

ANALYSE DE DISCOURS ET INFORMATIQUE.

Utilisation du Langage DEREDEC dans des études sur l'argumentation et de raisonnement en langue naturelle.

Alain LECOMTE

AVANT-PROPOS

Cet article doit beaucoup aux avis pertinents de J.M. Marandin et de C. Péquegnat pour la partie linguistique et de P. Plante pour la partie informatique. Nous sommes conscients du fait que deux ensembles de concepts s'y côtoient et souvent entrent en conflit: des concepts linguistiques d'une part (anaphore, coréférence...) et des concepts logico-informatiques d'autre part ("logique naturelle" pour la partie logique et concepts du langage DEREDEC pour la partie informatique). Cet entrelacs de références hétérogènes ne rend pas très facile la lecture de ce texte. Nous pensons néanmoins qu'un effort devrait être continué dans la voie d'une confrontation toujours plus poussée de ces différents langages. C'est à ce prix en effet qu'un sens nouveau peut être donné à la notion de formalisation en linguistique et en Analyse du Discours. Indiquons ainsi que c'est par parti-pris que nous avons adopté en de nombreux endroits la terminologie DEREDEC plutôt que la terminologie classique lorsque celle-ci était peut-être plus adéquate: nous éprouvions le besoin de parler en termes de ce que nous savions pouvoir réaliser en DEREDEC. (Par exemple la fonction "THEME", la relation de "DETERMINATION" ou la notion de "COMPLEMENT" (P1 ou P2)).

1. INTRODUCTION - PROBLEMES DE LA CATEGORISATION D'UN TEXTE EN TERMES D'UNE "LOGIQUE NATURELLE": 1)

Le travail qui est présenté ici s'inscrit dans un projet concernant l'utilisation de DEREDEC dans le domaine de l'argumentation et du raisonnement en langue naturelle. Comme on sait²⁾, le logiciel de Pierre Plante a le très gros avantage d'admettre directement en entrée

des procédures qu'il permet, du texte écrit en langue naturelle. Son fonctionnement, basé sur des reconnaissances de structures (modèles d'exploration) et sur des automates à états finis (possédant quelque parenté avec les ATR de Woods mais autorisant davantage d'opérations) permet de traiter le texte d'entrée de différentes façons: application d'une grammaire (telle que la GDS de P. Plante), introduction (sous forme d'"EXFAL" - expressions admissibles au lexique) d'informations de type sémantique ou textuel³⁾, projection de certaines de ces informations en des endroits appropriés (possibilité de représenter des mécanismes d'inférence), indexation et segmentation. Un avantage important, par ailleurs, est de ne pas faire dépendre l'analyse d'un texte d'une théorie fixée définitivement (grammaire, découpage morpho-syntaxique...) mais de permettre de réaliser des algorithmes qui soient la concrétisation de théories variables supposées adaptées à divers problèmes qu'on se pose.

Ainsi dans l'optique de J.-B. Grize et de ses collaborateurs (trices) du Centre de Recherches sémiologiques de Neuchâtel⁴⁾, peut-on viser à concevoir une batterie d'automates qui déposeraient à la surface d'un texte les traces des opérations dites "de la logique naturelle" qui se réalisent dans l'occurrence des items dont le texte est composé.

Soit par exemple le court texte suivant, extrait d'un article paru dans la Recherche no 106, 1979.

Exemple 1

Si de nombreuses larves de papillons peuvent être utilisées pour leur production de soie, seul le ver à soie est élevé systématiquement. Domesticqué depuis des origines fort anciennes, il n'existe plus à l'état sauvage. Pour se transformer en papillon la larve s'abrite en tissant un cocon formé d'un seul fil ininterrompu dont la longueur peut dépasser 1000 mètres. Déroulé, le cocon donnera la soie grège utilisée dans les filatures. Aujourd'hui, le ver à soie est devenu un instrument d'étude privilégié pour les chercheurs en raison de sa domestication et de ses caractéristiques biologiques.

On souhaiterait parvenir à introduire une catégorisation des unités lexicales usant des symboles α_i , γ , θ et δ , ces symboles pouvant être interprétés de la façon suivante:

α_i : opération d'ancrage (mise en place d'"objets"⁵⁾ autour desquels le texte s'organise); i désigne le rang de l'objet ancré et sert à

coindexer d'autres unités (dans le sens de "tokens") du texte (par exemple les répétitions lexicales, mais aussi les liens de coréférence).

- γ : opération d'ingrédient (enrichissement des objets ancrés). Par γ se trouve introduit un ensemble de traits appartenant à l'objet ancré par α .
- θ : opération de spécification. Le texte précise dans quel contexte est pris l'objet ancré par α (exemples: "pour se transformer en papillon", "dont la longueur dépasse 1000 mètres" - signifie que "le fil est pris sous l'aspect de sa longueur - "Aujourd'hui", etc.).
- δ : opération de détermination. δ sert à fabriquer, à l'aide d'un objet préalablement ancré et d'un prédicat, ce qu'on appelle une détermination: il s'agit en somme d'un noyau d'énoncés à partir duquel les autres énoncés du texte peuvent être mis en relation (de paraphrase, de présupposition, d'opposition, etc.).

Cette catégorisation étant effectuée, on obtiendrait ensuite en extrayant, à l'aide de "modèles d'exploration", les syntagmes affectés de chacun de ces symboles ou d'une concaténation de certains d'entre eux, des représentations de ce que sont les objets de discours pris dans ce texte, sous forme de classes-objets structurées:

1. α_1 : <le ver à soie > δ : <est élevé systématiquement>
 γ : <larves de papillon,>
 <production de soie >
 <domestiqué depuis des origines fort anciennes>
 <ne...plus état sauvage>
2. α_2 : <l'état sauvage>
 γ : <ne...plus>
 <domestiqué depuis des origines fort anciennes>
 <il (le ver à soie)>
3. $\theta\alpha_1$: <la larve > δ : <. s'abrite>
 γ : <un cocon>
 $\gamma\alpha_1$: <un seul fil ininterrompu>
 $\theta\alpha_1$: <longueur>
 $\gamma\theta\alpha_1$: <1000 mètres>

4. α_3 : <le cocon> δ : <déroulé, donnera>
 γ : <la soie grège>
 $\gamma\gamma$: <les filatures>

5. $\theta\alpha_1$: <le ver à soie> δ : <est devenu>
 γ : <un instrument d'étude>
 $\gamma\gamma$: <les chercheurs>
 <sa domestication>
 <ses caractéristiques biologiques>

Cette liste (ou liste de listes) pourrait alors se lire ainsi: le premier objet ancré est <le ver à soie>, il est composé d'un certain nombre d'ingrédients ou "aspects de l'objet" (glosables en "être une larve de papillon", "être producteur de soie", "être domestiqué...", "ne plus être à l'état sauvage"... autant de traits pouvant entrer dans une sorte de définition contextuelle de l'objet envisagé) et il supporte une opération de détermination: "est élevé systématiquement" (il s'agit de l'information considérée comme pertinente à cet endroit du texte, le reste, à savoir les ingrédients, participant seulement de la définition de l'objet sur lequel une information est apportée).

Un des aspects peut être considéré comme un objet de discours qui est ancré à son tour: <l'état sauvage>. Il possède alors lui-même comme ingrédients des aspects appartenant déjà au premier objet, ainsi qu'une reprise anaphorique de cet objet. Ceci marque une bipolarité du texte à cet endroit, lequel peut être vu comme parlant de <ver à soie> ou de l'objet <état sauvage/domestiqué>. (Interprétation: il est concevable qu'à ce moment du texte, celui-ci enchaîne soit sur l'un, soit sur l'autre de ces deux objets... soit sur les deux). $\theta\alpha_1$ indique qu'une spécification est introduite par-dessus un ancrage. La spécification sera ici: "pour se transformer en papillon", dans ce contexte, l'objet <ver à soie> devient l'objet <larve> ou encore "la larve" doit être conçu comme reprise anaphorique de "le ver à soie". Toujours dans ce contexte, l'objet ainsi spécifié reçoit une détermination: "s'abrite". La liste qui suit donne les ingrédients, les ingrédients des ingrédients et des spécifications de ces ingrédients. Ainsi " $\theta\gamma\gamma$: <longueur>" signifie que l'ingrédient "un seul fil ininterrompu" est vu, à cet endroit du texte sous l'aspect de sa "longueur", laquelle recevra alors un ingrédient particulier ("1000 mètres"). Et ainsi de sui-

te. Si on tentait de parcourir la classe-objet structurée associée à *<ver à soie>* on obtiendrait ainsi une suite d'enrichissements, de déterminations et de spécifications traduisant autant d'étages dans un fil argumentatif, autant de transformations "cognitives" d'un thème de discours pour parvenir à un état conclusif.

Bien sûr, un tel repérage est encore très intuitif et on nous objectera la difficulté d'isoler quelques critères formels permettant de distinguer un γ d'un α voire un γ d'un θ . En fait, il s'agira, dans ce qui suit, d'éclaircir quelques mécanismes de reconnaissance auxiliaires dans la perspective de production de telles analyses.

2. TOPOLOGIE DES SEQUENCES

2.1 Interprétation et topologie

Nous partirons de l'hypothèse que la forme "raisonnement" d'un texte (comme d'ailleurs la forme "récit", ou la forme "description", ou la forme "biographie", etc.) n'est pas directement inscrite dans le texte à la façon dont sont inscrits les mots et leurs rapports grammaticaux. Une telle forme n'est pas donnée à voir comme l'est par exemple le phénomène d'accord en genre et en nombre: ceci signifie que son repérage ne ressort pas d'une quelconque "grammaire de texte" (au sens donné à ce terme par exemple par Van Dijk). Il y a raisonnement parce qu'il y a lecture et interprétation. Pour reprendre les expressions du sémioticien Par Aage Brandt (1973), nous dirons avec lui que "tout discours inclut à titre d'actants constitutifs un DIRE qui est du côté du destinataire et un VOIR qui est du côté du destinataire". Une "forme" (raisonnement, récit, description...) se détache donc à partir d'un rapport entre ces actants qu'on peut dire "rapport d'interprétation". Pour Brandt, l'énoncé est un transfert d'un DIRE dans un VOIR, ce qui implique pour nous la possibilité d'un modèle topologique de l'interprétation: ce qui est dit occurrent alors comme un ensemble de signaux déclenchant pour le récepteur des opérations d'ouverture, de fermeture, de compactification et de mise en relation dans des espaces. Dans ces opérations d'interprétation, il se produit un effet de séquentialité qui est alors identifié à un

raisonnement ou à un récit ou à autre chose encore.

Il va de soi cependant que les rapports topologiques qui se tissent à partir des signaux fournis par les mots, d'une part, ne sont pas déterminés avec unicité (plusieurs ensembles sont possibles) et d'autre part ne se tissent pas à partir de rien: il y a toujours déjà un espace informé. Comme le dit Pêcheux (1982: 35, 45), "lire c'est toujours réinscrire un énoncé dans l'espace d'une problématique". Cet espace-réceptacle où se projette un texte pour qu'il fasse sens, cela pourra être aussi bien celui d'un inconscient individuel ("poinçonné" d'un fantasme particulier (cf. Lacan)) que celui d'une idéologie. Dans un texte d'historien sur les placements d'enfants au XVIIIe siècle (LEROY-LADURIE, 1979: 15-21) la commutation quasi automatique de <mise en nourrice> et de <allaitement mercenaire> en dira long par exemple sur les catégories d'interprétation requises par un texte qui se voudrait tout d'"objectivité scientifique".

Dans un cas pareil, c'est une similitude de place dans des formes syntaxiques récurrentes qui fait apparaître comme "naturelle" la commutation des deux expressions. Ou bien encore, ce sera un enchaînement particulier de places dans des formes syntaxiques qui provoquera un effet semblable, effet qu'on peut qualifier de coréférentiel. Ainsi dans notre exemple traitant du ver à soie, nous avons une "forme" qui consiste en: un SN ("le ver à soie") dont on a supposé qu'il portait la marque d'un ancrage, suivi d'une détermination ("est élevé systématiquement"). Cette forme est suivie d'une autre forme comportant d'abord une spécification ("pour se transformer en papillon") puis un SN du même type que le précédent ("la larve") Puis enfin une détermination ("s'abrite"). On pourra alors supposer, par exemple, que l'enchaînement de ces deux formes crée pour les SN en position similaire une relation de coréférence⁶⁾.

Ce genre de schéma caractérisera alors ce qu'on appellera: une règle d'interprétation textuelle.

La reconnaissance de tels liens de coréférence est fondamentale pour l'analyse des discours, en tant que celle-ci se propose justement de repérer ces points singuliers où un signifiant glisse vers un autre (soit sous l'effet d'une métonymie: <mise en nourrice>→<allaitement> ou <ver à soie>→<larve>, soit sous celui d'une condensation (métaphore)

<nourrice>→<mercenaire>) et qui sont autant de points autour desquels s'organisent les textes (et les polémiques). Pour y parvenir, il nous faudra d'abord définir les unités de segmentation pertinentes du texte, à l'intérieur desquelles s'organiseront les expressions lexicales, dans un réseau de places.

2.2 L'espace de la séquence

Nous définissons d'abord un segment comme une chaîne de mots, de longueur arbitraire contenant une expression nominale en position de thème. La notion de thème est reprise ici aux travaux de Halliday notamment. Est thématiqué ce qui est "mis en relief" du point de vue informationnel, avant le groupe verbal, voire même comme constituant une unité autonome, s'il n'y a pas de groupe verbal. (Exemples: "Pierre est venu hier soir"; "Hier soir, Pierre est venu"; "Oh! catastrophe!"...) sous sa forme actuelle, DEREDEC permet de définir comme thèmes tous les GN ("Groupes nominaux") occurrents avant un GV ("Groupe verbal") dans un GP ("Groupe propositionnel"). Cette définition pourrait être raffinée, de façon à obtenir par exemple comme thèmes, des GN figurant dans un complément circonstanciel antéposé, ou bien des GN-sujets ayant été déplacés. Une expression nominale en position thématique possède une expansion sous forme d'une suite d'"aspects" (les γ^j d'une série α dans l'exemple que nous avons donné plus haut); ceux-ci sont formés de syntagmes nominaux munis de déterminants (indéfinis, possessifs, démonstratifs...) instaurant avec l'expression nominale thème (E.N.T.) des relations anaphoriques ou coréférentielles.

Repérer des segments est ainsi un préalable: à l'étape actuelle de notre recherche, nous nous centrons sur la fonction nominale. On peut raisonnablement penser en effet que celle-ci s'exerce en des lieux de stabilité référentielle. (On retrouve ici la notion travaillée par Thom de stabilité structurelle d'une forme, une des conditions de cette stabilité étant justement la nomination). Le discours produit un réseau de relations entre des unités (des "noeuds") stabilisées. D'où notre démarche: définir les segments (et au-delà: les séquences) à partir des liens anaphoriques et coréférentiels entre expressions nominales et entre expressions nominales et pronominales. Le réseau formé par ces liens constitue l'espace de la séquence, son système de coordonnées.

Nous définissons ensuite une séquence comme un arrangement de segments soit sous forme emboîtée, soit sous forme juxtaposée. Une séquence fait donc l'unité d'un faisceau d'expressions nominales et va l'orienter dans le cas d'un raisonnement, vers un point de convergence, qui sera une sorte de thème d'un niveau supérieur aux thèmes des segments.

2.3 Construction de l'espace d'une séquence: anaphore et coréférence

Avant de parvenir à un calcul des segments et des séquences, il nous faut préciser la nature de ces liens qui les constituent. Ces liens s'expriment en termes de coréférence, d'anaphore et d'anaphore associative, toutes relations présentes dans les phénomènes dits de reprise immédiate qui ont donné lieu déjà à une littérature abondante (cf. en particulier F. Corblin (1983), J.C. Milner (1976), C. Blanché-Benveniste et A. Chervel (1966)).

Nous tenterons dans ce qui suit de caractériser ces relations à partir d'une approche utilisant notre notion de découpage en segments et séquences et de voir en quoi elles sont connectées avec les notions introduites précédemment (ancrage, détermination, spécification, ingrédient...). Afin de respecter la terminologie usuelle au moins approximativement (car les définitions de ces relations sont très variables selon les auteurs), nous caractériserons la coréférence comme une relation symétrique entre des unités du texte.

(Cf. Milner (1976):

Il y a relation de coréférence entre deux unités référentielles A et B quand elles se trouvent avoir la même référence - ce qui peut arriver sans que l'interprétation de l'une soit affectée par l'interprétation de l'autre. (p. 32).

Milner donne l'exemple: "Napoléon/l'Empereur" où la paire est "constituée par des termes dont la référence actuelle est identifiée indépendamment" (p. 34)).

Nous caractériserons l'anaphore, au contraire, comme une relation asymétrique. Pour elle, en effet, l'interprétation d'un terme est dépendante de celle d'un autre terme, ainsi qu'il arrive dans le cas de l'anaphore pronominale, bien sûr, (un syntagme nominal est repris par un

pronom) mais aussi dans le cas où le deuxième SN d'une paire, étant par lui-même relativement indéfini, reçoit son interprétation d'un SN de tête qui, lui, est davantage défini.

Exemple (2)

Un ver à soie se transforme en papillon. La larve s'abrite alors dans un cocon.

Mais comme il a été déjà souvent remarqué (cf. Hawkins (1978), Corblin (1983)) des cas de reprise ont lieu sans qu'il y ait anaphore (au sens "strict") ni coréférence, par exemple dans:

Exemple (3)

L'avion décolle... Le palonnier vibrait...

où, là aussi, pourtant, l'interprétation du deuxième SN nécessite la connaissance du premier. On parle, en ce cas, suivant Hawkins, d'anaphore associative.

Parmi ces trois relations, un auteur comme Corblin ne considère comme autorisant la reprise immédiate que les deux espèces d'anaphore, à l'exclusion de la coréférence. Il caractérise en effet les phénomènes de reprise par deux propriétés:

- A - Il s'agit de rapports qui s'instaurent en vertu d'une succession: l'interprétation du SN défini exige la prise en compte d'un déjà-là du texte.
- B - La notion de phrase n'est pas pertinente pour repérer le domaine à l'intérieur duquel l'autre terme du rapport peut être trouvé. (p. 119).

Or ces deux propriétés sont absentes de la coréférence telle qu'elle a été définie plus haut.

Nous retenons également du travail de F. Corblin l'idée selon laquelle

l'anaphore consiste non à répéter quelque donnée du texte antérieur, mais à situer nécessairement l'interprétation de tout nouveau SN d'un texte relativement à la mémoire des interprétations antérieures, à doter nécessairement chaque SN d'un rapport explicite aux interprétations déjà produites. (p. 120).

La notion de mémoire paraît en effet centrale dans l'analyse de ces phénomènes. Il s'agit d'une mémoire à court terme et propre à un texte qui se constitue lorsque se structurent ce que nous avons appelé les objets de discours par l'effet d'une sériation d'expressions nominales.

Il apparaît toutefois quelques difficultés lorsque nous voulons distinguer anaphore et coréférence. A ce propos, les notions de "plus ou moins défini" ou de "plus ou moins identifié" qui interviennent entre autres dans la caractérisation faite par Milner de ces relations, semblent garder un certain flou. Si, usuellement, dans

Exemple (4)

Le ver à soie se transforme en papillon. La larve s'abrite dans un cocon,

le deuxième SN paraît anaphoriser le premier, on peut toutefois tout aussi bien accepter une lecture différente d'une telle séquence, si on considère par exemple que suivant le critère proposé par Milner, "la larve" est un terme dont la référence actuelle est identifiée indépendamment de celle du syntagme "le ver à soie". Le cas serait encore plus frappant dans une suite du genre :

Exemple (5)

Bombyx muri est élevé pour produire de la soie. Le ver à soie...

Supposons en effet qu'il y ait plusieurs espèces de ver à soie et que Bombyx muri en soit une, l'interprétation anaphorique ou seulement coréférentielle de cette séquence dépendrait alors du fait qu'on possède ou qu'on ne possède pas cette connaissance.

Dans ce cas en effet, "le ver à soie" apparaît comme un indéfini relativement à Bombyx muri (cas d'anaphore) dans l'autre au contraire, ils sont aussi définis l'un que l'autre et leur relation est symétrique (cas de coréférence sans anaphore).

De fait, il apparaît assez nettement lorsqu'on considère ces phénomènes de reprise caractérisés par la non-pertinence de la notion de phrase, qu'ici encore plus qu'en tout autre domaine linguistique d'investigation, on ne saurait se contenter de raisonner sur des exemples tirés hors de tout contexte, et qu'il est nécessaire de construire ces notions transphrastiques que nous baptiserons ici "segment" et "séquence".

Si, à ce propos, les concepts précédemment introduit (p. et provenant des essais de logique naturelle tentés par J.-B. Grize et al. ont quelque pertinence, alors on peut les référer à la dimension de la séquence d'une part et on peut, d'autre part, tenter de les relier à ces phénomènes d'interprétation que sont l'anaphore et la coréférence. Dans le contexte d'une ou plusieurs séquences, en effet, un exemple tel

que (4) perd de son ambiguïté: on pourra en effet "calculer" que "*le ver à soie*" est le principal thème du discours et que le syntagme "*la larve*" occurrent en position thématique dans une phrase qui succède à l'introduction de "*le ver à soie*", ne pourra qu'en être une reprise au sens où on devra l'interpréter comme "*la larve consistant en le ver à soie*". Dans le court texte que nous avons choisi comme exemple, cette interprétation sera rendue encore plus évidente par la "mémoire" d'une séquence antérieure où "le ver à soie" est situé parmi "*de nombreuses larves de papillons*". Notons alors combien cet exemple illustre la thèse de Corblin sur les "domaines lexicaux structurés": nous avons bien un défini ("*le ver à soie*") dont le rôle est d'"opérer sur l'ensemble des domaines structurés par le texte depuis le début de celui-ci" (à savoir dans notre cas le domaine lexical constitué autour de "*larves de papillons*"), ce "travail (se faisant) à partir du contenu lexical du SN défini" (p. 121). Ceci étant, le défini exerce en effet son rôle sur le domaine précédemment introduit, mais ici, c'est-à-dire lorsque c'est la première fois où un défini "occure" dans le texte -ce rôle est de spécifier ce domaine sous l'aspect d'une introduction de ce dont le texte va parler. C'est là ce que nous avons appelé une opération d'ancrage. Il doit être possible de soutenir -bien que cela sorte des limites de cet article- que tout ancrage suppose du déjà-là textuel.

Autrement dit, on n'ancre un objet qu'à partir d'un domaine déjà introduit même s'il est relativement indéfini (ce qui se repère au système des déterminants, par exemple: "*de nombreux...*", "*quelques...*", "*certains...*"). Si cette introduction préalable n'a pas été explicite, l'effet de l'ancrage sera de créer rétrospectivement cette idée de déjà-là, en donnant une impression de raccord à d'autres textes antérieurs.

Lorsqu'un ancrage a eu lieu, l'unité lexicale supportant cet ancrage admet nécessairement une expansion: celle-ci consiste soit en des reprises anaphoriques ("*il*", "*la larve*"...) soit en des termes reliés par une relation associative, celle-ci étant alors entendue en un sens plus général que l'anaphore associative, (on y admettra des SN déterminés par un possessif ou des SN indéfinis figurant éventuellement en position de complément d'objet, par exemple: "*le ver à soie*" → "*la larve*" → "*un cocon*" constitue une expansion).

L'expansion se termine lorsque apparaît un nouvel ancrage.

3. CALCUL SUR LES SEGMENTS ET LES SEQUENCES

Nous proposons ici de jeter quelques bases de ce calcul. Il doit évidemment partir de quelques principes (même si ceux-ci sont destinés à être modifiés en cours de route) et des marques matérielles que le texte comporte. De ce point de vue, il est inévitable d'accorder l'attention aux déterminants nominaux. Pour simplifier, nous classifions les expressions nominales en quatre groupes:

- l-expressions: celles dont le déterminant est un défini ("le", "la", "les") leur occurrence indique qu'un objet est identifié "actuellement" (c'est-à-dire par la l-expression elle-même: cas d'ancrage) ou bien a été identifié antérieurement (cas de reprise).

- u-expressions: celles dont le déterminant est un indéfini ("un", "une", "des") leurs occurrences se répartissent en plusieurs cas qui ont trait à la distinction générique/situation. Exemples (non exhaustifs):

1. "un chat, ça griffe" (générique)

2. "un chat est entré hier dans ma cave" (situation)

2'. "un chat est entré hier dans ma cave. L'animal y a renversé toutes mes bouteilles" (situation).

1 et 2' comportent une identification, mais elle est conférée à posteriori par les reprises anaphoriques ("ça", "l'animal") 2 n'en comporte pas. En anticipant quelque peu sur la suite, nous pouvons dire que "un chat", dans ces exemples, reçoit le statut d'"expression nominale thématique" en 1 et en 2', mais non en 2 (où le "thème" sera davantage la situation en elle-même, plutôt que l'objet désigné⁷).

- c-expressions: celles dont le déterminant est un démonstratif ("ce", "cet", "cette", "ces"). Ce sont des expressions qui n'ont pas l'autonomie référentielle des l- et des u-expressions, puisque leur référence leur est toujours conférée par un segment antérieur. Le problème est alors de déterminer le champ de cette référence: s'applique-t-elle à une l- ou une u-expression seulement (cf.: "un chat...cet animal...") ou bien à un segment entier, voire à une séquence (cf.: "la lumière est blanche. Quand elle traverse un prisme, elle montre toutes les couleurs de l'arc-en-ciel. Ce phénomène a été étudié par Newton").

- p-expressions: celles dont le déterminant est un possessif. Nous parlerons de lien de coréférence associative pour désigner le lien exis-

tant entre une telle expression nominale qu'elle reprend sous l'aspect d'une unité dépendante.

Comme pour les u-expressions, les c- et les p-expressions reçoivent une interprétation thématique lorsqu'elles sont reprises dans la suite du texte soit par une l-expression (ex.: "...un cocon. Le cocon...") soit par un anaphorique (ex.: "...J'ai souvent rencontré cette femme. Elle m'a toujours paru bizarre" ou bien: "Paul a mangé son chapeau. Il était dur à digérer".)

Détermination des points-origines

Nous appellerons désormais Expression Nominale Générique (E.N.G.) une expression nominale supportant un ancrage et figurant donc comme tête de liste d'une expansion. Les liens de coréférence ou d'anaphore entre les expressions nominales quelconques et une E.N.G. se détermineront, en même temps que celle-ci, par un principe de hiérarchie.

Règle 1 - Un nom propre en position thème (ou complément si le thème n'est pas une expression nominale) est une E.N.G.

Règle 2 - Une l-expression en position thème (ou complément si le thème n'est pas une E.N.G. déterminée par une règle antérieure) munie d'au moins une autre détermination nominale que l'article est une E.N.G. si elle ne suit pas immédiatement une E.N.G. déterminée par la règle 1.

(Commentaire: il s'agit de ce que la littérature logique appelle "les descriptions définies". La clause restrictive a pour but d'identifier la description définie avec le nom propre qu'elle suit immédiatement, comme dans l'exemple: "A New York, j'ai rencontré Chomsky. Le célèbre linguiste américain faisait une conférence").

Règle 3 - Une u- ou une p-expression en position thème (ou complément si le thème n'est pas une expression nominale) est une E.N.G. si elle est immédiatement suivie par une c- ou une p-expression elle-même en position thème ou par un pronom-sujet anaphorique "il", "elle"... ou par un pronom relatif sujet.

(Exemple: "...des veuves prennent en charge un nouveau-né, elles ne peuvent l'allaiter...")

ou bien: "...des veuves prennent en charge un nouveau-né. Ces femmes..."

ou bien: "... nous avons posé la question à des ouvriers. Leurs témoignages..."

Règle 4 - Une l-expression en position thème ou contenue dans une expression nominale en position-thème ou en position complément s'il n'y a pas d'expression nominale en position thème est une E.N.G. si on peut construire à partir d'elle une expansion propre⁸⁾ non vide ne contenant pas d'E.N.G. du type des règles 1, 2 ou 3, soit vers l'avant du texte, soit vers l'après, soit dans les deux directions, soit dans l'espace restreint de la relative où elle figure comme sujet.

(exemples: "La lumière du soleil est blanche. Après avoir traversé un prisme, elle montre toutes les couleurs qui existent dans le monde visible. La nature reproduit le même résultat dans la belle gamme des couleurs de l'arc-en-ciel".

Soit l'expression: "la nature". Son expansion, s'il s'agissait d'une E.N.G. contiendrait:

soit (vers l'avant): "toutes les couleurs qui existent dans le monde visible", il s'agit-là d'une E.N.G. d'après la règle 2;

soit (vers l'après): "le même résultat" et "la belle gamme des couleurs de l'arc-en-ciel" qui sont des E.N.G. d'après la règle 2⁹⁾ cette expansion serait donc vide. D'où il résulte que "la nature" n'est pas une E.N.G. (donc le sera l'expression nominale "la belle gamme des couleurs de l'arc-en-ciel") (le cas de "le même résultat" devrait faire l'objet d'un traitement particulier en tant que constituant une unité fondant ce qu'ultérieurement nous caractériserons comme "thème transéquentiel", c'est-à-dire thème qui traverse deux séquences au moins. Si nous avions eu "la nature reproduit ce résultat...", il en eût été de même puisque alors "ce résultat" aurait appartenu à une autre expansion, déjà constituée).

"...un bord de l'arc-en-ciel est toujours rouge et l'autre toujours violet; entre eux se trouvent rangées toutes les autres couleurs. L'explication que Newton donne de ce phénomène est celle-ci..."

Ici "l'arc-en-ciel" est contenu dans une expression nominale thème ("un bord de l'arc"), on peut construire à partir d'elle une expansion propre qui contient "rouge", "l'autre", "violet", mais aussi: "un bord". C'est donc une E.N.G.).

