les *contraires* qui "dénotent les pôles extrêmes d'une dimension et autorisent la constitution d'une zone intermédiaire, non définie", qui regroupent chez nous l'exclusion et l'opposition.

Le dernier éclaircissement des notions présentées dans le modèle porte, dans le module de l'*opération logique*, sur la signification de l'étiquette *déduction notionnelle*. L'inférence porte dans ce cas sur des unités lexicales qui sont liées entre elles par une filiation nécessaire des sèmes inhérents ou afférents. C'est ici le domaine des figures de rhétorique, et en particulier de la métonymie et de la synecdoque. Par exemple le mot *ambulancier* évoque directement les mots *ambulance*, *malade*, et *hôpital* ... Entre le mot *ambulancier* et le mot *hôpital* rencontrés par exemple en contexte, une opération de déduction notionnelle a été réalisée.

Dans le même module, la catégorie d'*inclusion* qui est sans ambiguïté ne nécessite pas de commentaire.

Pour l'application à notre texte, certaines unités lexicales (cf "nouveau", "nombreuses", "longs" et "vers", deuxième instanciation, et "élevée") nous ont semblé pouvoir relever de deux interprétations différentes : la redondance des sèmes par rapport au contexte (créer / nouveau, géant / long, chaud / élevé), ou la répétition des mots (cf. "vers") peut être considérée comme une redondance ou une réactualisation non significative, ou au contraire comme une insistance. Pour mieux répondre à ces deux interprétations, les deux poids possibles ont été retenus dans le modèle.

Les unités lexicales "nombreuses" et "élevée" peuvent également être perçues comme des éléments de supplément de sens par opposition, la création d'un phylum, surprenante en raison de la quantité énorme des espèces à ce jour répertoriée, étant marquée par le terme "nombreuses", ou comme une simple redondance. Par ailleurs, la température moyenne élevée dans les profondeurs marines peut représenter un fait surprenant donc un "supplément de sens" par opposition, ou peut être simplement perçue comme complément de sens. Ces différences de signification ont été retenues dans la quantification.

Comme on le constate sur le graphique n° 3-5, le modèle procède des opérations cognitives jugées les moins complexes vers les plus complexes. Les poids reflètent donc cet ordre croissant (de haut en bas), à l'exception de la première catégorie ("insistance")

qui marque une intention particulière. Cet ordre croissant admet des paliers pour des catégories jugées comparables sur le plan de l'apport d'information et de son traitement.

Cette quantification est comme précédemment d'ordre iconique, puisqu'elle sous-entend que les valeurs les plus élevées de F0 ou l'écart de F0 le plus important, sont déterminés par les mots lexicaux qui possèdent le poids le plus élevé, marque du traitement cognitif le plus important.

5. CONCLUSION

Nous avons présenté la structure de 6 modèles linguistiques, leurs principes de quantification et leur application aux unités lexicales du texte. Ces modèles sont généraux et peuvent s'appliquer à de nombreux textes et types de texte. Nous ne prétendons pas que ces modèles soient les seuls modèles qui puissent rendre compte des réalisations des locuteurs : vu l'extraordinaire variété des textes et des discours, de leurs contenus, des intentions et des situations, nous sommes convaincue précisément du contraire. Les modèles retenus nous ont paru toutefois traiter des processus linguistiques centraux de tout texte, et par là offrir une légitimité à notre projet.

Les caractéristiques de ces modèles ont une portée générale dans la mesure où elles filtrent les indices porteurs de sens qui concernent tous les mots lexicaux du texte, mais à des degrés divers : c'est un point capital. D'autres processus linguistiques intéressants existent, mais parfois dépendants du type de texte, ils sont alors sans objet pour notre expérimentation.

Même compatibles ou indépendants du type de texte, leur granularité cependant n'est pas assez fine, autrement dit, ils ne permettent pas d'établir une différenciation suffisante des mots lexicaux : par exemple, le processus de catégorisation hyponymique et hyperonymique représente en soi une échelle tout à fait digne d'intérêt, mais même si l'étude arrive à distinguer plusieurs strates combinées avec plusieurs catégories (en intension et extension), ces différents plans risquent d'offrir une répartition très déséquilibrée des items à la fois dans l'espace du texte, et à la fois dans le nombre des items par catégorie. Il y a tout lieu de penser en effet que le texte fonctionne par niveau de catégorisation hyponymique ou hyperonymique semblable, et dans ces conditions les oppositions concernent plus des parties de texte entre elles que les lexèmes, ce qui ne semble pas compatible avec notre méthode.

Nous faisons l'hypothèse que les caractéristiques, la diversité des modèles que nous avons définis sont suffisamment proches des moyens que met en oeuvre la prosodie pour les mettre à l'épreuve, et c'est cette démarche que nous décrivons dans les prochains chapitres.

Les valeurs de quantification ont pour objet de prédire de manière contradictoire l'organisation prosodique des énoncés, contradictoire dans la mesure où les modèles proposent des stratégies d'élocution concurrentes. En effet les modèles portent sur la signification, c'est-à-dire sur un ensemble d'hypothèses laissées à l'interprétation du locuteur, hypothèses à partir desquelles il construit sélectivement le sens qui lui est personnel, et qu'il transmet à son auditeur par l'intermédiaire d'indices mélodiques, qui traduisent son "choix" sur le plan linguistique et pragmatique. Le Ny (1987) confirme d'ailleurs sur la base d'expérimentations psycholinguistiques solides, que "contrairement à ce dont témoigne notre conscience, il n'y a pas d'accès unique, immédiat et automatique à la signification contextualisée."

Notre étude se situant au niveau des constituants (modèles hiérarchiques) et au niveau du mot lexical (modèles de dépendance, de la complexité sémantique, de la connaissance supposée partagée), concernera dans une certaine mesure le plan de l'intonation, mais plus encore celui de la mélodie lexicale. Nous verrons notamment au

cours du chapitre IV comment le plan de la mélodie lexicale peut être étudié indépendamment du plan de l'intonation. Cependant que la mélodie résulte d'un traitement sur l'ensemble de la phrase ou plus local, les valeurs mélodiques sont saisies dans l'espace lexical. Cette méthode permet donc de mettre les modèles en compétition.

Avant d'analyser les relations entre modèles linguistiques et indices prosodiques, on posera le problème de l'homogénéité des données quantitatives d'une part entre les modèles eux-mêmes et d'autre part entre les modèles et les indices prosodiques. On ne peut évidemment comparer valablement les stratégies des locuteurs si la forme des modèles, c'est-à-dire la fourchette des poids, est différente, et si par ailleurs les faits prosodiques ne sont pas purgés des influences qui n'ont pas été envisagées par les modèles, à savoir ce qui relève des faits microprosodiques, du niveau socioculturel du locuteur, des différences intra- et inter-locuteurs, et même ce qui relève d'un niveau plus linguistique, celui de la modalité des phrases par exemple.

Le prochain chapitre est donc consacré à des points de méthodologie permettant de résoudre de manière simple ces différents problèmes.

ANNEXE : REPRÉSENTATION DES MODÈLES

Les pages qui suivent présentent les modèles appliqués au texte. Il existe deux types de modèles : d'une part les modèles qui présentent soit une structure hiérarchique, comme les modèles en constituants immédiats syntaxiques (modèle HR) ou sémantiques énonciatifs (modèles EN et ER) ou qui résultant d'une analyse syntaxique simple, s'appliquent directement sur les unités de surface (modèle DP) et d'autre part les modèles qui nécessitent une évaluation plus complexe, faisant intervenir pour chaque unité lexicale, plusieurs catégories. Ces modèles exercent leur analyse dans le domaine sémantique (modèle CM) et le domaine pragmatique (modèle CP).

Les quatre premiers modèles (HR, EN, ER et DP) seront illustrés dans le cadre de la phrase 1, de manière à pouvoir confronter les structures ou les analyses sur une même base. En ce qui concerne le modèle HR, les symboles utilisés sont les symboles standards. Pour ce qui est des modèles de l'énonciation EN et ER, le symbole "S" renvoie à la notion de support (ou thème dans la théorie standard), "A" à celle d'apport (ou rhème).

*Enfin pour les modèles hiérarchiques HR, EN et ER, les nombres en correspondance des unités lexicales indiquent la hauteur de leur niveau dans la hiérarchie. Une seule exception figure dans cette quantification : le dernier mot de chaque phrase ("vivantes" pour la phrase 1) a reçu une pondération qui correspond à la hauteur du premier groupe auquel il se rattache et non pas à celle qui devrait être la sienne (fin du constituant phrase). Ce fait est symbolisé par l'astérisque *.*

En ce qui concerne le modèle DP, les catégories, les symboles et la quantification utilisés sont les suivants :

- | | |
|----------------------------------|---|
| - Dépendance locale : DL → 1 | - Subordination Indirecte : SI → 3 |
| - Indépendance locale : IL → 2 | - Transsubordination : T → 3 |
| - Indépendance finale : IF → 2 | - Coordination et juxtaposition : C → 4 |
| - Subordination directe : SD → 2 | - Indépendance : I → 5. |

Les deux derniers (CM et CP) seront présentés non pas en fonction des phrases car les cheminements sont trop nombreux, mais dans leurs structures de définition. Le modèle CM dans son organigramme de fonctionnement prévoit que l'on passe successivement dans les deux modules d'analyse des relations intrasyntagmatiques et des relations intersyntagmatiques, alors que le modèle CP propose un cheminement par sélection exclusive des catégories.

Le texte de la phrase 1 est le suivant :

"D'éminents biologistes et d'éminents zoologistes américains ont créé pour des vers géants un nouveau phylum dans l'actuelle classification des nombreuses espèces vivantes".

	HR	1	2	3	4	
éminents	2		ADJ			P
biologistes	3		N	GN		
éminents	2		ADJ			
zoologistes	3		N	GN		
américains	4		ADJ	GN	SN1	
créé	3		V	GV		
vers	2		N			
géants	3		ADJ	GP		
nouveau	2		ADJ			
phylum	3		N	GN2		
actuelle	1	ADJ				
classification	2	N	GP			
nombreuses	1	ADJ				
espèces	1	N				
vivantes	2*	ADJ	GP	GP	SV	

Graphique n° 3-1
Modèle HR : Structure en constituants syntaxiques immédiats.

éminents	1	DL		ADJ			P
biologistes	4	C		N	GN	SN1	
éminents	1	DL		ADJ	GN		
zoologistes	2	IL		N	GN		
américains	5	I		ADJ			
créé	3	SI		V	GV		
vers	2	IL		N	GP		
géants	3	T		ADJ			
nouveau	1	DL		ADJ	GN2		
phylum	3	T		N			
actuelle	1	DL	ADJ	GP			
classification	3	SI	N				
nombreuses	1	DL	ADJ				
espèces	2	IL	N				
vivantes	2	IF	ADJ	GP	GP	SV	

Graphique n° 3-2
Modèle DP : Relations de dépendance des unités lexicales.

	EN	ER	1	2	3	4	5	
éminents	1	2	A					
biologistes	2	2	S	S				
éminents	1	2	A					
zoologistes	2	2	S	S				
américains	3	4		A	S			
créé	4	5			A	S		
vers	3	3			S			
géants	5	6			A	A	S	P
nouveau	3	4			A			
phylum	4	4			S	A		
actuelle	2	3		A				
classification	3	3		S	A			
nombreuses	2	3		A				
espèces	1	1	S					
vivantes	3*	4*	A	S	S	S	A	

EN1 2 1 2 3 4 3 5 3 4 2 3 2 1 3*
ER 2 2 2 2 4 5 3 6 4 4 3 3 3 1 4*

Graphique n° 3-3

Modèles EN et ER : Structure en constituants sémantiques immédiats.

C O M P L E X I T E S E M A N T I Q U E	Inter- syntag- -matique	Champ lexical	Pas ou changement	+2				
			Appartenance	0				
		Figuration	Réelle	+5				
			Lexicalisée	+3				
			Pas ou cliché	0				
		Réactua- -lisation	Rappel	-3				
	Sans rappel		0					
	Intra- syntag- -matique	Essence	Attribut	Extrinsèque	+2			
				Complémentaire	+1			
				Intrinsèque	0			
		Substance	Procès	Combinaison	+2	Neutralisation	0	
				Etat	+1	Combinaison	+1	
						Continuité	0/+1	
						Discontinuité	0/+1	
			Etat	0	Neutralisation	0	Combinaison	+1
					Combinaison	+1	Continuité	0/+1
					Continuité	0/+1	Discontinuité	0/+1
					Discontinuité	0/+1	Neutralisation	0
		Référent	Abstrait	+4				
					Concret/abstrait	+2		
Concret					0			
Spécialisé	+7							
			Standard	+4				
Fondamental	+1							

Graphique n° 3-4

Modèle CM : Structure de définition du modèle CM. Ce graphique propose l'analyse de la complexité des unités lexicales en contexte, mais il ne doit pas être lu comme un organigramme. En effet pour le calcul des poids, il faut balayer successivement les différents modules (présentés verticalement), puis faire un choix entre les différentes rubriques (présentées horizontalement) du module assorties d'un poids. La somme algébrique des poids qui se présentent dans ce parcours définit le poids attaché à l'item lexical.

T Y P E D E C O N N A I S S A N C E S	Evoqué	Insistance				6			
		Redondance				1			
		Réactualisation				2			
	Inférable ↑ +2	Opération	Inclusion	Partitif			2		
				Hierarchie	Genre / Espèce	3			
		logique		Dédution notionnelle	Espèce / Genre		4		
								5	
		Isotopie		de sens	Complément de sens			5	
					Supplément	Par addition			6
					Par négation	Opposition			7
						Exclusion			7
		Contradiction				8			
		Nouveau	Relatif	Contexte			9		
	Non utilisé								
	Inédit			Rattaché au contexte			9		
				Non rattaché au contexte			10		

Graphique n° 3-5

Structure de définition du modèle CP. Ce modèle analyse la connaissance partagée ou autrement dit encore les informations attendues ou inattendues. Ce modèle propose l'analyse des unités lexicales en contexte. Les encadrés grisés rendent compte de la première version de ce modèle, qui est proposée dans Prince (1983). Le mode de calcul des poids est identique à celui du graphique 3-4

