

Démarche générative

Les conditions de la génération

La génération du texte, n'a pas pour objectif de restituer à l'identique le texte initial. Cette restitution à l'identique n'est qu'une des possibilités théoriques de la génération.

Plusieurs conditions doivent être réunies.

Il faut d'abord disposer, suite à l'analyse sémique, de l'arbre thématique du texte étudié.

Il faut déterminer un point d'entrée dans cet arbre, différents points d'entrées pouvant conduire à plusieurs organisations possibles. En fait, il est possible de démontrer que les textes normatifs, contrairement aux textes littéraires ou journalistiques, ont un nombre limité d'organisations possibles.

Il faut également disposer du dictionnaire sémémique. Il a nécessairement été constitué pour construire l'arbre taxinomique. Le dictionnaire sémémique permet de déceler les synonymies éventuelles et de générer des variantes fondées sur ces dernières ou sur des périphrases¹.

Au plan méthodologique, il convient de s'entendre sur les entités à manipuler.

Les *entités initiales* nous semblent devoir être de deux ordres :

- d'une part des mots et lexies, que nous désignerons en tant que molécules sémémiques car composés de un ou plusieurs sémèmes, dont les sèmes sont soit activés soit désactivés du fait de l'association.
- d'autre part des schèmes linguistiques élémentaires liés et sémantiquement spécifiés (diathèse, vision, topicalisation, focalisation, etc.).

Les *entités finales*, syntaxiquement correctes, sont des énoncés simples ou complexes.

Le cheminement pour aller du schème à l'énoncé doit être explicité.

A partir de schèmes élémentaires on peut construire directement des énoncés simples. Et à partir d'énoncés simples on peut construire d'autres énoncés simples ou des énoncés complexes. A partir d'un énoncé simple et d'un énoncé complexe on ne peut obtenir qu'un énoncé complexe.

¹ Nous ne développons pas ici les difficultés propres à ce type de démarche étudiée de manière approfondie par Catherine Fuchs (1994).

Schémes liés → syntactèmes → énoncés simples → énoncés simples ou complexes

Mais on peut traiter au niveau sémantique les schèmes pour obtenir des schèmes intégrés puis des énoncés simples, à partir desquels on construira des énoncés simples ou complexes.

Schémes liés → schèmes intégrés → syntactèmes → énoncés simples → énoncés simples ou complexes

Il faut revenir sur les définitions des énoncés simples et complexes données page **Erreur ! Signet non défini.**

Ce qui différencie l'énoncé simple de l'énoncé complexe, c'est l'unicité de la base d'une part, l'unicité du groupe verbal d'autre part. S'il y a au moins deux bases ou au moins deux groupes verbaux coordonnés, on est en présence d'un énoncé complexe.

Cette observation en appelle deux autres.

Il convient d'abord d'approfondir la notion de coordination.

Celle-ci est envisagée dans ses trois aspects (Pottier 1974 § 244) : égalité (ou), addition (et), soustraction (mais). Nous aurons à vérifier si ces trois aspects épuisent les possibilités de formalisation. Par exemple, l'addition peut connaître deux variantes selon que l'on doit rendre compte de deux types de comportement : séquentialité ou simultanité.

En second lieu, il faut souligner que la coordination se manifeste pour tous les composants de l'énoncé :

Dans l'énoncé simple, le fonctème nominal peut comprendre, outre un groupe substantival obligatoire, un ou plusieurs fonctèmes adjectivaux facultatifs. Le fonctème adjectival peut comprendre, outre l'adjectif, un ou plusieurs fonctèmes nominaux. Le fonctème verbal apparaît seul ou accompagné d'un ou plusieurs fonctèmes nominaux. Il peut en outre être composé, outre d'un syntagme verbal, de un ou plusieurs fonctèmes adjectivaux.

Ajoutons que la subordination (modèle syntaxique), correspondant à la mise en dépendance (modèle actanciel), est admise dans l'énoncé simple. Or, on peut avoir des subordinations en cascade. L'énoncé simple peut donc en définitive être relativement complexe.

Enfin, au-delà des coordinations explicites, on trouvera dans le texte des coordinations logiques implicites qui seront prises en compte dans le modèle conceptuel mais dont le rendu sous forme d'énoncés pourra s'avérer délicat.

Remarquons par ailleurs que la définition du schème intégré donnée par Pottier (1974 p. 330) comme "schème réunissant plusieurs schèmes d'entendement, et n'en conservant qu'un seul en actance, les autres se situant en dépendance (subordination)", crée une limitation particulière à laquelle il nous paraît inévitable de déroger de façon à obtenir la séquence générative suivante :

Schèmes liés → schèmes intégrés → syntactèmes → énoncés simples
→ énoncés simples ou complexes

D'ailleurs B. Pottier admet la possibilité de cette extension, puisqu'à partir de deux schèmes Pierre/arriver (SE¹) et Paul/arriver (SE²), il construit un SI par factorisation donnant comme résultat par exemple « Pierre et Paul sont arrivés ».

Mais l'on voit immédiatement que s'engager dans un élargissement de la définition du schème intégré conduit inéluctablement à proposer également un élargissement de la structure du syntactème en autorisant la coordination au niveau de la base et/ou du groupe verbal pour obtenir la séquence suivante :

Schèmes liés → schèmes intégrés → syntactèmes intégrés → énoncés simples ou complexes

Dans ces conditions, la combinatoire qui sera à la base des différentes variantes possibles du texte recomposé sera appliquée au niveau sémantique de préférence au niveau syntaxique.

Dernière observation. On peut s'interroger sur la nécessité de partir du niveau des schèmes élémentaires, lorsque ceux-ci ont pour objet de "désintégrer" des complexes substantivaux, adjectivaux ou verbaux, constitués par adjektivation.

Ainsi, dans "L'éducation est la première priorité nationale." nous proposons de ne considérer que SI₁ et non SE₁₁, SE₁₂ et SE₁₃ (cf. page **Erreur ! Signet non défini.**).

La question est plus délicate concernant la seconde phrase qui présente un cas remarquable de double factorisation.

La phrase "Le service public de l'éducation est conçu et organisé en fonction des élèves et des étudiants" est un énoncé complexe construits à partir de quatre énoncés simples imbriqués.

Pour reprendre les observations de B. Pottier (1974 p. 224), on peut se demander si l'on doit supposer quatre schèmes conceptuels de base, ou dire que le procédé d'intégration est mémorisé (=immédiatement conçu), et l'on aurait alors affaire à une *syntaxie* (ou modèle mémorisé), parallèle à la *lexie*.

Autrement dit, "conçu et organisé" et "élèves et étudiants" sont constitutifs de deux syntaxies qui renvoient à un schème intégré (SI₂) (cf. p. **Erreur ! Signet non défini.**), lequel renvoie à quatre schèmes d'entendement élémentaires SE₂₁, SE₂₂, SE₂₃, SE₂₄.

La question concrète qui est posée est de savoir s'il convient dans l'analyse relationnelle, dans l'optique de la génération, de descendre dans tous les cas au niveau des schèmes élémentaires.

Enfin, il faut également se poser la question de la reconnaissance automatique des schèmes élémentaires à intégrer, faute de quoi, il conviendrait de conserver au cours de l'analyse relationnelle la structure du schème intégré et la référence de ce schème intégré dans chaque schème élémentaire, auquel cas la décomposition du schème intégré en schèmes élémentaires ne se justifierait pas. En effet, le SI apparaîtrait comme une unité de signification dont les

éléments constitutifs ne sauraient avoir d'existence indépendante sans perdre une partie de leurs propriétés.

Sans préjuger d'exemples plus complexes qui pourraient se présenter par la suite, il nous paraît que la recomposition du SI à partir des SE paraît envisageable mais au prix d'une combinatoire particulièrement complexe et aléatoire.

Nous ne considérerons donc pas, au moins provisoirement, les SE comme les unités élémentaires de signification manipulables dans le processus de génération, ce processus consistant dans la constitution et la reconstitution des phrases, puis de l'assemblage des phrases en paragraphes de différents niveaux.

Nous nous attacherons dans la suite à mettre en évidence des structures textuelles du niveau de l'énoncé simple ou complexe donnant le support nécessaire à l'analyse et à la génération.

Les structures de données nécessaires à la génération sont en définitive :

- le modèle global ou thématique;
- les structures textuelles à définir
- les schèmes d'entendement formant un énoncé simple et attachés aux nœuds du modèle thématique;
- le dictionnaire sémique que nous avons commencé à constituer, basé sur les molécules sémémiques à un ou plusieurs sémèmes;

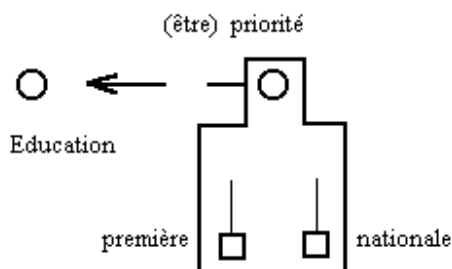
Les molécules sémémiques sont incluses dans les schèmes d'entendement, les schèmes d'entendement se rattachent aux nœuds du modèle thématique par la médiation des structures textuelles.

Application

Le modèle thématique ayant été donné précédemment, nous nous attacherons à la génération de phrases à partir des schèmes d'entendement.

Les étapes à suivre sont les suivantes :

Au premier nœud du modèle est attaché un seul schème d'entendement.



La base est par définition un complexe nominal. Le prédicat en l'occurrence est également constitué d'un complexe substantival.

Le syntactème est de type fN X fN'

A partir de la base et du prédicat, il s'agit de retrouver le fonctème nominal fN et le fonctème nominal fN'.

La structure du fonctème fN a été donnée plus haut.

La base ne comporte pas de relateur.

Le syntagme nominal SN, ne comporte pas de fonctèmes adjectivaux placés en dépendance.

Le déterminant est facultatif. Sans mention d'un quantificateur, il se limite à l'article dont il convient de déterminer le caractère indispensable ou non.

L'absence d'article se rencontre quand l'intention est de ne pas déterminer (Pottier 1974, p.182).

Ici, la voix descriptive appliquée à 'éducation' est incompatible avec l'indétermination.

S'agissant du prédicat, la même démarche s'impose.

Le schème n'indique aucun relateur.

Le syntagme nominal comprend deux fonctèmes adjectivaux correspondants à deux schèmes élémentaires connectés au schème de base.

La présence d'au moins un fonctème adjectival exclut l'indétermination, donc impose l'article.

Il faut déterminer l'ordre des fonctèmes adjectivaux par rapport au fonctème nominal.

"première" est un adjectif numéral ordinal. Ce type d'adjectif s'emploie (Jean Dubois/René Lagane, 1989, p. 72) :

- avant un nom (mais après son déterminant quand celui-ci est exprimé)
- comme attribut
- comme substitut

Ici, sa place est avant le nom. On peut penser qu'il conviendrait de compléter, dans l'analyse sémique, la définition du taxème de "première" en précisant "supératif ordinal".

S'agissant de "nationale", il convient d'appliquer la règle générale qui veut que l'ordre "base+adjectif" donne à l'adjectif une valeur "objective" (dite souvent "sens propre"), tandis que l'ordre inverse "adjectif+base" correspond à une intégration sémantique où l'adjectif apparaît avec une valeur "subjective" (dite "sens figuré").

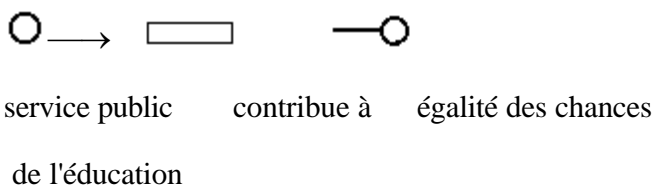
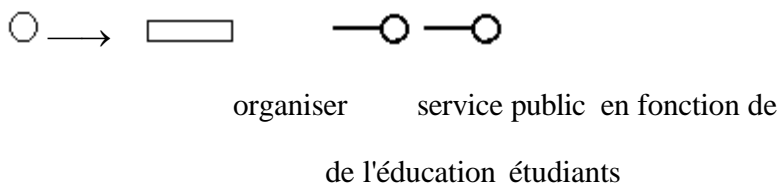
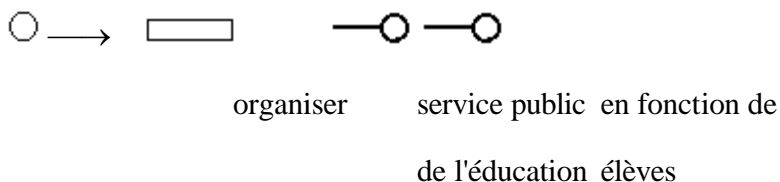
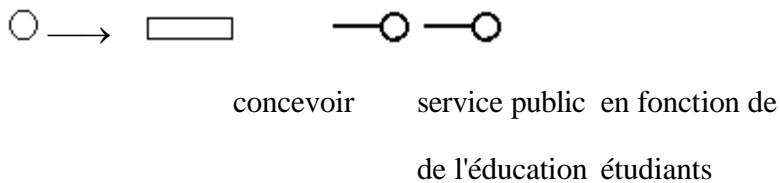
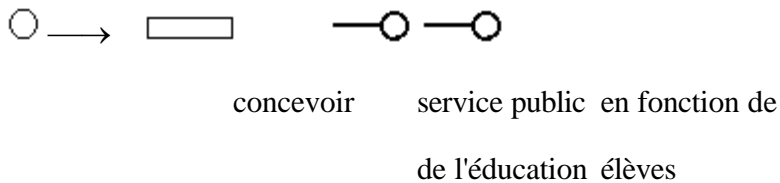
Ici, "nationale" a une valeur objective. On pourrait remplacer l'adjectif par "au plan national". S'il l'on ajoutait l'adjectif "vraie", avec une valeur subjective, on obtiendrait "la première vraie priorité nationale", formulation qui n'a pas réellement sa place dans un texte juridique, tandis que "la première priorité nationale vraie" n'aurait aucun sens dans la mesure où l'on ne voit pas ce que pourrait être une priorité objectivement vraie.

Nous pensons qu'il convient de considérer que par défaut tout adjectif (adjectif, adverbe) a une valeur objective. Ce principe se justifie particulièrement dès lors que l'on a affaire à un texte juridique.

A partir du SE, puis du syntactème correspondant, nous obtenons l'énoncé simple d'origine.

Il apparaît sur cet exemple simple qu'il est possible de parvenir à une génération qui respecte le sens original sans obstacle majeur.

Au second nœud du modèle sont attachés cinq schèmes d'entendement.



Les quatre premiers schèmes ont plusieurs points en commun.

- Leur base est vide (impersonnelle)
- Du point de vue de la vision :
 - l'orientation est inverse, d'où la forme passive
 - la visée porte sur le terme 3 : la lexie "service public de l'éducation"

INV 3 2 4 [1]

ou

INV 3 4 2 [1]

Le syntactème est de type fN X fA

La structure du fonctème fN a été donnée plus haut (page **Erreur ! Signet non défini.**).

La base ne comporte pas de relateur.

Le syntagme nominal SN, ne comporte pas de fonctèmes adjectivaux placés en dépendance.

Le déterminant est facultatif. Sans mention d'un quantificateur, il se limite à l'article dont il convient de déterminer le caractère indispensable ou non.

L'absence d'article se rencontre quand l'intention est de ne pas déterminer (Pottier 1974, p.182).

Ici, la voix descriptive appliquée à 'service public de l'éducation' est incompatible avec l'indétermination. L'emploi de l'article défini est donc obligatoire.

S'agissant du prédicat, la même démarche s'impose.

Le prédicat est fondé sur un fonctème verbal de la forme fV2 (au moins deux actants).

Le schème indique le relateur : "en fonction de ", caractéristique d'une mise en dépendance notionnelle.

Le fonctème verbal est accompagné d'un fonctème nominal introduit par le relateur.

Le fonctème verbal ne comprend aucun fonctème adjectival.

Le fonctème nominal ne comprend non plus aucun fonctème adjectival.

Le fonctème nominal représente la généralité des élèves ou des étudiants et non d'élèves ou d'étudiants en particulier. C'est un trait qui doit être saisi au niveau de l'analyse sémique, et attaché au sémème 'élève' ou 'étudiant' en tant que sème afférent.

Dès lors, l'emploi de l'article "des", pour "de les" s'impose.

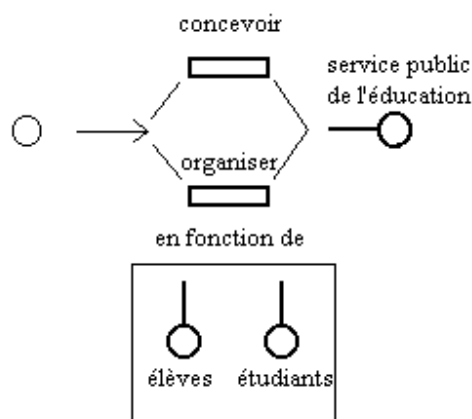
Nous obtenons ainsi les énoncés simples suivants :

"Le service public de l'éducation est conçu en fonction des élèves et des étudiants"

"Le service public de l'éducation est organisé en fonction des élèves et des étudiants"

« Le service public de l'éducation contribue à l'égalité des chances »

Les deux premiers énoncés se prêtent à une factorisation dont le schéma est le suivant :



D'où l'énoncé complexe résultant suivant :

"Le service public de l'éducation est conçu et organisé en fonction des élèves et des étudiants".

Les possibilités de variantes sont les suivantes :

"Le service public de l'éducation est organisé et conçu en fonction des élèves et des étudiants".

"Le service public de l'éducation en fonction des élèves et des étudiants est conçu et organisé".

"Le service public de l'éducation est conçu et organisé en fonction des étudiants et des élèves".

"Le service public de l'éducation en fonction des étudiants et des élèves est conçu et organisé".

"Le service public de l'éducation en fonction des étudiants et des élèves est organisé et conçu".

"Est conçu et organisé en fonction des élèves et des étudiants le service public de l'éducation".

"Est organisé et conçu en fonction des élèves et des étudiants le service public de l'éducation".

"Est conçu et organisé en fonction des étudiants et des élèves le service public de l'éducation".

"Est organisé et conçu en fonction des étudiants et des élèves le service public de l'éducation".

"En fonction des élèves et des étudiants est conçu et organisé le service public de l'éducation".

Le terme "organisé" ne peut pas précéder le terme "conçu". Il ne peut que le suivre. "Organiser" et "concevoir" ont l'un et l'autre un sème spécifique /prospectif/. Toutefois, sur l'axe de la chrono-expérience, l'organisation vient après la conception.

L'ordre de "élèves" et "étudiants" est indifférent, sauf à prendre en compte également la chrono-expérience qui veut que l'on soit élève avant d'être étudiant.

La place du fonctème nominal "en fonction des élèves et des étudiants" est normalement après le fonctème verbal. Le placer avant ou en début de phrase indique la recherche d'un effet littéraire qui n'a pas réellement sa place dans un texte normatif. Même remarque quand le fonctème verbal est placé en début ou en fin de phrase.

Le dernier schème d'entendement est rattaché au même nœud du modèle que les schèmes précédents. Ce nœud, « service public de l'éducation », est présentement la base réelle, alors qu'il n'était que la base apparente, imposée par la visée, dans les schèmes précédents. Sa place en fin de paragraphe ne s'impose pas en logique. Il pourrait tout aussi bien précéder l'énoncé précédent. On peut également le concevoir en dépendance par rapport à la base syntaxique de l'énoncé précédent, ce qui donnerait :

« Le service public de l'éducation, qui contribue à l'égalité des chances, est conçu et organisé en fonction des élèves et des étudiants. »

Cette mise en dépendance aurait pour effet d'affaiblir l'emphase mise sur la notion d'égalité des chances à travers un énoncé propre situé en fin de paragraphe. De plus, le lien logique avec le reste de l'énoncé n'est pas évident.

Cet exemple présente néanmoins un certain intérêt :

- il montre d'abord que l'énoncé, simple ou complexe, se prête à diverses manipulations ayant comme résultat un autre énoncé. Ainsi, avec deux

énoncés simples, on peut faire un énoncé simple ou un énoncé complexe. Avec un énoncé simple et un énoncé complexe, on peut obtenir un énoncé complexe. Avec deux énoncés complexes, on ne peut engendrer qu'un énoncé complexe.

- il montre également que la phrase simple n'est pas l'unité minimale de sens. L'unité minimale est le morphème (Pottier 1974, p.327), indécomposable dans un état synchronique donné (Rastier 1989, p. 279), mais formant un système complexe à l'image d'un atome. Les mots sont construits à partir d'un ou plusieurs morphèmes. Les mots peuvent se regrouper dans des entités complexes mémorisées, les lexies. Les mots et lexies forment des molécules sémémiques et sont en relation les uns avec les autres au sein de structures élémentaires d'entendement, les schèmes d'entendement liés les uns aux autres au sein de schèmes intégrés, schèmes élémentaires et intégrés étant susceptibles d'être actualisés en langue sous forme d'énoncés simples. Les schèmes d'entendement sont capables de s'assembler selon des configurations variées, exprimées par les modules actanciels, pour former d'autres schèmes présentant divers paliers d'intégration, tous les schèmes d'entendement étant reliés les uns aux autres au sein du texte soit directement, soit par l'intermédiaire des divers nœuds du modèle thématique sous-jacent et dont nous avons amorcé la construction. Dans les mécanismes d'intégration et de coordination qui mènent des schèmes d'entendement au texte global, nous découvrirons des régularités, structures conceptuelles et textuelles propres aux textes normatifs, dont la mise en évidence permettra une rationalisation des processus d'analyse et de génération.

Mais avant d'atteindre ces structures conceptuelles et textuelles, quelques compléments concernant les mécanismes sémantiques de formation des énoncés sont tout à fait nécessaires.

Une classification rigoureuse des mécanismes de formation des énoncés

TAL (p. 112-116) apporte une analyse rigoureuse et achevée des mécanismes de transformation des énoncés en fonction de l'intention de l'énonciateur. Il s'agit en réalité d'une grammaire de production.

Le niveau linguistiquement le plus bas est le niveau conceptuel qui s'exprime par un schème conceptuel susceptible d'être illustré par une image, un film ou un dessin. Ce schème conceptuel est nommé par Pottier *schème analytique* (SA).

Le second niveau implique le choix par le locuteur des lexèmes, ce qui représente un premier niveau de contrainte sémantico-syntaxique. Il s'agit du passage de l'événement conceptualisé à son expression en langue naturelle. On obtient ainsi des *schèmes d'entendement* (SE).

Le locuteur doit ensuite choisir dans les lexèmes des schèmes d'entendement ceux qui composeront la base et ceux qui composeront le prédicat. Cette opération consiste à

déterminer la *vision* du schème d'entendement. Le résultat de cette opération est un *schème prédiqué* (SP) qui permet de faire accéder le SE au statut d'énoncé..

En quatrième lieu, le locuteur peut appliquer au schème prédiqué une série d'opérations facultatives qui auront pour résultat un *schème résultatif* (SR) c'est-à-dire un schème complètement achevé du point de vue de l'intention du locuteur. Ces opérations sont notamment les suivantes :

- topicalisation : cette opération consiste à prendre comme topique un élément quelconque du SE, ce qui aboutit à créer une redondance dans l'énoncé relativement à cet élément qui est repris une fois en tant que tel, et une seconde fois sous forme d'un substitut, opération que Maurice Gross analyse comme un « détachement ». Ex. :

SP : les gendarmes ont arrêté les voleurs

SR : les gendarmes, ils ont arrêté les voleurs.

- focalisation : cette opération renforce l'emphase produite par la topicalisation : l'élément détaché est introduit par un présentateur. Ex. : Ce sont les gendarmes qui ont arrêté les voleurs.
- impersonnalisation : la tournure impersonnelle se traduit par la présence d'un présentateur tel que « il y a (que) », « ça fait (que) ». Ex. : « il a été arrêté trois voleurs par les gendarmes ».
- réduction : cette opération consiste à omettre un élément du module. Ex. : « les voleurs ont été arrêtés »
- hiérarchisation : la hiérarchisation consiste à articuler entre eux les énoncés, qui peuvent être simplement juxtaposés (« Pierre traversa le pont. Il aperçut Jean »), coordonnés (« Pierre traversa le pont et aperçut Jean »), ou subordonnés (« Pierre traversait le pont lorsqu'il aperçut Jean »).
- B. Pottier n'a pas développé dans une présentation volontairement concise (opus cit., p. 120) la modalisation qui y a évidemment toute sa place, et dont on peut regretter l'absence dans le schéma présenté, d'autant que B. Pottier lui accorde par ailleurs une place fondamentale dans la structure de l'énoncé (1974, p. 158 à 188, TAL p. 210 à 224). Au demeurant, les opérations appliquées au SP se rattachent à la modalisation dans la mesure où elles manifestent une prise en charge de l'énoncé par le locuteur.

Parmi toutes les opérations facultatives qui viennent d'être rapidement passées en revue, nous considérons que les opérations de topicalisation et de focalisation n'appartiennent pas au langage du droit sauf preuve du contraire. Par contre, la tournure impersonnelle est d'usage courant (« il est créé un établissement... »). Le procédé de la réduction d'actance qui se traduit par la forme passive est aussi d'usage fréquent car il permet de ne pas mentionner un actant trop imprécis. La seconde phrase du premier alinéa de l'article 1 de la loi du 10 juillet 1989

en constitue un exemple parfait. Dans « le service public de l'éducation est conçu et organisé... », comme nous l'avons déjà noté, la réduction d'actance permet d'éviter la platitude consistant à choisir comme base une notion peu significative comme « les autorités compétentes », car l'on se doute bien que seules les autorités compétentes sont en mesure d'organiser le service public et il est inutile de donner à cette notion évidente et de peu d'intérêt un poids sémique important attaché à la position comme base de l'énoncé. Cette opération de réduction est complémentaire de l'opération de choix de la base qui est effectué au niveau de la production du schème prédiqué, et revient à gommer un élément qui n'a aucune importance ou sur lequel l'on n'estime pas nécessaire de retenir l'attention. En conséquence, lorsque nous rencontrerons cette situation, lors de l'analyse sémique, nous prendrons soin de considérer comme base de l'énoncé, celle du schème prédiqué et non celle du schème d'entendement.

Enfin, la hiérarchisation est une opération qui est complètement intégrée dans notre analyse syntaxique où notamment un énoncé qui pourrait être indépendant devient un complexe adjectival par transfert ou un élément marginal au regard du noyau de l'énoncé. On remarquera une similitude entre l'opération consistant à choisir la base du noyau de l'énoncé et celle consistant à choisir dans deux énoncés celui qui sera le noyau et celui qui sera l'élément marginal. Ainsi « Pierre traversait le pont lorsqu'il aperçut Jean » n'est pas sémantiquement identique à « Pierre aperçut Jean alors qu'il traversait le pont ». Dans le premier cas, l'énonciateur met l'accent sur la circonstance que Pierre traversait le pont, alors que dans le second il attire l'attention sur le fait que Pierre aperçut Jean. Mais si l'on dit « Pierre traversait le pont et aperçut Jean », la relation entre les deux composants de l'énoncé complexe est une relation de succession dans le temps sur laquelle l'énonciateur n'a pas prise au niveau de l'interprétation de l'évènement.

En tout état de cause, les trois procédés de hiérarchisation ont pour effet de renforcer la cohérence textuelle, et il conviendra d'en tenir compte.

