HAL
open science

COURS DE LINGUISTIQUE MODULAIRE

Frédéric Torterat

To cite this version:

Frédéric Torterat. COURS DE LINGUISTIQUE MODULAIRE : Présentation de quelques approches en linguistique formelle et linguistique textuelle contemporaines. 2005. halshs-00067979v1

HAL Id: halshs-00067979

https://shs.hal.science/halshs-00067979v1

Preprint submitted on 9 May 2006 (v1), last revised 30 Sep 2008 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Frédéric TORTERAT
UNIVERSITE DE PARIS-SORBONNE (EA 3560)
UNIVERSITE D'ETAT D'HAÏTI
(FACULTE DE LINGUISTIQUE APPLIQUEE,
AGENCE UNIVERSITAIRE DE LA FRANCOPHONIE)

SÉANCES DE SÉMINAIRE 2005-2006
(PREMIÈRE MISSION)

LINGUISTIQUE MODULAIRE

(Frédéric TORTERAT)

SOMMAIRE :

SEANCE 1 : VERS UNE LINGUISTIQUE GENERALE ?
SEANCE 2 : CATEGORISATIONS ET CLASSIFICATIONS ;
SEANCE 3 : EN TERMES D'APPROCHES MODULAIRES ;
SEANCE 4 : QUELQUES PROBLEMATIQUES TRANSVERSALES ;
SEANCE 5 : LA PROFUSION DES METHODOLOGIES DESCRIPTIVES ;
SEANCE 6 : LES METHODOLOGIES EXPLICATIVES ET LEUR FORMULATION DIDACTIQUE.

SEANCE 1 :
VERS UNE LINGUISTIQUE GENERALE ?[1]

La linguistique, d'un point de vue épistémologique, renvoie certainement plus à un ensemble de terminologies descriptives, de démarches explicatives, de classifications typologiques, qu'à une science à proprement parler. Disons plutôt que si elle reporte surtout aux premières, avant de reporter à l'autre, cela tient notamment au fait que le linguiste s'impose d'abord comme un méthodologue, un terminologue, un philologue, mais que ses recherches intègrent des domaines comme la sociologie, la psychanalyse par exemple, sans qu'elles en soient extraites ou forcément distinctes. Cela étant, le fait d'envisager éventuellement la linguistique à travers ses modélisations n'empêche pas de lui reconnaître de véritables champs disciplinaires, de l'inscrire dans une historiographie, et moins encore de lui assigner une dimension philosophique, praxéologique ou autre. D'où la persistance de compromis comme les linguistiques dites de « trait d'union » (ainsi la sociolinguistique), ou celles qu'on appelle quelquefois « restreintes » (ainsi la linguistique textuelle, la linguistique sociale). Les années 1960-1990 ont d'ailleurs donné dans une forme de profusion de champs disciplinaires bien compréhensible : ce moment de l'histoire des sciences et techniques, combiné à l'accélération des innovations technologiques, a été aussi celui des rapprochements entre des équipes de recherche, des disciplines et des professions, et l'on pourrait en dire autant, en termes de « traits d'union », de l'électromécanique ou de l'aromathérapie. Il n'est pas hors de propos de noter, dans cet esprit, que l'on soit passé dans les années 2000 d'une question d'interculturalité à une problématique de coculturalité dans des domaines variés : en termes philosophiques et psychanalytiques, l'*autre*, c'est *ce qui n'est pas moi*, en 1990 comme en 2000. Or, là où les années 1990 se préoccupent de savoir ce qui se passe *entre nous*, les années 2000, elles, cherchent à savoir ce qui se passe *quand l'un avec l'autre,* ou plutôt *quand l'un va vers l'autre* : les faits sont analogues, mais les approches, de leur côté, ne sont plus vraiment similaires.

L'un des autres enjeux principaux de la linguistique contemporaine est facile à discerner, pour peu qu'on l'envisage dans une vue propédeutique et prospective : on assiste peu à peu à la mise à l'écart du *distingo* des sciences exactes et des sciences dites « humaines ». D'un côté, les premières reviennent à un certain souci de concrétude et un certain culturalisme, là où les autres, simultanément, se tournent vers plus d'exactitude et d'empirisme. Pour ne donner qu'un exemple anecdotique, le fait qu'un bon journaliste aujourd'hui soit aussi médiologue nous paraît s'imposer, tout comme on imagine peu qu'un économètre ne soit pas, aussi, un économiste, et pour poursuivre dans la rubrique des faits de société, il est certains cas aujourd'hui où tout le monde déclarerait légitime que le juriste collabore avec le psychiatre, l'architecte avec l'urbaniste, le politicien avec l'historien.

Pour ce qui nous concerne ici plus particulièrement, la linguistique des années 2000 et suivantes s'inscrit très nettement dans des problématiques à la fois actionnelles et opérationnelles, mais aussi dans un courant de transversalité, lui-même complété par un élan de modularité[2]. Ainsi, la production de sens implique-t-elle une action, autrement dit des emplois effectués en contexte, et la question se pose désormais de savoir à quelle(s) opération(s) linguistique(s) renvoie telle ou telle activité de discours. En outre, les approches que les linguistes peuvent mettre en place de la phrase, du paragraphe, du texte, sont transversales dans ce sens où elles sollicitent plusieurs méthodologies descriptives, plusieurs appareils taxinomiques, et, sans fabriquer pour autant une généralisation hasardeuse, nous admettrons ensemble que tout indique, dans ces manières de pratiquer la linguistique, un

[1] Qu'on me permette dès le début de témoigner toute ma reconnaissance envers mes collègues de l'Unité de Recherche Pradel Pompilus (URPP), à savoir Délano Frantz Alexandre, Moslyn Constant, Jean Marcel Georges, Wesner Mérant, Wilhem Michel, Léonard Etienne, Stéphane Pierre-Paul, Serge Sylvestre, Louinès Volny et Lemète Zéphyr. Qu'ils soient remerciés ici d'avoir bien voulu corriger mon créole haïtien hésitant et quelquefois erroné, et d'avoir contribué au bon déroulement de cette mission d'enseignement par leurs questions et leurs suggestions.

[2] Il est recommandé de consulter sur le web le rapport EAGLES, paru en 1996, sur les formalismes multimodulaires (url : http://www.ilc.pi.cnr.it/EAGLES/home.html).

véritable souci de productivité phénoménologique. De plus, les conclusions se veulent intermédiaires, les déductions sont présumées soumises à discussion, mais au moins personne ne part de rien pour tout remettre à plat. Un tel comportement d'ailleurs, aujourd'hui, paraît des plus risibles et ferait *illico presto* l'objet d'une marginalisation. Dans cette vue, la modularité permet de combiner plusieurs composantes d'un (ou de plusieurs) champ(s) disciplinaire(s) (comme la phonologie et la syntaxe), pour aborder un même fait pour ainsi dire dans un ensemble unifié (d'où les grammaires d'unification, dont nous reparlerons, ainsi que de la présence affirmée des interfaces).

En termes de suivi de la recherche, on assiste par là-même à ce que nous appellerions, pour faire simple, des regroupements méthodologiques. Ici une approche appartient plutôt à la sémiotique des valeurs, là où une autre reprend la problématique des opérations énonciatives, et où telle autre encore sollicite le matériel représentatif des grammaires syntagmatiques. Ainsi, tel phénomène cognitif est abordé tantôt d'un point de vue strictement grammaticographique, tantôt à l'appui de certaines modélisations schématiques propres au traitement automatique. Or des liens sont toujours possibles entre ces regroupements, l'enjeu consistant de plus en plus à distinguer cette fois-ci ce qui est de l'ordre de l'indéformable et au contraire du déformable, de l'invariance et, à l'inverse, de la variabilité.

Les grammaires syntagmatiques, comme on s'en doute, placent le syntagme au premier plan, lequel syntagme rassemble des constituants en général autour d'une unité LEX avec laquelle ils entretiennent des relations variées (comme celle de dépendance). En termes chomskyens, la grammaire peut ainsi répondre à des *principes* fermement établis, comme l'existence de catégories grammaticales ou des contraintes qui pèsent sur les relations entre les constituants (Chomsky 1995, Chomsky et Lasnik 1993, Sportiche 1997). C'est ensuite au niveau local que différents *paramètres* peuvent être inventoriés, lesquels ressortissent au domaine de la déformabilité, de la variabilité et de la spécificité. En termes d'approches modulaires en grammaire générative, celle-ci comporte des composantes qui appartiennent, sur un premier plan, aux formes dérivationnelles et flexionnelles notamment, mais surtout, sur un plan segmental, aux formes *P* (*phonologique* : chaînes linéaires de sons perceptibles, *grosso modo*) et *L* (*logique* : relations entre constituants et déplacements).

Si on pose la phrase *kijan ou rele ?*, on schématisera le régime du verbe *rele* par une variable *x*, qui n'apparaît pas dans la chaîne linéaire du syntagme verbal :

1. Kijan ou rele *x* ?

Or, c'est bien *ou* ici qui correspond au régime verbal, et le fait qu'il s'inscrive dans une tournure interrogative provoque son déplacement (qui reporte au principe du DEPLACER α, bientôt remplacé par le principe *MOVE*) :

1'. Kijan ou rele *x* ?

L'emplacement du régime étant vide, la grammaire générative a d'abord indiqué ce fait à l'appui de la notation *e* (pour *empty*, cat. vide), soit :

2. Kijan ou rele *e* ?

mais cette terminologie a bientôt été délaissée pour son inappropriété à dégager les liens qui s'établissent entre les unités (ce qui sera ensuite le cas de la notation PRO) : la position n'est pas vide, à proprement parler, mais elle est occupée par une unité qui s'est déplacée à l'occasion d'un mouvement (l'ang. occupe toutes les positions avec *what is your name ?*, et le fr. apparaît pléonastique avec *comment vous appelez-vous ?*).

Le module dit de la chaîne des traces permet de représenter les mouvements opérés dans le syntagme, comme suit :

3. Kijan ou_i rele *i* ?

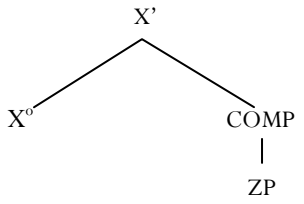
On reconnaîtra facilement une relation de sous-jacence entre le pronom prédiqué (ou_i) et sa trace (*i*) dans le syntagme verbal. Cette contrainte locale démontre le lien qui existe entre la position qu'occupe matériellement l'unité déplacée et celle(s) qu'elle occupe effectivement au niveau L. Dans la phrase

qui dis-tu qu'on a bousculé ?, cette relation de sous-jacence sera formalisée ainsi :

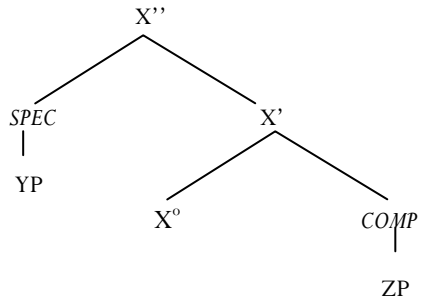
Qui$_i$ dis-tu qu'$_i$ on a bousculé i ?

L'unité *x* régime du verbe *bousculer* se concrétise dans deux complémenteurs (*qui* et *que*) et, même si elle constitue bien un argument (ici un complément) du verbe, elle n'en occupe pas matériellement la position. Quoi qu'il en soit, cette formalisation permet de *retracer* à proprement parler par liage les relations qui s'établissent entre les deux pronoms C (pour complémenteurs, ang. *complementizers*) et l'argument complément du verbe[3].

Le module nommé X-barre, de son côté, attribue à toutes les constructions une condition d'endocentricité, autrement dit l'existence d'une polarité dans la représentation des syntagmes, ce qui minimalise la gamme des schématisations envisageables en grammaire. Le principe en est très simple : chaque unité se concrétise ou non dans sa représentation formelle (appelons-la X°), qui coïncide en général avec son instanciation, son emploi, sa matérialisation dans le syntagme. Celle-ci peut être complétée ou poursuivie par un complément quelconque (ici ZP), et former avec lui un syntagme (X'), qu'on représentera ainsi :



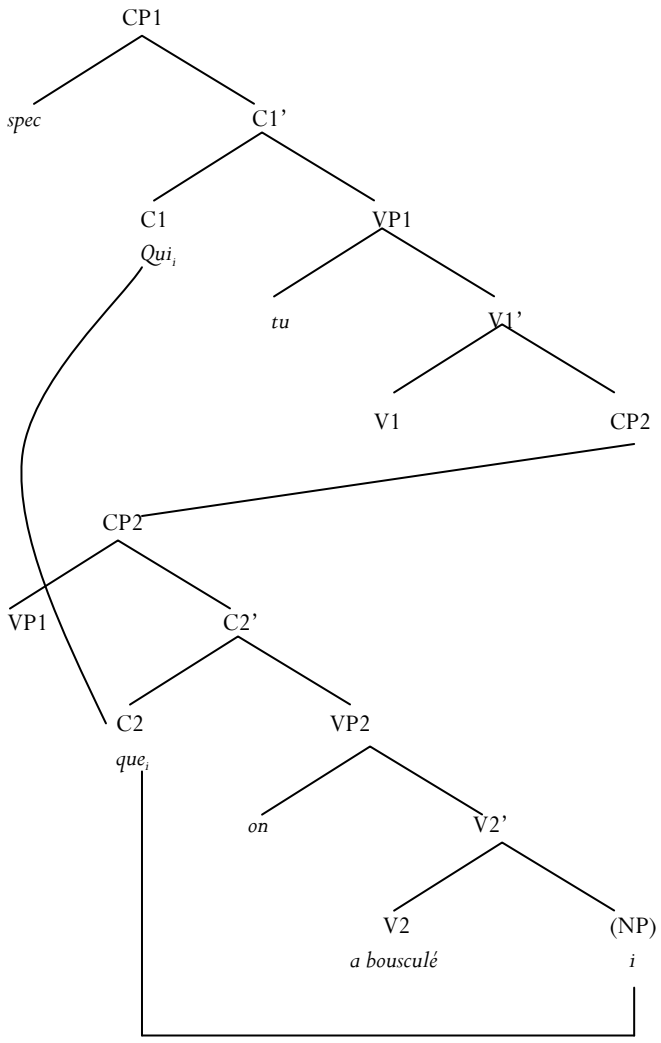
Ce syntagme comprend une ou plusieurs unités, et prend éventuellement appui sur un cotexte quelconque là aussi, qui fait office de spécifieur (ici YP) et qui permet surtout de projeter le syntagme X' dans un ensemble plus grand, et donc dans un réseau de relations avec d'autres constituants :



La projection des unités se fait alors en plusieurs temps, ce qui nous laisse suffisamment de place pour indiquer les relations qui existent entre les composantes du syntagme et donc de la phrase. Elle nous invite surtout à dégager les relations, dans le syntagme verbal, entre le verbe et ses arguments (par exemple sujet et compléments éventuels), de même que dans tous types de syntagmes (nominaux, adjectivaux ou autres), ainsi qu'à sélectionner les têtes argumentales à partir desquelles se tisse un réseau de relations.

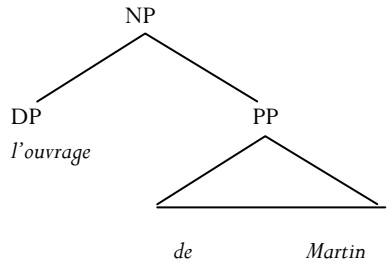
Pour l'exemple cité *supra* (*qui dis-tu qu'on a bousculé ?*), on estimera que le complémenteur *qui*, repris dans *que* et dans la position complément du verbe *bousculer*, constitue bien la tête d'un syntagme (*phrasal*) à partir duquel on va être en mesure de projeter les autres unités de l'ensemble (appelé éventuellement inflexion : I) :

[3] Cette phrase étant segmentable en deux temps (*tu dis quoi ? / qui a-t-on bousculé ?*), on estimera que les deux complémenteurs sont effectivement des pronoms, même si le second apparaît comme conjonctionnel à la suite de *dire*. Mais cela peut se discuter.

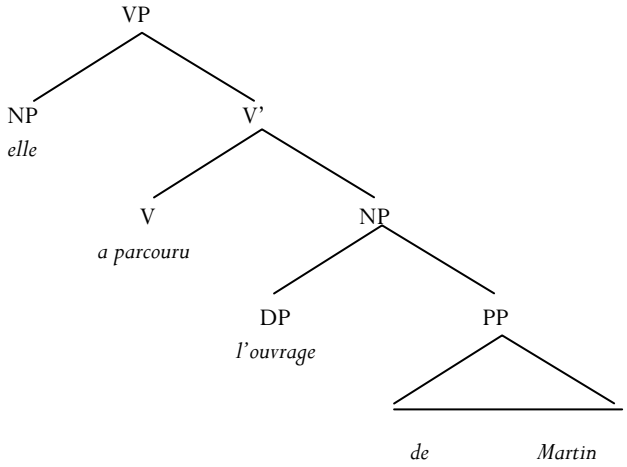


Quand elles sont privées d'arcs, les schématisations auxquelles on recourt en général reprennent, quoi qu'il en soit, cette coindexicalité qui caractérise en partie la représentation en arbre tout en permettant de flécher le parcours phorique des termes en présence. Tout cela n'est d'ailleurs pas complètement en contradiction avec les représentations que l'on dressera de la phoricité même en sémiotique (Fontanille 2001).

Il est facile dans ces termes de noter la relation de c-commande qui opère sur toutes les composantes renvoyant à la variable *x* tracée par l'indice *i*, dans le schéma, et de réaffirmer la présence d'un autre principe, celui de la bijection, qui exige une relation bi-univoque entre le complémenteur et la variable qu'il lie. D'une manière générale dans le programme minimaliste de Chomsky, il s'agit d'écarter tout ce dont on peut faire l'économie ; ainsi conserve-t-on les composantes de la grammaire qui ont une application directe dans la représentation des unités syntagmatiques, et notamment des phénomènes comme celui de l'accord (*agree*), que nous verrons en d'autres occasions. Les deux opérations qui s'avèreront les plus productives en termes de description seront donc l'assemblage (*merge*) et le déplacement (*move*). L'opération d'assembler (*merge*) se substitue en partie par simplification au module X-barre, mais répond au même principe : il s'agit de montrer comment deux éléments se combinent pour former un ensemble plus grand qui les résume. Ainsi I par exemple contient A et B qui sont sœurs (*sisters*). Quand nous disons *l'ouvrage de Martin*, nous prédiquons un SN (*NP*) qui contient d'un côté le SN 1 *l'ouvrage* et de l'autre le SN 2 (prépositionnel) *de Martin*, qui sont donc sœurs. En projetant ainsi SN 1 et SN 2 dans un même SN qui les contient, nous les fusionnons et en même temps nous en donnons la projection maximale qu'on puisse en donner.



Ce sous-ensemble (appelons-le ainsi), dans un VP tel que : *elle a parcouru l'ouvrage de Martin*, donne ceci :



Pour une question du type *kilès ou vle ?*, dont la représentation linéaire simplifiée serait [CP $_{C'}$ [C[kilès$_i$] [$_{VP}$[V' $_{NP}$[ou] $_V$[vle] *i*]]]], on note encore une fois que l'indication *i* témoigne du fait que l'unité qui constitue le régime du verbe est montée dans le complémenteur à l'appui d'un déplacement (*move*). L'effacement (*delete*) est effectif, mais on admet que l'unité laisse une copie (*copy*) à la position qu'elle occupe dans la chaîne syntagmatique. Ce principe de *delete / copy*, comme on l'aura facilement compris, remplace à présent opportunément le principe des traces précédemment exposé.

On notera que dans le cadre du principe de projection étendu (*EPP*, pour *extended projection principle*), les catégories du temps (T), du verbe (V) et du complémenteur (C) ont la possibilité d'attirer (*attract*) les caractéristiques d'un NP par exemple non présent comme spécifieur, même si cette position est déjà occupée. Au contraire, il se peut que celle-ci n'étant pas occupée, on soit contraint de faire remonter les caractéristiques (ou l'une d'elles) d'une unité postérieure dans la chaîne, et ainsi de suite, mais nous verrons plus exactement ce point dans le chapitre où il sera question des traits.

Centrés sur le syntagme ou non, certains formalismes prennent en compte le contexte (*Context Sensitive Grammars*), avec donc un certain nombre de paramètres complémentaires, là où d'autres n'en font pas cas (ainsi les *Context-Free Grammars*, CFG). Dans le cadre des CFG, on note certaines réécritures qui conduisent à simplifier les constructions grammaticales au-delà même de la linéarité de l'énoncé, ainsi :

VP : V NP

VP : V PP

De cette manière, un *verb phrasal* (*VP*, soit SV) est la réécriture, pour ainsi dire la reformulation récursive, d'un *verb* (verbe) et d'un *noun phrasal* (*NP*, pour SN) dans le premier cas, et d'un *prepositional phrasal* (*PP*, pour SP) dans le second. Il ne s'agit donc pas, suivant le programme minimaliste, de reproduire un inventaire de tous les énoncés envisageables, mais de dégager un ensemble limité (*finite set*) de réécritures libres du contexte. On peut affirmer dans ce sens que ce programme n'appartient plus aux premières vues de la grammaire générative, ce qu'on a pu reprocher à Chomsky lui-même. L'autre reproche que l'on peut formuler à l'égard de cette démarche revient à sa faible représentativité descriptive, et c'est pourquoi certains linguistes ont modifié les grammaires syntagmatiques dans le cadre de ce que Cori et Marandin 2001 appellent la *réforme*.

Nous n'exposerons pas intégralement les positions de Bresnan (*LFG*), Gazdar-Klein-Pullum-Sag (*GPSG*), Joshi (*TAG*), ni Shieber (*CBG*), qui dans les années 1980-1990 ont réformé, entièrement ou en partie, le programme minimaliste : ces reformulations méthodologiques font désormais partie d'un fonds commun auquel tout le monde a accès. L'appareil ontologique n'a pas bougé, mais ces dissidences ont occasionné l'apparition en un temps assez court de champs explicatifs distincts de la grammaire syntagmatique chomskyenne, tout en conservant certains des modules de cette approche. Le groupement des grammaires lexicales et fonctionnelles (*Lexical-Functional Grammars, LFG*), prolongé pour une bonne part dans les grammaires généralisées (*Generalized Phrase Structure Grammar*, *GPSG*) préconise d'envisager les catégories syntagmatiques et celles des unités à travers des ensembles de traits caractéristiques, chacun des traits étant réparti en deux temps : *attribut* d'un côté, *valeur* de l'autre (laquelle peut être attribuée par défaut : Da Sylva

1998). Cette bipartition nous rappellera bientôt celles des catégories et sous-catégories (<CAT, SUBCAT>) dans d'autres types de formalismes. Certaines catégories étant plus ou moins spécifiées, disons caractérisées par des traits, cela nous amène à conclure différents niveaux de généralité, d'autant que la théorie d'Optimalité qui s'applique aux LFG (Prince et Smolensky 1993) admet que toutes les régularités déclarées comme telles sont envisageables comme des règles (plus ou moins) universelles qui dépassent donc les spécifications particulières que l'on peut attribuer à une langue[4]. Plus particulièrement sur les schématisations en arbres, les plus grandes réformes méthodologiques reviennent aux *TAG* (pour *Tree Adjoining Grammars*) ainsi qu'aux Grammaires d'Arbres Polychromes (GAP). Mais avant tout rappelons que le genre, le nombre, la personne, la transitivité, le temps, sont autant de traits spécifiables (Cf. DeGraff 2000). Dans ce cas, prenons l'une des deux réécritures citées *supra*, à savoir *VP : V NP* en l'exemplifiant : on admettra qu'il est possible d'assigner des traits binaires (<attribut, valeur>) à chacun des constituants, notamment de la manière suivante et au plus simplifié (pour *fais tes devoirs*) :

[VP [V (<*mode*, imp.>, <*temps*, présent>, <*nombre*, singulier>, <*personne*, deuxième>)] [NP (DET (<*cat*, possessif>, <*nombre*, pluriel>, <*personne*, deuxième>)) (N (<*genre*, masculin>, <*nombre*, pluriel>))]]

Dans le cas d'un accord (*agreement*) entre un NP et un V dans une suite du type *VP : NP V*, quelques uns des traits spécifiés peuvent fusionner ensemble, d'autres étant éventuellement inopérants à un niveau supérieur, ainsi :

[VP [NP (<*nombre*, w>, <*personne*, x>, <*genre*, y>)] [V (<*nombre*, w>, <*personne*, x>, <*temps*, z>]] ⇨ [VP (<*nombre*, w>, <*personne*, x>, <*temps*, z>)]

On peut aussi envisager d'indiquer ces traits spécifiés dans le cadre d'une notation par contraintes (Shieber 1992) : ceux-ci doivent se révéler effectivement compatibles les uns aux autres, ce que l'on schématisera ainsi :

$$\begin{pmatrix} \text{NP} & \begin{pmatrix} \text{ACCORD} & \boxed{1} & \begin{pmatrix} \textit{nombre}, \text{ w} \\ \textit{personne}, \text{ x} \\ \textit{genre}, \text{ y} \end{pmatrix} \end{pmatrix} \\ \text{VP} & \begin{pmatrix} \text{ACCORD} & \boxed{1} \end{pmatrix} \end{pmatrix}$$

Ce qu'on appelle le mécanisme d'unification consiste à regrouper les traits ainsi spécifiés quels que soient les types de contraintes retenues. On assigne alors toutes les valeurs autour de caractérisations communes, par exemple aux LFG et aux GPSG, que l'on unifie afin de maximiser la portée descriptive des schémas. Effectivement, le principe d'économie dérivationnelle (Chomsky 1993) refuse en un sens l'approche de certains phénomènes laissés « libres » pour l'énonciateur, comme les déplacements optionnels des complémenteurs (*tu viens quand ? quand viens-tu ? / quand tu viens ?*) ; de ce fait, certains linguistes sont amenés à décrire ces phénomènes à l'appui de la phonologie plutôt que de la syntaxe. L'enjeu consiste donc, pour certaines constructions qui peuvent sembler *a priori* problématiques à première vue, de partir éventuellement de l'interface la plus simple entre linguistique et théorie computationnelle (Cf. Miller et Torris 1990). Cela n'est pas des plus faciles, par exemple dans le domaine méthodologique des GPSG. Une application de ce formalisme en particulier, comme le *bottom-up filtering*, revient à déterminer le premier constituant de la phrase, à partir de celui de gauche (le début, en somme). Ce type d'analyse procède donc en deux temps : d'une part, il s'agit de délimiter le début de la phrase ou du syntagme (*bottom-up*) pour ensuite envisager ce qui constitue le reste de cette unité (*top-down*). A partir de là, la règle de précédence linéaire nous conduit à positionner les arguments par rapport à la tête, pour admettre des règles générales (*phrase-structure rules*). Ainsi aurons-nous des suites du type :

S : NP, VP

NP : Det, N, AP, PP

[4] Pour reprendre les termes de K.P. et T. Mohanan, de l'Université de Singapour (« Universal and Language Particular Constraints in OT-LFG », 2003), on se place donc dans une problématique d'invariance et de récurrence.

NP : N
VP : V, NP, PP
Que nous linéariserons par exemple de la manière suivante :
S, <NP, N>, <VP, V>, <PP, Prep>, <NP, Det>, N.

Les analyseurs (*parsers*), notamment dans le domaine du traitement automatique, passent par la reconnaissance d'une structure préenregistrée qui inventorie les parties du tout (*part-whole paradigm*) en constituants, ce qui revient au même dans bien des cas. Ceux-ci déterminent comment les unités les plus grandes sont construites en termes de séquences d'unités plus petites (*smaller constituents*), mais ces descriptions formelles demeurent elles aussi limitées devant la flexibilité de la praxis linguistique (variation structurale et discontinuité). Grootjen 2001, pour réparer ces apories, nous invite à mettre en place un modèle relationnel entre les unités : à l'input apparaît une sélection des constituants à l'appui d'un *tagger*, suivi d'un classifiant (*classifier*) et ensuite d'un relateur (*relater*), l'ensemble débouchant sur une représentation arborescente. Mais le plus porteur selon nous dans cette modélisation, c'est qu'elle nous invite à répartir des schèmes de relation basique (*basic relation schemes*) à travers trois posés opératoires : la prédication d'une part (la relation entre le nœud prédicatif et ses arguments), la modification d'autre part (entre un modifieur et ses arguments), et enfin la qualification (entre le nœud et ses qualifiants). Ces trois schèmes relationnels peuvent être appliqués récursivement, et rassemblent ainsi leurs caractéristiques dans l'input. De cette manière, le tagger regroupe l'ensemble des numérotations pour un mot donné (avec la formule $T_i = \text{tag}(w_i)$). Même formalisation pour les classifiants :

$$C_i = \bigcup_{t \in T_i} \text{classify}(t)$$

ainsi que le relateur :

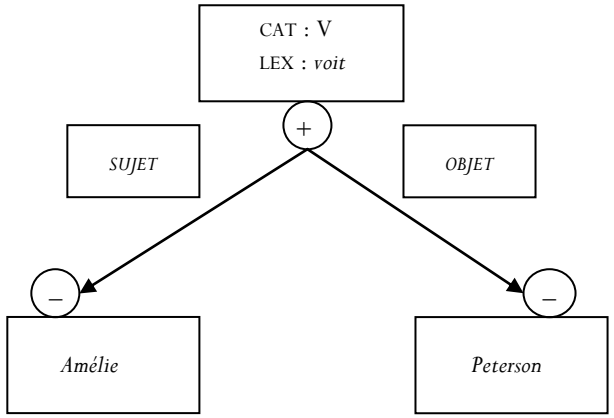
$r = C_1\, C_2 \ldots C_n$, avec $c_i \in C_i$

Ce type de parsage présente l'atout d'être simple et très abordable à la programmation. Nous ferons également remarquer qu'il admet l'existence d'invariants linguistiques en faible nombre.

Le paradigme génératif se réfère, quoi qu'il en soit et de l'aveu de Chomsky lui-même, à la notion de système constructionnel : ce qui organise une grammaire ne tient pas d'un répertoire de formes et de matérialisations linguistiques, mais de la forme logique d'un posé théorique. Ainsi, à l'appui d'un nombre fini de principes (auparavant des règles de réécriture), il est admis qu'une langue implique un nombre infini d'énoncés possibles que la grammaire formelle (syntagmatique, soit *Phrase Structure Grammar, PSG*) permet d'engendrer de manière récursive. Les principes sont combinables (on peut déterminer un *Pr >a* et un *Pr >b*, ainsi qu'un *Pr a,b*, ou *a,a*, ou *b,b,a*) mais aussi compressibles, selon que la grammaire en question est sensible au contexte (pas vraiment de restriction de classes, de catégories) ou non, selon la hiérarchie de Chomsky. L'ensemble projette donc de relever le maximum de caractéristiques communes à plusieurs faits linguistiques, mais aussi de dépasser les problématiques embarrassantes comme la constituance non bornée (*unbounded branching*), la discontinuité de certains syntagmes, ainsi que l'éventuelle non-coïncidence entre l'ordre linéaire de la chaîne et la constituance argumentale. En outre, au-delà des unités lexicales, les unités fonctionnelles (comme le temps ou l'accord), apparaissent comme autant de spécifications qui s'imposent (dans le cadre de ce qu'on a appelé au début l'*Augmented Transition Networks, ATN*). Nous nous plaçons donc loin d'une linguistique générale à proprement parler : nous sommes plutôt dans une phase intermédiaire, avec la mise en place d'un certain nombre de principes (*principles*), de formalismes (*grammars*) et de programmes (*schedules, programs*). Cela étant, on aura bien compris que c'est d'abord par souci d'exactitude (*accuracy*) et d'appropriété descriptive (*adequacy*) que les grammaires syntagmatiques vont donc peu à peu étendre le champ des caractéristiques notées et des types répertoriés (ainsi dans les LFG, TAG et GPSG). Ces deux exigences méthodologiques, lesquelles sont plus ou moins explicitement revendiquées par les auteurs concernés, rapprochent considérablement la linguistique de l'informatique dans le domaine désormais répandu des *non-transformational grammars* : les formalismes, déclaratifs, impliquent un nombre variable de composantes descriptives basées sur l'unification (Shieber 1986, auteur du *Parse and Translate : PATR II*). Les GPSG, dans cette vue, posent elles aussi les catégories comme des ensembles de traits spécifiés à travers la

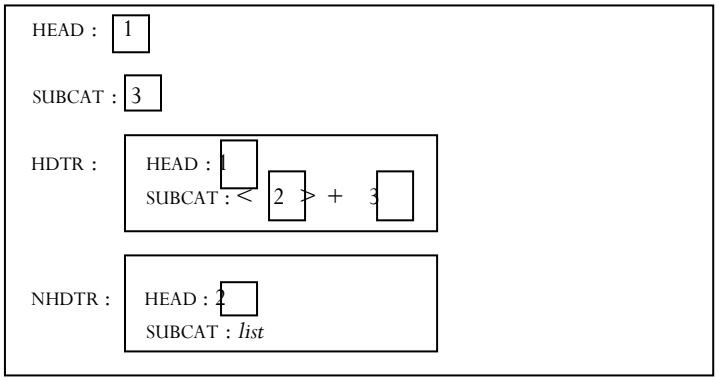
notation <attribut, valeur>, ce que fait en un sens le module X-barre en ceci qu'il répartit les catégories syntagmatiques en types catégoriels et niveaux syntagmatiques (par exemple en termes de spécification de traits, nous aurons <N, +>, <V, –>, <BARRE, 2>, <PERS, 3>, <PLU, –>, <CAS, DAT>, pour le pronom *lui* dans *je lui donne à boire*). Ce genre de spécifications unifiées permet certaines généralisations (d'où les *G*-PSG), comme les règles de dominance immédiate et de précédence linéaire. Le trait SOUS-CAT, par exemple, constitue une règle de précédence linéaire qui permet d'ordonner une expression sous-catégorisée après l'expression qui la sous-catégorise (voir Cori et Marandin). Or, ces réformes à la fois rédactionnelles, notationnelles et schématiques passeront dans le domaine de la composition des arbres, à travers surtout les grammaires d'arbres ajoints (*TAG*) et les grammaires d'arbres polychromes (GAP). La GPSG ne modifiant pas les formulations schématiques arborescentes, ces deux approches méthodologiques opèrent (sans être en marge des *PSG*), pour le premier, par composition et par adjonction, permettant ainsi de traiter les relations entre constituants à distance, et, pour le second, par la mise en ordre des branches de l'arbre par rapport à un pivot (lequel peut être la tête d'un constituant endocentrique), tout en fusionnant les niveaux hiérarchiques dans le cas d'un même pivot pour plusieurs types de relations[5]. Nous avons aussi cette structure de traits dans les *Functional Unification Grammar* (*FUG*), à ceci près que cette fois-ci, ce sont les spécifications de traits qui priment, et non la grammaires et ses règles. Les traits sont donc typés, notamment pour décrire les *frames* à disposition de l'énonciateur (ensemble de schèmes actionnels linguistiques).

Comme nous l'avons vu, les formalismes basés sur des contraintes s'appuient sur des ensembles de variables regroupées dans des propriétés. Ils passent par une caractérisation générale des phénomènes relevés (ou supposés) et par des représentations (*category, set of properties, set of features* : Blache 2001). Dans un même esprit, une structure s'appuie sur des objets : pour un graphe par exemple, les objets sont des nœuds ou des arcs, et ces deux types d'objets sont liés entre eux dans le cadre des grammaires d'unification polarisées (GUP) notamment, où l'on attribue à chaque objet une polarité. Ainsi, les structures générées par le formalisme renvoient à des structures neutres combinées à des structures élémentaires en nombre restreint. On dispose donc des arbres où les nœuds sont (sous)spécifiés à travers une structure de traits : en distribuant des polarités (positives et négatives) aux traits <cat> plutôt qu'aux nœuds directement, on ordonne les branches de l'arbre de manière à dégager les structures élémentaires sous forme de cellules dont le bord supérieur est une chaîne d'arcs positifs qui correspondent à la partie gauche et qui sont appelés à s'unifier avec les autres arcs :



Il en est de même pour les grammaires d'adjonction d'arbres (*TAG*), lesquelles sont les premières grammaires de combinaisons de structures (et d'arbres ordonnés), à savoir qu'elles peuvent être composées d'arbres intermédiaires (ou quasi-arbres) dans lesquels d'autres quasi-arbres peuvent intervenir. Le rapprochement avec les HPSG est évident, surtout avec des schémas de base en *head-daughter-phrase*, qui présentent les constituants de tête avec des sous-constituants-filles catégorisés :

[5] On emploie volontiers en Haïti, et cela dès les classes du primaire, le terme de pilier (*pilyè*), pour désigner à proprement parler le pivot d'un constituant ou d'une suite de constituants. Mettons la phrase en cr. h. *pou ki es ou deyè kay la ?*, l'adverbe *deyè* fera office de *pilye* du syntagme verbal (certains parlent d'un *terme prédical*), cette suite étant paraphrasable en *pou ki es ou vin cheche kay la ?* ou encore *pou ki es ou ap cheche kay la ?*.



Cette fois-ci, chacune des listes de traits est représentée de manière récursive en deux temps et de manière caractéristique : celle de la tête et celle de ses arguments-filles (non têtes). D'ailleurs, la notation SUBCAT est en elle-même une liste de constituants sou-spécifiés, alors que les constituants-filles sont soit instanciés (*list*), soit non instanciés (*elist*). La liste SUBCAT du constituant-tête, elle, se rapproche d'une concaténation de traits. Dans cette vue, la polarisation permet de noter des listes SUBCAT, et donc de contrôler l'effection de la valence (verbale, notamment), quand, de leur côté, les LFG et les grammaires synchrones nous conduisent à synchroniser des grammaires syntagmatiques hors-contexte avec des grammaires dites de dépendance, à savoir que celles-ci ne dégagent pas d'arbres, mais des structures de traits réentrantes, autrement dit des structures fonctionnelles :

S > NP VP
$\downarrow = \uparrow$ *sujet* $\downarrow =$

Chaque nœud syntagmatique apparaît de manière synchrone avec un nœud fonctionnel, là où les notations en $\uparrow$ et $\downarrow$ renvoient au nœud fonctionnel synchronisé avec un nœud syntagmatique. Cette équation démontre bien que le nœud syntagmatique courant et son *père* sont synchronisés avec le même nœud *f*.

Les approches énumérées ci-dessus sont donc autant de formalismes qui ne mènent pas vers ce qu'il conviendrait d'appeler une linguistique générale, dont elles sont plutôt une verbalisation intermédiaire, généralement non philosophique, laquelle demeure d'ailleurs suspendue à des programmes d'unification régulièrement revus et corrigés. On ne peut toutefois saisir complètement ces démarches sans avoir constamment présent à l'esprit ce qui provoque cet intense bouillonnement de caractérisations et de classifications, à savoir un véritable souci de productivité.

Bibliographie sélective :

Blache Ph., 2001 : *Les Grammaires de propriétés : des contraintes pour le traitement automatique des langues naturelles*, Paris, Hermès (coll. *Sciences*).

Chomsky N., 1993 : « A minimalist Program for linguistic theory », in *The View from Building 20 : Essays in linguistics in honor of Sylvain Bromberger* (*dir.* K. Hale et S.J. Keyser), Cambridge, MIT Press.

Chomsky N., 1995 : *The minimalist Program*, MIT University Press.

Chomsky N., Lasnik H., 1993 : « The Theory of principles and parameters », *in Syntax : an international handbook of contemporany research* (*éd*. J. Jacobs, A. Von Stechow, W. Sternefeld et T. Vennemann), Berlin, éditions de Gruyter.

Cori M., Marandin J.M., 2001 : « La Linguistique au contact de l'informatique : de la construction des grammaires aux grammaires de construction », *Histoire Epistémologie Langage (HEL)* 23, 1 : 49-79.

Da Sylva L., 1998 : *Interprétation linguistique et computationnelle des valeurs par défaut dans le domaine syntaxique*, Université de Montréal, Département de Linguistique et de traduction, thèse de Ph.D.

DeGraff M., 2000 : « A propos de la syntaxe des pronoms objets en créole haïtien : points de vue croisés de la morphologie et de la diachronie », *Langages* 138, Larousse : 89-113.

Fontanille J., 2001 : « La Sémiotique est-elle générative ? », *Linx* 44 : 107-132.

Gootjen F., 2001 : « Relational Indexing Using a grammarless parser », in *Proceedings of the 2001 IEEE Systems, Man, and Cybernetics Conference*.

Miller P., Torris T. (*éd.*), 1990 : *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Paris, Hermès.

Prince A., Smolensky P., 1993 : « Optimality Theory : constraint interaction in generative grammar », *RuCCS Technical Report 2*, Piscateway, NJ Rutgers University Center for Cognitive Science.

Shieber S.M., 1986 : *An Introduction to unification-based Approaches to grammar*, Stanford, CSLI.

Shieber S.M., 1992 : *Constraint-Based Grammar Formalisms*, MIT, MIT Press.

Sportiche D., 1997 : « Reconstruction and constituent Structure », *Linguistics Colloquium Series Talk*, october 1997, MIT.

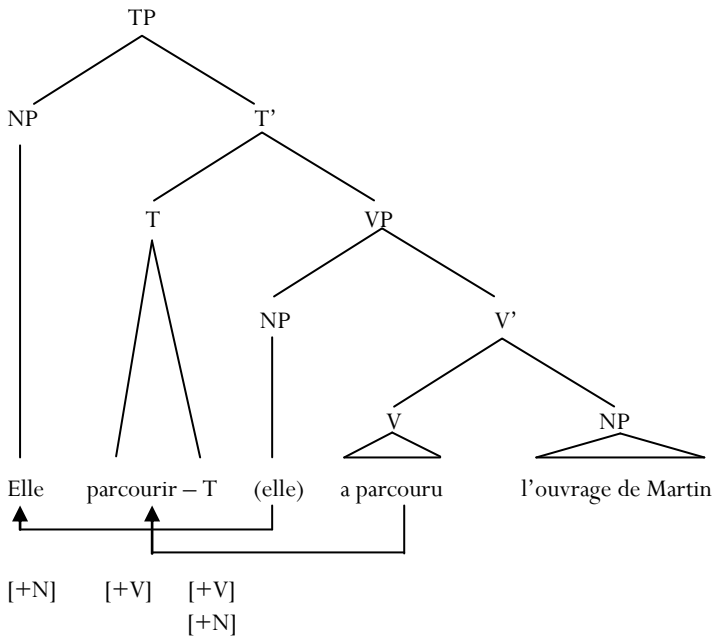
Evaluations possibles :

- discuter l'intitulation de « linguistique générale », en référence éventuellement à Saussure, et en la confrontant à la question des formalismes linguistiques contemporains ;
- exposer ce qui distingue un formalisme (*grammar*) d'un paradigme (*paradigm*) et d'un programme.

SEANCE 2 : CATEGORISATIONS ET CLASSIFICATIONS

Contrairement à la distinction classificatoire *syntaxique / lexical*, qui s'avère tout à fait opératoire en grammaire générative, celle qui consiste à séparer le grammatical du lexical aboutit irrémédiablement à la mise à l'écart du second hors de la grammaire. Dans le projet de grammaire universelle, cette marginalisation du lexique, lequel est peu formalisable et donc peu schématisable, assimile les éléments lexicaux à des unités qu'il suffirait d'instancier le moment venu (Rastier 2001), alors que dans les grammaires dissidentes sur cette question, comme les grammaires lexicales et fonctionnelles et les grammaires généralisées (LFG et GPSG), ceux-ci sont ouvertement réintégrés dans les schémas (Cf. Sauzet et Zribi-Hertz 2003). La grammaire générative a suscité bien des questions sur la place qu'il convient d'attribuer aux unités lexicales dans les représentations, tout en les laissant peu à peu en marge. Cette position a d'abord présenté un caractère très embarrassant, car il est évident que les niveaux lexical et syntaxique participent conjointement, et non subsidiairement pour l'un ou pour l'autre, à la signification de l'unité. Désormais, il est admis que ces unités s'inscrivent dans un contexte, et dans d'autres termes à l'intérieur d'une microstructure de la cognition. On leur assigne donc des traits (*features*), par exemple phonologiques, tout comme on en assigne aux items syntaxiques.

Dans ce cadre, l'assignation de traits est tout aussi présente que dans les autres formalismes linguistiques. Par exemple ci-dessous :



On assiste dans ce schéma à la représentation de deux déplacements (montées, ang. *raising*), d'une part du NP *elle* qui monte en spécifieur de TP, et d'autre part du V *a parcouru* qui vient s'adjoindre à la projection minimale de T. On peut dès lors assigner des traits qui viennent expliciter ces mouvements de NP et de V dans TP et T : le NP *elle* comporte un trait [+N] et le verbe un trait [+V]. La projection du temps (T) combine les deux traits ([+V] [+N]) qui s'assemblent en lui et viennent s'y adjoindre, ce qui serait aussi le cas dans la projection de l'accord. On dira que T attire (*attract*) les traits de l'un et de l'autre (l'*EPP* opère de la sorte sur plusieurs niveaux).

On se place donc ici en marge (pour ainsi dire dans l'*après*) des grammaires transformationnelles, à l'appui des grammaires d'unification, qui sont basées sur des structures de traits (*feature-based approaches*), comme les *GPSG*, *HPSG* et *LFG*, mais aussi à l'appui des grammaires relationnelles (*relational approaches*), et à celui des grammaires catégorielles, lesquelles combinent des règles schématiques pour aborder les expressions (formalismes de Lambek ou de Montague, CG). Dans la première approche, des traits notamment communs aux constituants syntagmatiques sont unifiés, à proprement parler amalgamés (Blevins 2002). Par exemple en GPSG, les traits d'une phrase finie peuvent être hérités d'un verbe suivant la supposition que les phrases sont des projections verbales endocentriques.

Impulsées par Pollard et Sag (Cf. Sag et Wasow 1999, Ginzburg et Sag 2001), les grammaires guidées par la tête (*Head-driven Phrase Structure Grammars, HPSG*) demeurent conformes en certains points au programme génératif, et s'affirment comme modulaires en ceci qu'elles intègrent les principales composantes linguistiques du formalisme de Chomsky, comme la phonologie, la syntaxe, le lexique et la sémantique ; de plus, elles ont pour objet téléonomique de pouvoir générer une infinité de phrases possibles dans la langue. Les innovations des HPSG reviennent à la représentation d'ensemble de ces composantes, et de leur combinaison, à travers notamment les structures de traits dans un formalisme binaire d'*attribut/valeur* (*AVM*). En outre, elles s'établissent sur une exigence de bonne formation des phrases (dans ce sens, elles sont déclaratives : elles comportent des déclarations de *wellformness* [6]), et concèdent une place centrale au lexique (les principes afférant aux structures de mots sont distinctes de ceux qui concernent celles des phrases, lesquelles n'influencent qu'occasionnellement la signification des mots). Chacune des phrases ainsi représentée correspond à une seule représentation, laquelle est monostratale, et ne comprend pas de catégorie vide. Les traits sont les suivants : phonologique (PHON : la composante phonématique et prosodique du mot ou de la suite), à la fois syntaxique et sémantique (SYNSEM : comprenant le niveau LOCAL (LOC), à savoir la catégorie (CAT), le contenu (CONT), le contexte (CTXT), et le niveau NON LOCAL (N-LOC), qui représente des données comme des arguments non instanciés ou des dépendances non bornées à travers le mécanisme *slash*).

[6] Ces déclarations de bonne formation des énoncés en général prennent appui sur des contraintes à satisfaire simultanément, comme tous les modèles basés sur des contraintes.

L'ensemble de ces traits se regroupe dans une hiérarchie typée : les types, à travers l'héritage des (sous)spécifications, répertorient les relations entre les signes relevés. Par exemple, le trait de la catégorie (CAT) permet secondairement de sous-spécifier quels sont les traits de TETE et ceux de VALENCE, avec, pour cette dernière, une sous-spécification des arguments comme le sujet et les compléments du verbe.

Résumons : une suite peut être typée comme étant un mot, un syntagme, une phrase, et répond à des contraintes de bonne formation. Pour ainsi dire sous-typé, le signe représente la matérialisation de ce qui est énoncé. Un mot, dans ces termes, aura par exemple pour contraintes d'appartenir à une catégorie et d'avoir tel ou tel argument, dont on convoque la liste :

TYPE	*CONTRAINTES*
mot	[CAT [ARG-ST *list(synsem)*]]

or, il en est de même pour le syntagme, à ceci près qu'il ne répondra pas aux mêmes contraintes à l'emploi :

TYPE	*CONTRAINTES*
syntagme	*signe* DTRS [HD-DTR *signe* NHD-DTRS *list(signe(s))*]

La notation *argument-structure* (ARG-ST) résume l'ensemble des sous-catégorisations du mot (SUBCAT), avec une liste de caractéristiques SYNSEM qui renvoient aux arguments dégagés. Concernant le syntagme, la notation *daughters* (DTRS) spécifie les arguments instanciés du syntagme (dans des branches, donc, aussi appelées *filles*), avec la tête (HD-DTR) et les arguments non-têtes (NHD-DTRS). Dans ce cas, il convient de déterminer d'abord la capitalité du syntagme (quelle est sa tête : *headedness*) et quelle est sa clausalité, autrement dit le type de syntagme ou de phrase matérialisé. Nous voici donc au carrefour de deux méthologies descriptives qui jusque là ont pu se confronter : d'une part, le fait que ce formalisme soit guidé par les têtes le rapproche du formalisme X-barre en termes de constituants, avec donc des syntagmes dominés par une tête lexicale (à noter qu'il peut y avoir des *headed phrases* comme des *non-headed phrases*) ; d'autre part, la problématique de la clausalité rapproche les HPSG des grammaires de construction, qui assignent des contraintes spécifiques à chaque type de phrase (qui sont « classiques » à proprement parler, et discutables).

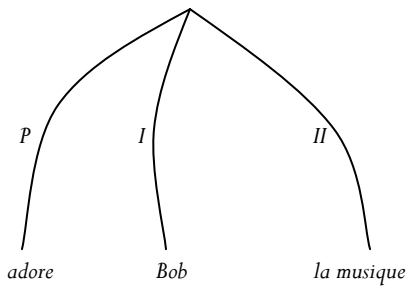
D'une manière générale dans les HPSG, un syntagme consiste donc dans la combinaison de mots (instanciés ou non) qui sont autant de branches argumentales du schéma, et qui sont (sous)spécifiés, et la plupart du temps numérotés en [1], [2],…[*n*] (les numéros sont appelés des *tags*, et sont listés (ou concaténés) à l'appui de la notation $\oplus$). De même pour un mot (ici un verbe transitif) :

mot [SYNSEM / LOC / CAT [VALENCE [SUJ [1] COMPS [2]] SOUS-CAT [1] $\oplus$ [2]]]

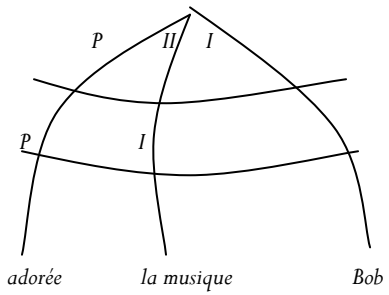
Dans les cas de coordination de plusieurs à plusieurs, d'extraction, d'extraposition, d'oblicité multiple (quand par exemple un verbe a pour compléments un comp. direct, indirect, attributif ou autre ou plus), ou encore de cliticisation, les constituants n'ont pas forcément à être linéarisés (on traite ainsi plus facilement la discontinuité des constituants), et les différents niveaux de représentations n'ont pas à coïncider dans tous les cas. Ce qui importe, c'est d'assigner le maximum de traits pour que l'ensemble du mot ou du syntagme soit effectivement décrit.

Dans le cadre des grammaires relationnelles initiées notamment par Perlmutter en revanche, on pose d'une part que les relations argumentales

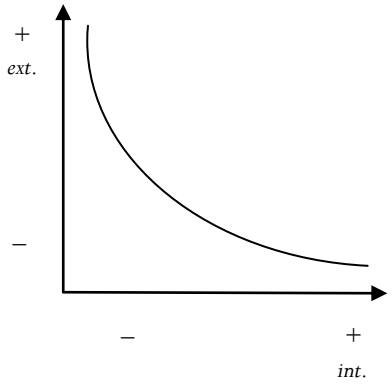
sont des *constructs* qui ne peuvent être ramenés à des propriétés juste thématiques, phonologiques ou autres, mais à des relations autour du noyau (*core relations*), résumables dans des termes (*terms*) qui les apparentent aux listes ARG-S des HPSG et aux R-arguments des LFG. D'autre part, cette approche présuppose que les systèmes grammaticaux sont intrinsèquement multistrataux, avec plusieurs niveaux syntaxiques dans lesquels les expressions peuvent se voir assigner plusieurs relations distinctes. Il s'opère donc tout un réseau de relations (*relational networks*) qui sont représentables par différentes strates, lesquelles sont elles-mêmes schématisables sous la forme d'arcs. Ainsi une phrase (*S*) telle que *Bob adore la musique* tisse un réseau de relations entre les constituants : le verbe *adore* est le noyau prédicatif de la phrase, et il est marqué par un arc nommé *P* (*predicate*), alors que deux autres arcs (I et II) indiquent que *Bob* est sujet, et que *la musique* est complément :



Une opération telle que la passivation sera donc représentée par la présence d'une strate supérieure, qui indique notamment le réagencement des arguments :



On notera que le terme de *valence* désigne, notamment en sémiotique tensive, l'ensemble des liens tensifs qui unissent le noyau aux constituants qui se rapportent à lui. Un terme épilinguistique comme celui de *genre* est lié à ses hyponymes par le biais d'une valence objectale (ainsi *masculin* et *féminin*), tout comme le terme *moun* et ses dérivés, et comme, dans une construction locutionnelle, *moun* dans la suite *moun du deden* est lié en tant que noyau au syntagme nominal qui le caractérise (*du deden*) et réduit par là-même son extensité. D'ailleurs, sur un plan paradigmatique, on peut schématiser facilement dans un diagramme les gradients d'intensité et d'extensité d'un terme quelconque, comme suit, avec l'exemple *siège* :



La variable d'extensité (*ext.*) reporte à l'appariement du vocable à sa catégorie, à sa classe (par exemple quand un terme se rapproche plus ou moins du terme le plus typique qui peut lui être assigné, comme le *féminin* du *genre*, le *vétiver* de l'*épice*), alors que la variable d'intensité (*int.*) renvoie de son côté à la signification même du terme et de ses modulations (Fontanille et Zilberberg 1998). Par exemple, si je déclare *Tom est un agneau*, j'emploie le vocable d'*agneau* non pas à travers son extensité et en tant que tel, mais comme un terme intense par un biais analogique. La variable d'intensité, ainsi, qu'on la

place d'ailleurs sur un plan paradigmatique ou sur un plan syntagmatique, concerne donc la manière qu'on aura de moduler la signification même du vocable, contrairement à celle d'extensité, qui permet à proprement parler de répandre (Hjelmslev 1972) cette signification.

Cette schématisation par diagramme et à l'appui de cette variabilité présente un enjeu véritable dans le cas des catégorisations. Effectivement, cette déformabilité même nous invite à admettre que ce n'est pas la relation paradigmatique qui conditionne la catégorie, mais ce vers quoi la catégorisation aboutit, d'autant plus que celle-ci renvoie à un processus de typification qui présente forcément un caractère culturel plus ou moins affirmé (un chat, à Port-au-Prince, n'a certes pas la même intensité affective qu'à Paris, de même que la vache en Europe et en Inde).

Deux types d'approches semblent correspondre à ces caractéristiques d'intensité et d'extensité : d'une part la classification onomasiologique, qui intègre l'unité dans une classe, une catégorie en particulier, et la projette dans un ensemble plus ou moins fini d'autres unités de la même classe (ce qui renverrait plutôt à l'extensité) ; d'autre part, la classification sémasiologique, telle qu'elle est pratiquée en lexicographie par exemple, qui consiste à assigner des acceptions différentes à l'unité suivant l'emploi qui en est effectué en contexte.

Les unités que nous employons en discours sont discrétisées, c'est-à-dire qu'elles sont envisageables en termes d'emploi spécifique, contextualisé, mais aussi qu'elles s'intègrent dans une phrase. Prenons un exemple : *belmè Jan va pase*. Dans cette phrase, *belmè (Jan)*, sujet du verbe *pase* et qui s'inscrit dans sa valence verbale (disons comme spécifieur), s'inscrit dans un syntagme nominal dont il est la tête, et prend une acception particulière dans ce contexte : en tant que tel, il s'agit d'une unité discrète. Si nous déclarons, analogiquement, que *Daline two belmè*, nous extrayons du terme de quoi caractériser *Daline*, et grammaticalement, nous adjectivons le nom-tête du SN *two belmè* et le faisons rentrer dans la valence de *Daline*. Or, cette fois encore, un tel emploi concerne *belmè* comme unité discrète, alors que le lien que nous pourrions établir entre la belle mère de Jean, qui est bien vraie, et cette manière qu'a Darline d'en être une, concerne un réseau d'acceptions possibles qu'il est possible d'attribuer au terme : ce qu'on appelle son schéma continu.

Les catégorisations relèvent aussi, dans bien des cas, du phénomène culturel. On se rappellera ce qu'en disent Rosch et Lloyd à propos de ce désir irrépressible de segmenter le monde qui nous caractérise, mais qui révèle aussi une manière de nous l'accaparer. Envisageons juste la catégorisation des temps verbaux : celle-ci, au-delà même du fait que des langues favorisent les affixes flexionnels là où d'autres emploient des opérateurs distincts, ne renvoie pas aux mêmes vécus perceptifs des individus (Cf. Bentolila 1987, Damoiseau 1989). En termes grammaticaux par exemple, que dire de la préposition en chinois ? de l'adverbe en peul ?

Beaucoup de linguistes, confrontés bien entendu à la variabilité linguistique, se sont donc entendus pour attribuer au phénomène de catégorisation (*cat*) une dimension hyperonymique (ainsi les concepts de spécifieur, de modifieur, de complémenteur, qui transcendent les catégories grammaticales comme l'adverbe et la préposition) [7]. Les variables hyponymiques de ces catégorisations s'appelleront donc des sous-catégorisations (*sub-cat*). Les premières classifications catégorielles sont donc non déformables, alors que les classifications sous-catégorielles sont déformables. A ce titre, si personne ne conteste l'affirmation de degrés de typicalité des objets, cet élan n'est nullement unanime aussitôt qu'il s'agit de recourir au présumé « meilleur exemplaire », le prototype. Celui-ci, effectivement, peut être considéré comme une abstraction, une fiction (Rosch), un objet qui présente un niveau maximal de typicalité parmi les autres membres de l'ensemble dont il est une représentation idéale, mais aucunement le principe organisateur (Rastier 2001). En marge de cette problématique, et pour revenir à ce que nous disions plus haut, il est plus indiqué sans doute de parler de termes extenses et de termes intenses. Un terme extense regroupe suffisamment d'acceptions pour désigner un certain nombre de sous-catégories : dans ce sens, il présente un haut degré de typicalité (par exemple, les termes de *légume* ou encore de *locomotion*).

[7] Voir là-dessus en particulier Kriegel 2003, ainsi que Coveney, Hintze et Sanders 2004.

Quant à la catégorie PHRASE (*sentence*) n'est pas facile à cerner, en ceci qu'elle dépend de données prosodiques, et donc phonologiques, mais aussi syntaxiques (quand il s'agit par exemple de faire la part de l'ellipse ou de représenter les mécanismes d'extraposition), sémantiques (qu'est-ce qui forme sens dans la phrase, entre phrases ?), discursives (dans le cas notamment des énoncés co-construits par plusieurs énonciateurs, ou dans celui d'une polyphonie énonciative), éventuellement lexicales (Cf. la problématique des « mot-phrases »). Le formalisme génératif insiste ainsi sur le fait que S n'est pas la projection d'une tête, de sorte que certains linguistes parlent d'une catégorie exocentrique ternaire (Pollock 1998), ce qui signifie que nous avons là une catégorie sans tête (ou exocentrique), et que cette notation domine des constituants comme le verbe et ses compléments, la préposition et ses compléments[8]. A la question de savoir, dans ces termes, si la phrase représente une structure plate ou une structure hiérarchisée, répondons qu'en grammaire syntagmatique et dans le cas d'un SV contenant un sujet, un verbe et ses éventuels compléments, le SV lui-même domine tous les constituants, qui sont alors des filles de SV, tout comme elles sont des filles de S. On s'extrait ainsi de la grammaticographie classique et de la problématique d'une bipartition entre le sujet et les compléments du verbe : ici, le SV exerce une dominance du même ordre, et les constituants filles ont le même poids structural (le sujet ne se distingue des compléments qu'en ceci qu'il apparaît linéairement avant eux). En GG, les déplacements (*move*) des constituants dans le SV plaident à leur manière en faveur de ce principe : seuls les constituants filles peuvent être déplacés. De même, les traits de valence démontrent que des rôles thématiques sont assignés en interne à des catégories têtes en positions de spécifieur ou de complément (par exemple, un verbe transitif comme *vouloir* exige deux rôles thématiques internes (sujet/objet), auxquels on peut attribuer des valeurs telles que *agent, source* ou autre, dans une composition de relations argumentales qui coïncident à des rôles qu'il est possible de sous-spécifier de plusieurs manières). Quant à l'ordre des constituants, il est admis que les variations relevées dans ce domaine montrent combien elles sont à même d'intégrer une typologie des langues au-delà des computations syntaxiques qui leur sont supérieures (par exemple, les sujets initiaux du français, de l'espagnol et du créole haïtien tranchent avec les sujets finaux du coréen et du japonais). Autre exemple, le trait [+ Q], qui renvoie à un quantifieur (instancié en *wh* en ang., en *qu* dans les langues romanes, en *k* en créole haïtien – d'après ici une langue romane) a des chances d'être déplacé dans le syntagme verbal, mais il conserve une position argumentale quoi qu'il en soit attribuée avant son instanciation. Les avis formulés en la matière sont assez bien résumés notamment par Baltin et Collins 2001, lesquels abordent des sujets aussi variés que le sont les question des relations thématiques, de la passivation, du A-mouvement, de l'extraposition et de la négation de phrase.

[8] La suite *une femme* contient une tête qui domine un constituant ([DET *une* [N *femme*]]), elle est donc unaire, alors que *pour une femme* est binaire (la suite précédente apparaît comme le régime de la préposition : [PREP *pour* [DET *une* [N *femme*]]]. Le SP ainsi dégagé domine alors deux constituants, autrement appelés *sœurs* (*sisters*).

Bibliographie sélective :

Baltin M., Collins Ch. (*dir.*) 2001 : *The Handbook of contemporany syntactic theory*, Malden, MA & Oxford, Blackwell.

Bentolila A., 1987 : « Marques aspecto-temporelles en créole haïtien : de l'analyse synchronique à la formulation d'hypothèses diachroniques », *La Linguistique* 23, 1.

Blevins J.P., 2002 : « Nontransformational Grammar », in *The Linguistic Encyclopedia* (éd. K. Malmkjaer), Routledge.

Coveney A., Hintze M.A., Sanders C. (*dir.*), 2004 : *Variation et Francophonie*, Paris, L'Harmattan.

Damoiseau R., 1989 : *Système aspecto-temporel, applications à l'enseignement du français*, Port-au-Prince, Imprimerie le Natal.

Fontanille J., Zilberberg C., 1998 : *Tension et Signification*, Paris, Pierre Mardaga éditeur.

Ginzburg J., Sag I.A., 2001 : *Interrogative Investigations : the form, meaning and use of english interrogatives*, Stanford, CSLI.

Hjelmslev L., 1972 : *La Catégorie des cas*, Munich, W. Fink.

Kriegel S. (*dir.*), 2003 : *Grammaticalisation et Réanalyse*, Paris, CNRS éditions.

Pollock J.Y., [1997] 1998 : *Langage et Cognition*, Paris, PUF (coll. *Psychologie et sciences de la pensée*).

Rastier F., 2001 [1991] : *Sémantique et Recherches cognitives*, Paris, PUF « Formes sémiotiques » (2e édition revue et corrigée).

Sag I.A., Wasow T., 1999 : *Syntactic Theory, a formal introduction*, Stanford, CSLI Publications.

Sauzet P., Zribi-Hertz A. (*dir.*), 2003 : *Typologie des langues d'Afrique et Universaux de la grammaire*, Paris, L'Harmattan.

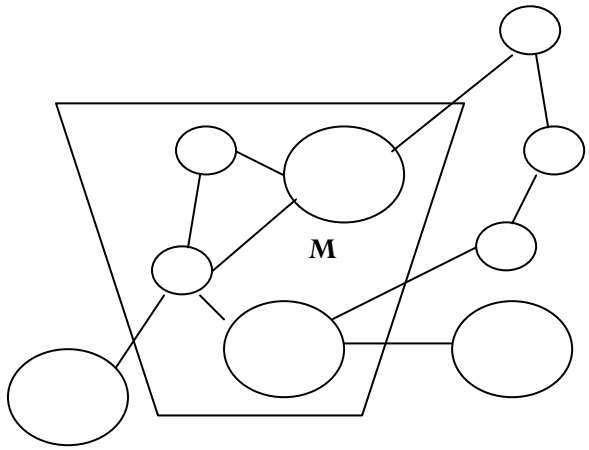
Evaluations possibles :

- expliquer, schémas à l'appui, l'opportunité de l'assignation de traits caractéristiques (*features*) en linguistique ;
- dégager une bibliographie commentée de trente entrées au moins sur la notion de *phrase* en linguistiques créole et romane.

SEANCE 3 : EN TERMES D'APPROCHES MODULAIRES

Une approche modulaire présente une méthodologie qui part de sous-systèmes qu'on peut appeler des modules, et qui permettent de traiter une à une des problématiques restreintes (Nølke 1999). Ces modules répondent à des règles locales (*local rules*), lesquelles sont subsumées dans le système par des métarègles (*metarules*) qui les relient entre elles. Cette modularité consiste donc à traiter des phénomènes distincts, mais aussi délimités, avec une valeur explicative plus grande et plus effective. Si nous tentons de traiter un phénomène local, par exemple la clitisation dans le syntagme verbal ou encore la subordination dans la phrase, nous approchons ces phénomènes à travers un module qui s'attachera à cette problématique en particulier, sans dispersion. Ce n'est qu'ensuite que l'on pourra intégrer les conclusions de la recherche menée dans un système linguistique plus grand et surtout plus général.

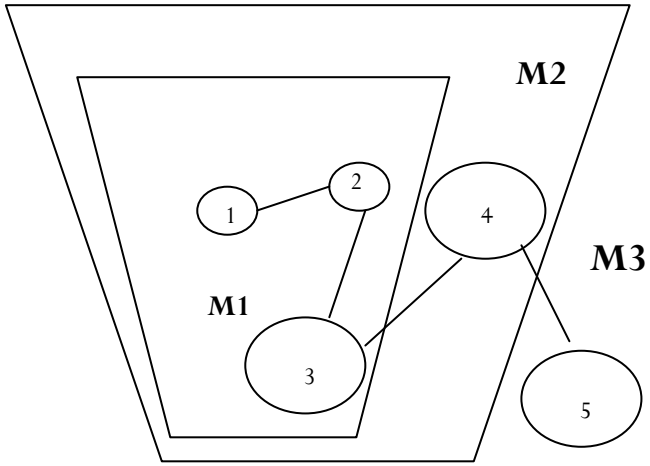
Comment, effectivement, rapprocher la forme du sens ? Question rebattue, poncive, mais à laquelle il est toujours difficile d'apporter une réponse. L'approche phrastique et l'approche textuelle semblent comporter encore les modules les plus porteurs et les plus significatifs pour ce rapprochement, mais la phrase et le texte sont encore des systèmes trop généraux pour permettre des problématisations restreintes. Nølke 1999 et Roulet 1998 parlent ainsi de supermodules, intermédiaires, comme on l'aura compris, des modules restreints et des systèmes étendus, et que l'on peut représenter de la manière suivante :



On entend quelquefois dire que telle ou telle problématique se place *à un autre niveau*, et la démonstration prend tout à coup un caractère bancal. Mettons que quelqu'un ait établi ses conslusions sur un fait linguistique, un autre peut lui déclarer : non, cela ressortit au domaine de la sociolinguistique, ou encore au culturel, au phonologique pur et dur, à la grammaire de phrase. C'est le cas notamment des sujets de recherches que l'on appelle par moments mixtes, dans ce sens où ils partent de phénomènes par exemple purement phonologiques ou syntaxiques pour exemplifier ou appuyer des données d'ordre littéraire, sociologique, philosophique ou autre (ainsi des sujets comme *la phrase verbale dans le dialogue romanesque*, *les dissimilations spontanées chez tel ou tel chroniqueur de radio*, *Snoop Dog et le code switching*).

Il est généralement conseillé de faire appel, au cours des descriptions de faits, à plusieurs composantes qui agissent en interface les unes avec les autres. Dans le domaine du discours, ce principe de compositionnalité, frégéen à la base, rapporte la signification de tel emploi en contexte d'après des données d'ordre lexical et sémantique, là où tel autre exige d'être abordé à travers l'interface du syntaxique et du pragmatique par exemple. Prenons une phrase déclarative à première vue, du type *j'entends que vous partiez* : à partir du moment où l'on ne fait qu'établir l'inventaire des affixes flexionnels verbaux, ou encore la subordination, on demeure dans le champ de la composante lexicale ou syntaxique, et ce n'est qu'en élevant l'appropriété descriptive de la phrase au niveau de la composante pragmatique qu'on établit que cette phrase déclarative constitue en fait une injonction. En affirmant donc qu'il s'agit d'une phrase déclarative qui s'inscrit dans une action verbale, illocutoire, d'injonction, on replace l'approche modulaire de la phrase dans le supermodule des actes illocutoires, ce que l'on peut schématiser linéairement de la manière suivante : *[INJONCTION [PHRASE DECLARATIVE [(prop 1 (sub prop 2)]]]*.

Les modules 1 et 2 correspondent donc ci-dessous aux problématiques des affixes flexionnels et de la subordination, lesquels sont placés dans le supermodule M1 de la phrase, laquelle est recadrée dans la problématique des actes illocutoires (supermodule M2), que l'on a la possibilité de positionner dans une problématique plus générale (supermodule M3), comme par exemple celle du domaine politique :



Ellipse, constructions détachées, connotations verbales sont autant de problématiques qui seront abordées dans des domaines méthodologiques distincts, mais leur approche modulaire permet de discerner quelles sont les intrications des différents niveaux de contraintes, de constructions possibles, de structures de traits que l'on peut leur attribuer. Si nous disons par exemple *elle, elle est heureuse*, cela aura éventuellement des incidences sur la manière dont on sera amené à envisager alors la suite *elle est heureuse* : la réduplication a-t-elle des implications sémantiques ? pragmatiques ? Ces questions appartiennent aussi au traitement modulaire.

Le programme minimaliste chomskyen procède lui aussi par modules, cycliques pour leur part, et sur un modèle dérivationnel. Les contraintes de localité n'étant pas toujours opérables (par exemple quand un NP monte dans la catégorie du T, puis la cat. CP ou IP), notamment quand il s'agit de représenter des relations de traits distantes, ce programme préconise l'existence de cycles à partir desquels la représentation des dérivations syntagmatiques peut s'effectuer. Le système général à partir de quoi vont s'établir les modules descriptifs complémentaires demeure quoi qu'il en soit le niveau syntaxique (les formes P et les formes L appartenant de leur côté à des sous-systèmes, les unités lexicales étant réduites à leurs traits de sélection et à leurs traits phonologiques). C'est pourquoi Jackendoff 1997 a mis en place, en dissidence pourrait-on dire, ce qu'il convient d'appeler la modularité représentationnelle. Celle-ci combine les systèmes phonologique (avec des données prosodiques, syllabiques notamment), syntaxique (à travers les catégorisations, la sélection des nœuds et des constituants adjacents entre autres), et conceptuel (la « forme logique » de Chomsky). Dans ce système, les modules sont reliés les uns aux autres par des règles de Correspondance, qui révèlent autant d'interfaces possibles. Dans la terminologie de Jackendoff, un *événement* se conclut sur plusieurs formalisations possibles (ainsi emploiera-t-on un verbe ou un nom, voire un adjectif, pour désigner un même événement). L'interprétation compositionnelle, qui consiste il est vrai en partie à séparer les modules pour les assembler ensuite et non les prendre comme un ensemble *a priori*, en est elle-même contestée : l'emploi de tel ou tel circonstant dans un syntagme verbal va par exemple avoir une incidence directe sur la signification du verbe, tandis que, positionné à un autre moment de la phrase, il aura une incidence sur une autre unité, comme on le verra ici, en ang. :

a. *He was certainly not earnest.*

b. *He was not certainly earnest.*

Tantôt, ci-dessus, l'adverbe porte sur l'adjectif attributif, tantôt sur le verbe-tête, ce qui aura une incidence, comme on s'en doute, sur la signification d'ensemble de la phrase.

Cette position méthodologique n'est pas tellement en dissidence par rapport au modèle chomskyen, qui dénonce en soi déjà quelques contradictions de ces ordres. Ce que Jackendoff 1997 nous invite à admettre, en revanche, c'est qu'effectivement la composante phonologique a beaucoup plus d'influence sur la signification des unités décrites que le peu qu'on a pu lui attribuer, car elle entre directement en ligne de compte pour ce qui concerne la segmentation de la phrase et donc dans la délimitation des segments dégagés (dans la phrase, le texte) pour en faire la description (Cf. le parsage, ang. *parsing*). Chomsky place la morphologie et la phonologie après le point d'*epel*, après donc les composantes de la syntaxe et de la forme L. Toutefois, on ne se méprendra pas sur les grandes lignes du programme minimaliste, qui n'ambitionne pas de permettre l'interprétation des phrases dans leur intégralité, mais de déterminer les mécanismes qui permettent de les engendrer[9]. C'est pourquoi Chomsky 1998, encore une fois par simplification et par principe d'économie,

[9] Jusqu'il y a peu les chaînes prosodiques ont été considérées comme strictement linéaires, avec une organisation pouvant être abordée dans les termes d'un formalisme syntagmatique simple. La notion de syllabe, par exemple, est absente de l'approche générative « classique ». Dans les compositions dérivationnelles, les représentations phonologiques procèdent d'une organisation unilinéaire, et c'est pourquoi les approches contemporaines de la discipline insistent sur l'architecture des représentations phonologiques et sur l'organisation interne des constituants, au premier abord desquels on peut placer la syllabe. L'approche autosegmentale (Goldsmith 1990) , mais aussi la phonologie computationnelle et la théorie de l'optimalité (Antilla et Cho 1998) suggèrent toutes de prendre la syllabe pour une unité centrale de la description. Par ailleurs, la dimension multilinéaire de l'architecture phonologique (avec les phases, les synchronisations et désynchronisations) est à présent reconnue, avec la syllabe pour organisateur principal (Durand *et alii* 1998).

suggère de favoriser l'opération d'assemblage plutôt que celle du déplacement (c'est le désormais répandu *merge over move*).

La Grammaire Fonctionnelle (*Functional Grammar*, *FG*) de Dik 1997 constitue l'une des grammaires génératives dont l'appareil descriptif et notamment schématique est ouvertement modulaire, tout en ayant pour principal enjeu de cerner les traits sémantiques de la proposition et de la phrase avant même les traits syntaxiques, lesquels sont présentés comme corollaires. Cette approche reprend le système des cas pour typifier les traits spécifiés dans les représentations grammaticales (ainsi les catégories AGENT, PATIENT, EXPERIENCER), mais aussi les caractéristiques théma-rhématiques de la phrase (THEME, RHEME), de même que la terminologie des actions verbales hypéronymiques (ou *speech acts*, actes illocutoires). Dans le format de sa présentation d'ensemble, Dik *op. cit.* ordonne trois supermodules : le plan lexical (LEX), le plan syntaxique (SYN), et celui des règles d'expression (qu'on peut abréger en SEM). Certains linguistes se réclamant de la FG, comme Hannay 1998, ont très vite complété cet ensemble par un supermodule pragmatique, qui découle de la combinaison des précédents. Cela permet notamment de fonder l'emploi de mots comme intenses ou extenses par exemple, mais aussi de constructions présntatives ou passivées entre autres, ainsi que de certains recours complémentaires à l'expressivité.

Ce formalisme favorise l'entrée lexicale à travers l'assignation de traits qui prennent pour d'aucuns la forme de restrictions à l'emploi en termes de dénotation(s), de constructions argumentales et de valence. Par exemple, si l'on prend le verbe TISSER, celui-ci présentera les caractéristiques suivantes :

$$\text{TISSER}_{V}\langle\text{concret, textile}\rangle\ (X_{I})_{PAT}$$

Il apparaît que le terme appartient à la classe des prédicats verbaux (V), et qu'il réclame une position argumentale (X_I), avec pour restrictions les traits indiqués à la suite du verbe (si X désigne par exemple un crayon, il satisfait à la première restriction, mais non à la deuxième : on ne tisse pas un crayon). En outre, le trait PAT (patient) indique que cette unité concernera un référent qui subit l'action de tisser (le trait PROC, lui, n'implique pas forcément qu'il y ait une action et un agent). Dik inscrit ensuite ces restrictions dans le champ des réécritures de sensibilité chomskyenne, à travers le mécanisme d'INPUT / OUTPUT. L'exemple des tournures réflexives avec des verbes pronominaux est à cet égard particulièrement explicite. Si nous disons *Darline se déteste*, nous représenterons, dans le cadre de la FG, la proposition à l'INPUT comme suit :

$$\text{DETESTER}_{V}\ (X_{I})_{AG}\ (X_{J})_{PAT}$$

ce qui donne à l'OUTPUT (dans la phrase concrète et matérialisée) :

$$\text{DETESTER}_{V}\ (X_{I})_{AG}\ (\text{REFL}\ X_{I})_{PAT}$$

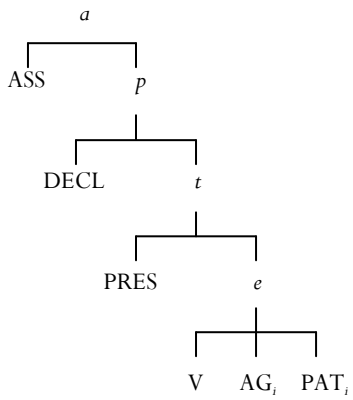
Dans le module de l'expression, l'INPUT $(X_{J})_{PAT}$ sera réécrit par l'OUTPUT $(\text{REFL}\ X_{I})_{PAT}$, et donc matérialisé dans un pronom réfléchi. On imagine la productivité de cette démarche descriptive, qui, avec l'assignation de traits caractéristiques, dégage des restrictions d'emploi que l'on peut renommer des contraintes *à l'emploi*. Ainsi des opérations syntaxiques comme la passivation, la réflexivation, l'adjonction nominale ou adjectivale, les constructions factitives ou attributives sont autant d'occasions pour exemplifier économiquement l'opportunité de ce type de notations. L'INPUT que nous avons reporté ci-dessus ($\text{DETESTER}_{V}\ (X_{I})_{AG}\ (X_{J})_{PAT}$), et dans lequel on glissera si l'on veut des restrictions (telle que <concret> ou encore <individu>, pourquoi pas) ne comporte pas encore d'autres caractéristiques, comme celle du temps verbal ou de l'aspect, et c'est au niveau suivant qu'on les reportera (celui de la prédication noyau, ou *core predication*), avec donc un trait complémentaire :

$$[\ \text{PRES}\ (\text{DETESTER}_{V}\ (X_{I})_{AG}\ (X_{J})_{PAT})]$$

On complètera ainsi la suite au gré des modules représentés (ici les modules LEX et SYN sont les seuls présents) en les hiérarchisant à la droite des notations de manière étendue. Si effectivement cette suite apparaît dans une phrase (Dik emploie indistinctement *clause* pour proposition et pour phrase) déclarative (niveau SYN) et dans un acte d'assertion (niveau EXP, PRAG), cela se résumera en :

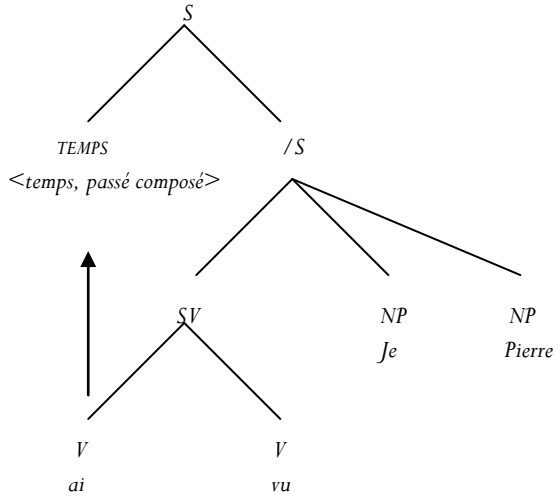
[ASSERTION [DECLAR [PRES (DETESTER $_V$ (X $_I$) $_{AG}$ (X $_J$) $_{PAT}$)]]]

De nombreux opérateurs caractéristiques sont de ce fait envisageables, comme le topique donné (*given topic, GT*), le sous-topique (*sub-topic, ST*), ou encore le focus (*focus, FOC*). Dans ces termes, apparaît un principe qui en somme va de soi, celui que Dik replace dans le contexte d'une *multifunctional theory of constituent ordering*, à savoir que l'ordre des constituants est influencé par un certain nombre de principes fonctionnels qui se combinent et quelquefois entrent en concurrence les uns avec les autres. Or, cela ne remet pas en cause la complémentarité des modules indiqués, que certains auteurs schématisent de manière arborescente, comme suit (nous simplifions au maximum les notations de traits) :



[ASSERTION [DECLAR [PRES (DETESTER $_V$ (X $_I$) $_{AG}$ (X $_J$) $_{PAT}$)]]]

Un autre type de formalisme qui sollicite également une approche modulaire, et dont nous avons brièvement parlé en note, est la *Radical Construction Grammar* de Croft 2001. Celui-ci combine opportunément la schématisation en arbre, le trait SLASH et le binaire <attribut, valeur> propre notamment aux grammaires d'unification. Ci-dessous un exemple de représentation :



Ce schéma, ici très sommaire, montre néanmoins que les modules descriptifs hétérogènes intègrent parfaitement, et sans l'alourdir de trop, une représentation schématique dans laquelle on peut à tout moment spécifier d'autres traits, par exemple les cas ou les positions argumentales en rapport avec la valence verbale.

On n'apprendra à personne que les grammaires non-contextuelles étant insuffisantes pour décrire certains phénomènes linguistiques (Cf Shieber 1986), les grammaires faiblement contextuelles (*mildly context sensitive grammars*, d'après l'expression de Joshi) permettent de pallier un nombre non négligeable de ces apories descriptives. C'est le cas notamment d'une autre approche modulaire : celles des grammaires à concaténation d'intervalles (*RCG,* pour *Range Concatenation Grammars*), qui sur des chaînes verbales, comme le syntagme ou la phrase (ou les deux combinés bien entendu), ne procèdent pas d'une segmentation en termes de sous-chaînes, ainsi que cela se fait dans d'autres types de formalismes, mais assignent à chacun des prédicats

relevés une position, nommée intervalle[10]. Un même mot ou un syntagme analogue qui serait présent plusieurs fois dans une phrase (et donc dans la chaîne) correspond ainsi à un seul intervalle répété. Une RCG dégage donc des *clauses*, qui prennent la forme suivante :

$P > P_1, P_2 \ldots P_n$

P et les P_i qui suivent dans la chaîne linéaire de l'énoncé sont à proprement parler des prédicats, lesquels ont un ou plusieurs arguments. Chacun de leurs arguments, en revanche, est présenté en RCG comme la concaténation d'une ou de plusieurs composantes, d'une ou de plusieurs variables si l'on veut, chacune de ces variables représentant un intervalle de la chaîne. Ainsi aura-t-on $P_1(X)$, $P_2(XY)$, $P_3(X, Y)$, dans lequel le prédicat 1 porte sur l'intervalle X, P2 sur XY, P3 sur le couple X, Y. L'exigence principale des RCG réside dans la notation du maximum d'arguments dans la chaîne (une concaténation d'intervalles), ainsi que dans la détermination des relations marquées que ces formalismes abordent à travers une question de modularité. Par ailleurs, le fait de sélectionner non pas des sous-chaînes non récursives, et abordées une seule fois, mais des intervalles qui dépassent la linéarité même de la chaîne, cela permet de combiner différents points de vue sur une même position. L'atout méthodologique des RCG consiste donc à pouvoir attribuer plusieurs caractéristiques à un même intervalle, comme le type de relation (complémentation, modification, répétition par exemple), de même que les phénomènes d'accord et de constituance (Boullier 2003). De plus, ce type de formalisme rapproche l'analyse syntaxique de caractéristiques plus discursives, lesquelles sont de plus en plus impliquées aujourd'hui dans les notations linguistiques, même les plus schématiques (Avrutin 1999).

[10] La RCG renvoie aussi à un autre formalisme, ou plutôt à une théorie typologique qui nous vient de William Croft (*Radical Construction Grammar : syntactic theory in typological perspective*, 2001, New York, Oxford University Press). Cette grammaire défend notamment le fait que les catégories comme le nom, le verbe et l'adjectif ne sont pas des catégories propres à telle ou telle langue en particulier, mais bel et bien des invariants typologiques, en tant que parties invariables du discours. Celles-ci sont typifiées selon leur possibilité d'avoir certains rôles dans certaines constructions grammaticales, et cela de manière translinguistique (*cross-linguistic categories*, que Croft considère comme des *methodological opportunities*). Cf. le brillant cours de J.A. Bickford, *Morphology and Syntax* (1998), diffusé à Dallas par le *Summer Institute of Linguistics*.

Bibliographie sélective :

Antilla A., Cho Y.M.Y., 1998 : « Variation and Change in optimality Theory », *Lingua* 104 : 31-56.

Avrutin S., 1999 : *Development of the Syntax-Discurse Interface*, Dordrecht, Kluwer Academic Publishers.

Boullier P., 2003 : « Counting with Range Concatenation Grammars », in *Theoretical Computer Science* 293 : 391-416.

Chomsky N., 1998 : *Minimalist Inquiries*, The Framework, MIT Occasional Papers in Linguistics 15, MIT.

Dik S.C., 1997 : *The Theory of Functional Grammar*, Berlin-New York, Mouton de Gruyter.

Durand J., Laks B., Angoujard J.P., 1998 : *Théorie de la syllabe : rythme et qualité*, Paris, CNRS éditions.

Goldsmith J., 1990 : *Autosegmental and metrical Phonology*, Oxford, Blackwell.

Hannay M., 1998 : *The Utterance as unit of description : implications for Functional Grammar*, Université Libre d'Amsterdam.

Jackendoff R., 1997 : *The Architecture of the Language Faculty*, Cambridge Mass., MIT Press.

Nølke H., 1999 : « Linguistique modulaire : principes méthodologiques et applications », *in Approches modulaires, de la langue au discours* (*dir*. H. Nølke et J.M. Adam), Lausanne, Delachaux et Niestlé : 17-74.

Roulet E., 1998 : « Dialogism and Modularity : the topical organization of dialogues », in *Dialoganalyse VI. Referate der 6. Arbeitstagung. Prag 1996* (éds. S. Cmejrova *et alii*), Tubingen, Niemeyer : 49-60.

Evaluations possibles :

- traiter deux phrases en créole atlantique de langue lexificatrice romane dans une approche multimodulaire, et en tâchant de dégager au moins une généralisation de traits ;
- exposer les grandes lignes de la grammaire dite *fonctionnelle* en prenant pour support une ou plusieurs problématique(s) transversale(s).

SEANCE 4 : QUELQUES PROBLEMATIQUES TRANSVERSALES

La schématisation d'une chaîne théma-rhématique en *THEME n [RHEME 1, RHEME 2, RHEME n…]* et en linguistique textuelle dégage un apport descriptif très opportun à partir du moment où il s'agit d'aborder l'organisation du discours. D'un autre côté, l'apparition d'un *thème* ou d'un *rhème* demeure un moment particulier de la chaîne discursive, alors que la thématisation, pour ce qui la concerne, ne peut être ramenée dans tous les cas à l'extraction d'une unité. Par rapport aux thèmes non marqués, seule la dimension textuelle favorise alors une juste compréhension des phénomènes de thématisation, d'autant qu'elle nous conduit à envisager plus concrètement les mécanismes de l'extraction elle-même (Lazard 2001). D'aucuns parlent d'ailleurs à ce sujet de disjonction dans la trame thématique, que d'autres reformulent à travers une question de gestion des ruptures thématiques.

Dans le discours, la verbalisation de ces indications, c'est-à-dire leur matérialisation, dénonce un tournant spécifique. Or, concernant le mécanisme thème-rhème en général, on notera qu'il est assez exceptionnel qu'une macro-taxinomie de ce type fasse autant l'unanimité chez les linguistes, étant avéré que, comme celles d'*actants – circonstants* autour de la notion d'actance, et notamment celle de *valence* pour la transitivité verbale entre autres, ces notions présentent l'attrait tout à fait particulier d'avoir une dimension interdisciplinaire. Au sens linguistique, l'incontournable Blumenthal 1980 montre combien cette combinaison appartient à des faits sporadiques dans le discours, dont la linéarité même demeure un point à débattre (d'autant qu'en grammaire se pose la question de sa multiplicité – et donc de son unicité).

En marge de la problématique des caractères thématisant et rhématiseur des opérateurs de l'organisation du discours, rappelons que ces derniers agissent sur des plans segmental et suprasegmental, en complément desquels un terme, celui de transition, est une voie possible pour explication.

Effectivement, on aura tort de réduire l'expressivité de certains emplois des opérateurs de l'organisation du discours à des données strictement apotaxiques ou diataxiques, bien que les procédés d'extraction et de substitution soient très ordinairement sollicités dans ce domaine (Van Voorst 1988).

Quoi qu'il en soit, on peut estimer que, dans le champ allocutif, les variables thématique / rhématique et leurs versions opératoires thématisation / rhématisation appartiennent tantôt au domaine de l'interlocution (avec des cas d'opérations égocentrées – autocentrées ou non – et allocentrées), tantôt à celui de la délocution (généralement allocentrées). Dans le dialogue, le passage d'un thème à un autre provoque une transition qui s'impose ou ne s'impose pas, et c'est ce qui se passe lorsque l'intervention d'autrui vient perturber le champ allocutif (il est dans ce sens *événementialisé* en discours). Dans tous les cas, nous sommes bel et bien ici confrontés à ce que C. Morris appelait un *comportement manifeste* des co-énonciateurs (Cf. Weinrich 1993 et la notion d'*événement*). Ces transitions concernent bien entendu, par exemple en réponse à une contradiction, ou à une objection que le locuteur se ferait à lui-même ou à quelqu'un d'autre, cette transition thématique dont nous avons parlé. On comprendra qu'à ne prendre la présence de ces opérateurs disséminés entre les segments phrastiques que pour de minuscules procédés oratoires, on méjuge sensiblement la portée, et surtout le vrai rôle de ces emplois instantanés, ne serait-ce par exemple que vis-à-vis des corrections intermédiaires, qui représentent quelquefois bien plus que de petites incidentes explicatives, ou quelques digressions de circonstance. On peut appeler certaines (re)thématisations *égocentrées*, et non hyponymiquement *autocentrées*, dans ce sens où, plus qu'un simple et restrictif recadrage sur *soi*, elles consistent dans une concrète exhibition du *moi* du locuteur. Dans le dialogue verbalisé, et donc en général dans l'interlocution, on verra sans difficulté combien ce genre d'indications linguistiques spontanées chez les locuteurs en présence, et tout à fait instantanées, a persisté dans la question : – *On m'a donné des mangues.* – Et *moi* alors *? je n'ai rien ?*. A ce titre, nous emploierons ainsi le terme de *concaténation*, au sens restrictif, dans le cas où l'opérateur marque une rhématisation des segments dans la chaîne de l'énoncé, et notamment une suite d'actions verbales, voire de comportements verbalisés ou non. Partant de là, l'emploi spontané de l'opérateur participe invariablement d'un comportement verbal spécifique. Le texte comprend donc bel et bien des temps (nous parlons pour notre part plus volontiers de *moments*), qui sont de l'ordre de l'impondérable (et dont la sporadicité et/ou la périodicité sont tout à fait révélatrices en sociolinguistique).

On peut distinguer par ailleurs facilement *types* et *genres* de discours, tout en insistant sur le fait que les énonciateurs recourent à des représentations intertextuelles qui, au-delà du fait qu'elles les invitent à connoter certains moments du discours, les renvoient dans une perspective actionnelle. On rapprochera ces conclusions de celles de Rastier 1998, pour qui le genre permet de relier le texte et son contexte, tout en s'élevant en véritable principe organisateur. Le texte est ainsi une rencontre entre le contexte et l'intertexte.

Sur un autre plan, aux questions de savoir jusqu'où va le cotexte d' « avant » et d' « après », et de déterminer ce qu'il convient de prendre en compte dans le contexte, les réponses ne vont pas de soi. Pour les plus réticents, les termes concernés désigneraient un matériau textuel et extratextuel, linguistique et extralinguistique extensible à l'envi, mais qu'en est-il alors de celui de « phrase » ? La phrase disposerait d'une majuscule « ouvrante » et d'une ponctuation « clôturante », alors que le cotexte et le contexte n'auraient aucune ponctuation caractéristique, aucune « pause » ou aucun « blanc graphique » ? Rien n'est moins sûr a posteriori (Dahlet 2003). D'autre part, où est le linguistique par rapport à ce qui est extralinguistique ? Les réponses sur ce point sont peu satisfaisantes (Rastier 1998, Kleiber 1998).

Quoi qu'il en soit, sur les notions de *texte, cotexte, contexte*, nous admettrons le même principe subsumant de *cohérence*, qu'ont eu l'occasion de commenter G. Guillaume (1992 : *6 déc. 1956*), ainsi qu'Adam 1990 et Weinrich *op. cit.*. En outre, nous remarquons très nettement qu'au moment où l'énonciateur se détermine dans la spontanéité la plus avérée (G. Guillaume ayant justement parlé en ce cas de *momentanéité*), et décide d'apporter à ce qu'il prédique une expressivité certaine, affirmée, son comportement verbal (*illocutoire* si l'on veut) devient *marqué*, et ce de manière spécifique. Tout ceci nous renvoie donc aux *conditions d'ancrage*, qui sont bien exposées dans les *Leçons du 4 juin 1945, série C*, puis du *14 mars 1946*, de G. Guillaume (*op. cit.*). D'autre part, ces conditions « circonstancient » linguistiquement et/ou extra-

linguistiquement l'expression *E*, laquelle renvoie à un contexte donné, de même qu'elle s'insère favorablement dans ce qui constitue ses « avant » et « après » cotextuels (par exemple les termes co-présents avec l'unité).

Ainsi en termes de textualité et de contexte, plusieurs tentatives, plus ou moins opportunes, ont été produites en vue de répondre à la question de savoir si certaines constructions ou dérivations (pour d'aucunes transcatégorielles) correspondent ou non à des universaux. Nous citerons les cas de la Grammaire Relationnelle de Perlmutter, qui à travers une terminologie abstraite a tenté de caractériser la passivation sur un plan translinguistique, de la Grammaire Fonctionnelle de Dik, qui a par exemple abordé cette problématique à l'appui de la transitivité verbale, l'enjeu pour ces formalismes (la linguistique de Culioli est elle aussi formelle dans un sens) étant de dégager des mécanismes invariants (comme l'extraposition, la détermination, les constructions nominatives / accus. et ergatives), ou dans tous les cas d'ouvrir la voie des généralisations. Par exemple, Langacker et sa Grammaire Cognitive (encore un autre formalisme !) se saisit de l'appareil théma-rhématique de l'énoncé par l'intermédiaire des notions de *trajector* (le repéré) et de *landmark* (le repère) : l'énonciateur, de manière générale, s'approprie un repère à partir duquel vont se rassembler, dans la chaîne (multi)linéaire de l'énoncé, un certain nombre, donc, de repérés. On représente ainsi la signification des unités linguistiques en prenant pour supports des schémas qui dérivent les uns des autres, et qui s'appuient sur des notions subsumantes telles que la spécification, la détermination, la modification). Au-delà de ces paramètres, on notera une véritable innovation dans le champ des représentations épilinguistiques et des problématiques transversales : les catégories grammaticales, comme l'adjectif et le verbe, ne sont plus caractérisées à travers des traits descriptifs, mais des opérations abstraites. On sort ainsi du phénomène classificatoire pour déboucher sur des invariants fondamentaux (que rejoint par exemple, pour décrire le rapprochement des faits d'adjonction nominale et de coordination, la *co-jonction* de Rebuschi, ou encore bien avant, concernant le système des temps verbaux, le *point* ou *intervalle de référence* de Reichenbach). Or, ces invariants ont justement une dimension transcatégorielle. Par exemple, si nous prenons un extrait de corpus haïtien :

De bon zanmi sa yo deside yon jou pou yo leve bonè.

les circonstants constituent un appui non négligeable pour la signification de l'opérateur temporel, dont ils intensifient l'acception dans le sens d'une certaine restriction. En termes de structures de traits, les compléments de contenu, et avec eux le contexte, auront donc une incidence certaine sur les données sémantiques du verbe lui-même.

L'un des formalismes les plus à même de traiter des problématiques à ce point transversales est sans doute la grammaire du rôle et de la référence (*RRG*, pour *Role and Reference Grammar*), que l'on complètera éventuellement avec la Grammaire Applicative Universelle (GAU) de Shaumyan 1987. Les composantes de la RRG sont la structure logique, la projection des constituants, celle des opérateurs, et enfin la pragmatique. Ces projections, qui n'ont rien à voir avec celles du module X-barre génératif, sont les composantes descriptives des interfaces qui existent entre le lexique, les constituants en tant que segments syntagmatiques (*templates*), les catégories (comme le temps et la quantification) et les données contextuelles (rôles topical ou focal, notamment). Dans l'énoncé, il s'agit donc de dégager les segments syntagmatiques sur lesquels portent éventuellement des données contextuelles : un cœur (noyau de la structure argumentale) en-deçà de la phrase (*sentence*) et de la proposition (*clause*), lequel se répartit dans le noyau même prédicatif (NUC – PRED) et ses différents arguments (ARG). Il s'agit donc d'une grammaire *distribuée* (et non dérivationnelle, comme celle préconisée dans le programme minimaliste de Chomsky), et dépourvue de mouvements. La RRG ne prévoit aucun mouvement de constituants, mais donne une position principale aux constituants prédicatifs ainsi qu'à la fonction référencielle.

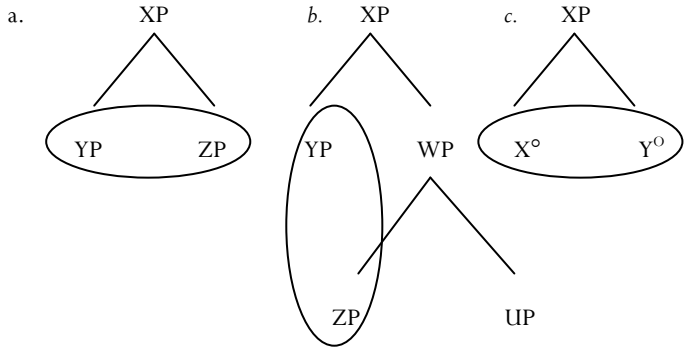
Certaines programmations d'analyse en linguistique, inspirées notamment du programme minimaliste des principes et paramètres, sont capables de traiter des problématiques interlinguistiques à travers une approche modulaire et des modèles d'analyse/étiquetage de corpus définis. C'est le cas notamment du projet IPS genevois (dirigé par Eric Wehrli : *Interactive Parsing System*), qui combine des modules universels et des modules spécifiques, en partant de principes abstraits et de paramètres qui varient selon les langues. Ce qui, dans ce projet, appartient aux modules universels ressortit surtout à la formalisation X-barre, à la règle des projections, celle des chaînes et de l'attachement, ainsi que des cas et des rôles thématiques. Le mécanisme de

formation des chaînes par exemple (A, A-barres, clitiques), complète heureusement celui du X-barre, lequel note le syntagme phrase en tant que *tense phrase* (*TP*) ou *inlfection phrase* (*IP*). Or, le constituant *FP* (*functional phrase*) correspond dans ce modèle à des structures fonctionnelles, comme les propositions réduites (*small clauses*), et l'ensemble, consultable sur http://latl.unige.ch, renvoie à un algorithme de *parsing* qui comprend une analyse lexicale et des unités de segmentation du texte et de la phrase.

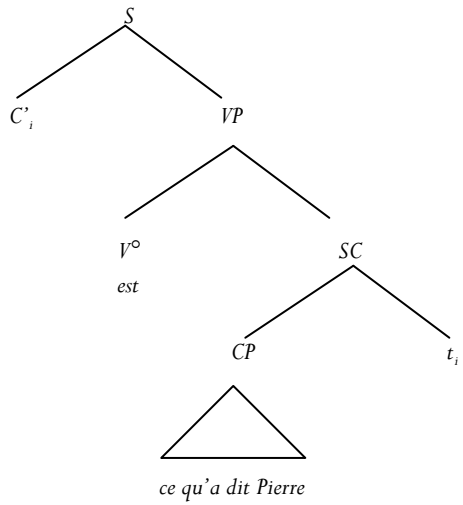
Une autre approche transversale, pour un certain nombre de problématiques linguistiques, consiste à partir d'un fait phonologique, comme par exemple le module TCRS. Ce modèle place la grammaire dans une perspective de phonologie lexicale (strates, interface morphologie / phonologie) autour des contraintes qui s'imposent à une langue en particulier, et tout en considérant la formation des mots comme la première opération de dérivation (Cf. Lacharité et Paradis 1993, Boltanski 2002). Dans ce cadre, il convient de hiérarchiser les contraintes relevées à travers l'organisation de strates descriptives et de caractérisèmes spécifiques, ce qui entraîne un certain ordre d'application. Ainsi les éventuelles violations des principes universels (par exemple l'existence des affriquées ou des voyelles nasales) seront elles-mêmes modulées selon des structures de traits (exemple : + ou – nasal).

On peut alors envisager de partir d'un point de vue *gestaliste*, c'est-à-dire en considérant les objets dans la perception qu'on en a, et, surtout, en partant de l'action de prédiquer. Les unités linguistiques et leurs relations sont ainsi abordées comme un tout unifié, dont on ne peut extraire l'un des deux. La langue est regardée dans sa réalité perceptive, et cela permet de partir de modélisations linguistiques variées pour en tirer des procédures dynamiques tout en insistant sur la synchronisation des opérations accomplies sur les différentes composantes de la linguistique.

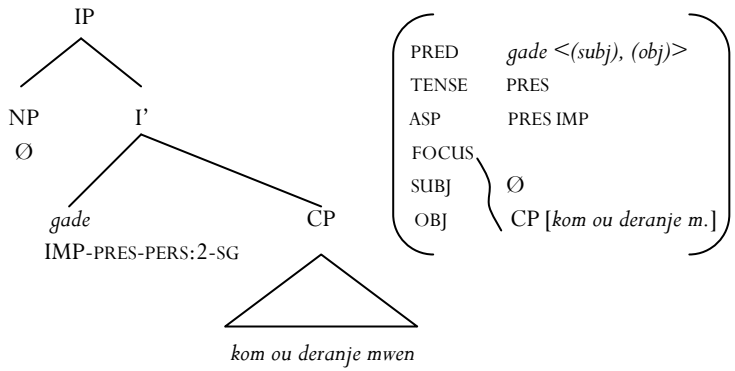
Ces démarches nous mènent notamment vers une approche assez innovante des problématiques transversales dans le domaine du mouvement et des constructions phrastiques : il s'agit de celle qui consiste à traiter de syntaxe en termes d'antisymétrie. Cette méthodologie, kaynienne à la base, revient à généraliser la relation de c-commande à la distinction catégorie/segment et à l'axiome de correspondance linéaire (*LCA*, pour *Linear Correspondance Axiom*). A titre d'illustration, Moro 2000 déclare dans cette vue que le mouvement n'est pas guidé par la morphologie, mais mû par l'obligation de neutraliser un point de symétrie né de deux éléments symétriquement c-commandés l'un l'autre par le fait de linéarisation de l'énoncé. Le mouvement provoque donc une micro-rupture dans la symétrie de la phrase, que l'on peut schématiser des manières suivantes :



Dans certaines constructions par exemple, le point de symétrie constitué par un verbe (et sa projection V° dans le schéma) et d'un *qu*-élément/phrase, requiert que l'élément adjoint le plus distant remonte en IP dans une construction multispécifiée, qu'on notera de la même manière pour les cas de constructions avec extraction thématique d'un constituant :



Dans les approches dérivationnelles, et parmi elles le formalisme génératif, la solution pour traiter ces déplacements consiste à poser une suite linguistique telle que, par exemple, *gade kom ou deranje mwen* en premier temps, et d'énoncer une règle (le *scrambling* par exemple), afin de décrire les éventuels mouvements qui se produisent dans l'énoncé construit tout en s'appuyant sur des traits comme le temps, le cas ou les rôles argumentaux, que le traitement en LFG complète avec une série de spécifications multicatégorielles :

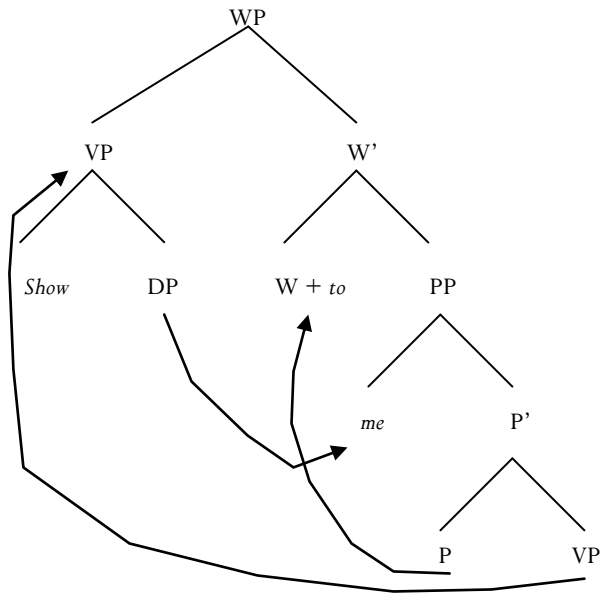


Si les LFG s'en accommodent volontiers, un formalisme tel que celui des RRG rejette, de son côté, le modèle X-barre et sa multi-projection : basé sur une théorie de la structure propositionnelle qui permet de schématiser les relations, notamment argumentales, entre les constituants à travers des branches fortement hiérarchisées et unifiées, au-delà même de toute question de linéarité des constituants (c'est au pivot de faire démarrer l'ensemble des spécifications), celui-ci demeure profondément monostratal.

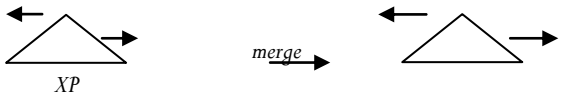
Les dérivations génératives, pour leur part, admettent une influence directe du plan linéaire de l'énoncé et de la problématique de l'ordre des mots dans la phrase qu'elles formalisent à travers les opérations de *merge* et *move*. Un exemple nous en est brillamment donné par Edward P. Stabler, dans son article « Minimalist Grammar and Recognition » (disp. sur le web), lequel montre combien le formalisme génératif s'accommode très explicitement d'un principe de chaîne linéaire de l'énoncé (*chain-based syntax*). Dans la phrase *They showed a picture to me that I like*, on admettra que la proposition subordonnée est adjointe au syntagme nominal *a picture* (*right adjoined to the noun phrase* x), tout en étant extraposée en position finale (*right extraposed to its final position*). En simplifiant tous les autres aspects de la construction, il devient possible de poser la dérivation suivante :

[show me] > merge P
to [show me] > P attracts DP
me$_i$ to [show t_i] > merge W, W attracts P
to$_j$ + W me$_i$ t_j [show t_i] > VP raises to spec, WP
[show t_i]$_k$ to$_j$ + W me$_i$ t_j t_k

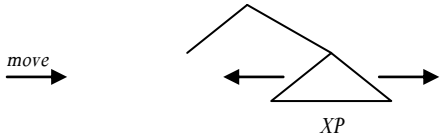
Pour mettre en lumière la linéarisation des constituants de la chaîne dans cette dérivation avec un arbre, on peut ainsi compléter l'ensemble avec des arcs plutôt que de coindexer les unités qui se rapportent les unes aux autres mutuellement :



Ainsi, dans la suite [*a picture*$_i$ [*show* t_i]], on assiste à une fusion (*merge*) du verbe et de son complément, que l'on peut aussi combiner avec l'autre argument complément (*to me*), en *merge V, DP* ou bien *merge W, DP*, à la suite de quoi l'on déterminera quels arguments conditionnent d'autres arguments (ou plutôt les attirent : *attract*), ou encore lesquels sont éventuellement extraposés. La chaîne est alors envisagée comme une suite de liens (des arbres en général), parmi lesquels l'ordre des mots peut être représenté par des arcs. Ce que l'opération *merge* schématise en somme, c'est l'existence d'une paire de structures : quand deux chaînes sont combinées, il en résulte une suite de liens qui sont entièrement connectés dans ce sens où chaque unité est reliée à une (ou plusieurs) autre(s) par un trait de dominance, précédence linéaire ou l'inverse. Schématiquement :



L'opération *move*, pour sa part, demeure unaire, et consacre le fait qu'en se liant l'un à l'autre, un nœud se substitue à lui et le subsume :



De cette manière avec *merge*, chaque phase de la précédence dans la chaîne et chaque relation (éventuellement représentée par un arc) renvoie aux différentes positions qu'occupent les constituants, ce qui la rend descriptivement plus transversale et plus modulable que l'opération *move*. De même, et comme dans le programme PROLOG, les contraintes spécifiées n'ont pas besoin d'être ordonnées : combinées, elles rapportent un état de fait à travers des relations caractérisables de plusieurs manières (ceci dit, contrairement au PROLOG, qui détermine une suite de contraintes et qui se présente donc en cela comme déclaratif, les MERGE et MOVE génératifs sont basés sur des règles). On notera néanmoins qu'à la base, le programme minimaliste a fait la tentative de plusieurs initiatives d'unification de ces règles, à travers avant tout DEPLACER NP et DEPLACER *QU-*. Cet essai d'unification portera d'abord sur les contraintes de localité. De son côté, le modèle X-barre reprend ces tentatives en généralisant l'hypothèse suivant laquelle tous les syntagmes sont ordonnés par rapport à une tête (ils sont endocentriques), ce qui conforte la représentation au moins supposée d'une relation de subjacence entre les constituants, laquelle s'applique aux mouvements (on sent là toute l'influence de l'ouvrage de *Barriers* : une barrière et une seule est passée à chaque phase de l'opération, quelle qu'elle soit, ce qui représente une contrainte représentationnelle imposante, mais surtout une véritable hiérarchisation des constituants). Au niveau d'Epel (*spell-out*), c'est-à-dire au niveau des unités effectivement instanciées, prédiquées (celui de S.S.), la théorie de la *copie* de DEPLACER X, les unités impliquées (qui ne coïncident pas forcément avec des unités lexicales entières) sont spécifiées extérieurement aux unités discrètes : on prend en compte des opérateurs (le temps, la quantification par exemple) comme des sous-ensembles signifiants. Dans la phrase *Lequel*$_i$ *des deux ouvrages a pris Evelyne i ?*, l'opérateur *i* est copié avant le mouvement de montée dans le complémenteur extérieurement au noyau verbal *prendre* et au noyau nominal *ouvrages*. On peut estimer bien entendu que l'opérateur *i* n'est repris que dans l'opérateur *qu* de *lequel*, mais par un fait d'intégrité phonique, c'est l'ensemble du complémenteur et de ses arguments éventuels (ici *des deux ouvrages*) qui est entraîné[11].

[11] On parlera à cette occasion de *pied-pied* (il s'agit du *pied-piping*) : seuls les traits de *quel* ont été déplacés en fait, mais la suite (*des deux ouvrages*) est envisagée *in situ* (dans son site d'origine).

Bibliographie sélective :

Adam J.M.., 1990 : *Eléments de linguistique textuelle*, Liège, Pierre Mardaga éditeur.

Blumenthal P., 1980 : *La Syntaxe du message*, Tubingen, Niemeyer, Max Verlag (180) : 113-120.

Boltanski J.E., : *Nouvelles Directions en phonologie*, Paris, Presses Universitaires de France (coll. Linguistique Nouvelle).

Dahlet V., 2003 : *Ponctuation et Enonciation*, Ibis Rouge éditeur, *Presses Universitaires Créoles*.

Fontanille J., Zilberberg C., 1998 : *Tension et Signification*, Paris, Pierre Mardaga éditeur.

Guillaume G., 1992, in Valin R., Hirtle W., Joly A. *(eds)*, 1992 : *Leçons de linguistique de Gustave Guillaume*, Laval/Lille, PU.

Kleiber G., 1998 : « Dimensions du contexte : écrit *vs* oral », in *Analyse linguistique et approches de l'oral, Orbis Supplementa* 10, Leuven, Peeters : 123-134.

Lacharité D., Paradis C., 1993 : « The Emergence of constraints in generative phonology and a comparison of three current constraint-based models », *Canadian Journal of linguistics* 38-2.

Lazard G., 2001 : « Thème, Rhème : qu'est-ce à dire ? », in *Etudes de Linguistique générale*, Paris – Louvain, Peeters : 79-90.

Moro A., 2000 : *Dynamic Antisymmetry*, Cambridge, Massachussets, MIT Press.

Shaumyan S., 1987 : *A semiotic Theory of Language*, Bloomington, University of Bloomington Press.

Van Voorst J.G., 1988 : « Thematic Roles are not semantic roles », *Revue Québécoise de linguistique* 17, 1 : 245-259.

Weinrich H., 1993 : *Textgrammatik der deutschen sprache*, Mannheim, Dudenverlag.

Evaluations possibles :

- produire une bibliographie critique d'un minimum de dix entrées sur la Grammaire Relationnelle de Perlmutter ;
- démontrer la productivité de l'approche théma-rhématique de l'énoncé à l'appui d'une exemplification créolophone.

SEANCE 5 :
LA PROFUSION DES METHODOLOGIES DESCRIPTIVES

On notera que depuis les années 1990, les schémas se sont pour beaucoup substitués aux descriptions linéaires et aux formules notamment booléennes. C'est donc une certaine vision du traitement cognitif des unités qu'a bousculée la généralisation du phénomène d'assignation de traits caractéristiques, lesquels sont en général représentés en réseau. Dans le cadre des formalismes syntagmatiques déclaratifs, la première exigence, au-delà même de celle de la représentativité des constructions envisagées, consiste dans la productivité des généralisations des caractéristiques formelles que l'on attribue à tel ou tel énoncé (ou type d'énoncés). Les grammaires syntagmatiques énumèrent donc des constructions typifiées à travers des structures de traits (Cf. les HPSG), ou des arbres (ainsi, les TAG). Très vite, il en est ressorti qu'il est indispensable de hiérarchiser les indications portées sur les représentations linguistiques de ces constructions, et de se préoccuper de la manière de combiner les caractéristiques dégagées afin qu'elles héritent les unes des autres[12].

Parmi les formalisations envisageables en linguistique, laquelle n'est bien évidemment pas que descriptive (elle se présente aussi comme prédictive et comme explicative), des auteurs ont mené des approches dans une vision récursive et modulaire, dans le cadre de ce qu'on appelle, nous l'avons vu, les grammaires d'unification[13]. Dans ce cas, ce n'est plus la phrase qui occupe la place de l'unité principale de description des faits, mais les syntagmes eux-mêmes et toutes les composantes qui interviennent dans leur emploi, en plus des relations qu'ils entretiennent entre eux. Or, que cette démarche nous conduise vers des représentations schématiques, qui sont légion dans la discipline, ou vers des paraphrases linéaires, ce type de formalisations présente de grandes opportunités : les grammaires syntagmatiques guidées par la tête (HPSG) et les grammaires d'adjonction d'arbres (TAG) permettent de résumer les constructions sur les plans de la combinaison des arguments, mais aussi de leur composition (comme les SUBCAT, ou sous-catégorisations). En HPSG, ces sous-catégorisations indiquent par exemple la distribution des accords verbaux, les phénomènes d'auxiliarisation, d'attraction temporelle ou encore de phoricité. Dans ce cas, il suffit de sélectionner une catégorie (SN, SV notamment) et de lui adjoindre des caractéristiques graphiques (qui peuvent poser difficulté dans le cas des créoles), phonologiques (comme l'élision, la contraction, la dissimilation ou l'assimilation), mais aussi prosodiques (ainsi la protase et l'apodose, l'intrusion, la réduplication), et bien entendu argumentales. Ces dernières, qu'on appellera des propriétés combinatoires, forment l'une des indications principales des HPSG, qui intègrent aussi bien les composantes catégorielle que prosodique ou phonologique. L'apport majeur de ce type de formalisations revient au fait qu'elles nous invitent à délimiter les constituants syntagmatiques dans la phrase à travers la segmentation, mais aussi les contraintes qui pèsent sur leur emploi et sur leur position. Pour que celles-ci conservent une représentativité satisfaisante, les sous-catégorisations renvoient à des constructions argumentales rapprochées (locales), mais n'excluent pas pour autant les faits d'extraction ou de coordination de plusieurs à plusieurs, pour ne prendre que ces exemples.

Pour des cas comme *pou ki es ou cheche kay la ?*, la modélisation en HPSG va abstraire la représentation des arguments qui, bien que distants, sont adjacents ou « frères », pour laisser le champ à une description abordable, comme elle le ferait de relations moins distantes, et c'est là qu'intervient le mécanisme *slash*.

Prenons une catégorie majeure, comme le SV ou le SN. A l'intérieur d'un syntagme du type *Pèp ayisyen an se yon pèp ki pa gen pyès dwa*, au-delà de la

---

[12] Un bel exemple, à ce titre, de grammaire « électronique » (autrement dit paramétrable) nous vient d'Abeillé 2002. Celle-ci prend plus particulièrement appui sur les TAG, et dans le cadre des grammaires d'unification (*unification-based formalisms*).

[13] La linguistique est descriptive dans ce sens où elle permet de décrire et de représenter des faits, comme des constructions syntagmatiques ou des déplacements dans une chaîne linéaire de SV par exemple ; elle se révèle explicative quand elle consiste à rapporter ces faits dans une problématique plus générale ou des schémas continus, voire dans le cadre de phénomènes cognitifs ou culturels ; elle se présente aussi comme prédictive, en ceci qu'elle nous conduit à aborder des faits qui n'ont été ni décrits ni expliqués, voire des problématiques qui jusqu'ici n'ont pas été envisagées.

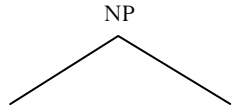
construction attributive, les HPSG nous incitent à admettre que le SN *Pèp ayisyen* et le SN *pèp ki pa gen pyès dwa* sont tous deux sous la portée du verbe attributif ; de même, et si l'on schématise le deuxième SN de la même manière, on admettra que le SV *ki pa gen pyès dwa* apparaît sous la portée argumentale du SN *yon pèp* (et plus particulièrement le nom *pèp*). On compose alors les syntagmes ainsi dominés en *n-uplets* sous la portée d'une tête, à partir de laquelle le mécanisme *slash* distribue donc les sous-catégorisations. C'est pourquoi il devient tout à fait possible d'envisager les relations de dominance argumentale à distance comme on le fait des caractéristiques combinatoires, et d'assigner aux constituants syntagmatiques la marque la plus concrète possible des relations d'opérateur à argument(s).

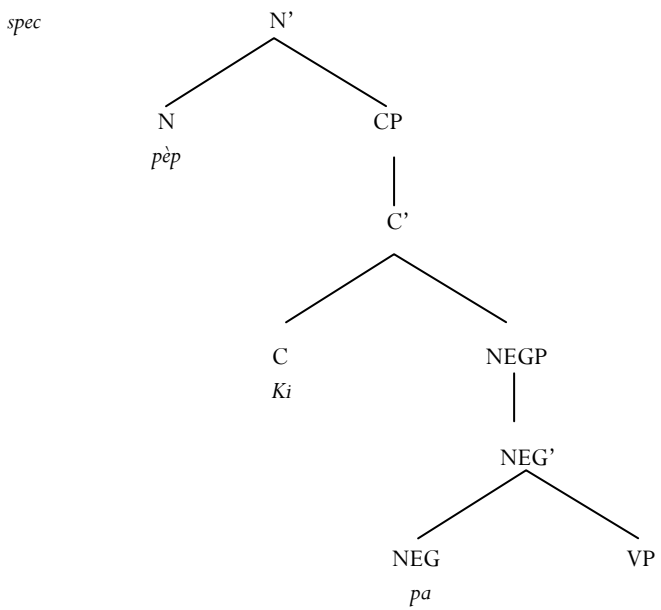
On partira donc au minimum un trait commun (catégoriel – deux SN par exemple –, mais aussi d'autres ordres), ce qui fait qu'en premier temps apparaîtra la sous-catégorisation commune aux deux (ou plus) constituants, après quoi viendront les sous-catégorisations distinctes entre constituants.

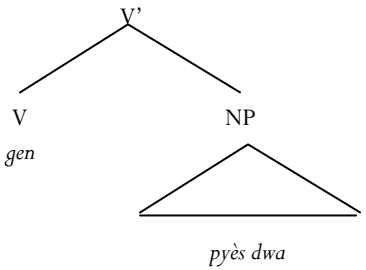
En cas d'ellipse du verbe par exemple, le mécanisme *slash* renverra à la (sous-)catégorisation commune aux constituants, et c'est dans un même esprit que l'arbre en formalisation TAG, que nous aborderons dans quelques lignes, représentera autant de structures argumentales qu'il y a de SV effectifs, qu'ils soient présents ou non. Ce qui peut éventuellement compliquer les représentations, pour les HPSG, concerne une certaine profusion des sous-catégorisations, qui résistent aux schémas trop sommaires, et pour les TAG, la possibilité qu'il y ait trop de dérivations indiquées.

Mais avant de revenir sur ces formalismes, notons que, lui aussi, le programme minimaliste de sensibilité chomskyenne procède bien par l'intermédiaire de niveaux et de relations entre les constituants et les têtes (par exemple les relations modifieur-modifié, spécifieur-spécifié). Le constituant supérieur (au plan argumental), est alors indiqué à travers la relation de c-commande, à laquelle se substitue chez certains auteurs celle de o-commande (une relation d'oblicité entre les arguments de la tête). De ce fait, cette modélisation dégage des dérivations, qui relèvent surtout des grammaires syntagmatiques, mais en restreignant la part des sous-catégorisations, et en insistant sur les positions argumentales des unités syntagmatiques par rapport à la tête. Or, comme en HPSG, les catégories majeures peuvent être fonctionnelles : ainsi l'inflexion, le temps ou la négation (avec leurs projections maximales *IP*, *TP* et *NegP*).

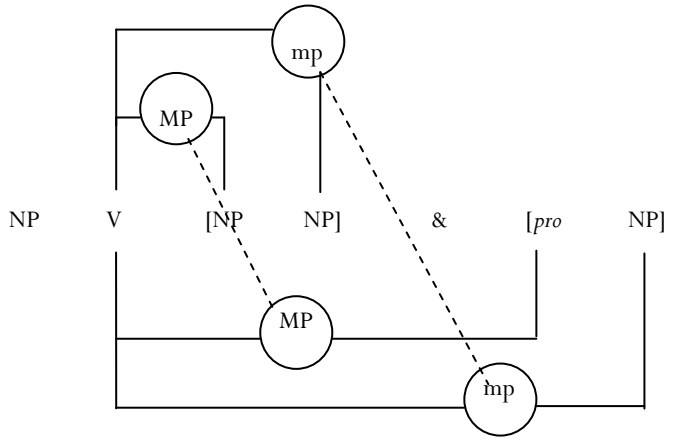
La problématique générale du *merge over move*, comme nous l'avons vu, permet de représenter schématiquement les relations des arguments à la tête en partant d'une projection de celle-ci en plusieurs temps, ce qui fait qu'une unité a donc la possibilité de devenir la tête d'une projection fonctionnelle. Suivant le schéma *XP : ZP – X' / X': X – YP*, il devient envisageable de la traiter en termes de projections maximale, intermédiaire et minimale. Pour la suite *pèp ki pa gen pyès dwa*, cela donne, sous forme linéaire sommaire [NP *pèp* $_{CP}$[C [*ki* $_{VP}$[*pa gen* $_{NP}$[*pyès dwa*]]]]], et, sous forme arborescente en termes de projections :







D'autres schématisations, plus ou moins algorithmiques, permettent elles aussi de dégager le réseau de relations qui existent entre les arguments. Par exemple dans le RM (*relational modelling*) suggéré par Sleator et Temperley 1993, les constituants forment des cellules (*cells*) qui possèdent des nœuds multiples (*multiple nodes*). S'ensuit toute une typologie possible de relations. Ainsi Grootjen, Kamphuis et Sarbo 1998 suggèrent de noter la relation entre la tête et les arguments contraints la *sous-catégorisation*, celle qu'elle entretient avec les arguments non contraints la *modification*, et enfin celle qui existe entre le spécifieur et la tête la *spécification*. Dans une suite coordonnée par exemple, comme *Pol esplike òganizasyon Fakilte ak Inivèsite a* :

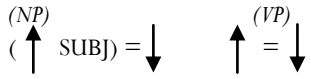


Cette schématisation nous invite à envisager plusieurs types de relations entre les constituants ainsi coordonnés, mais aussi plusieurs nœuds dans ces derniers. En tant que réseau de relations, elle permet aussi de représenter économiquement la coordination de plusieurs à plusieurs tout en allègeant le schéma des parties des constituants qui ne sont pas à mettre en relation ensemble. On assiste alors à un mouvement (*move*) qui se combine avec une fusion des caractéristiques mises en valeur par le schéma (*merge*).

Cela étant dit, rappelons qu'en dissidence quasi-complète avec les grammaires syntagmatiques de sensibilité chomskyenne (réformes incluses), les grammaires guidées par la tête (*Head-Driven Phrase Structure Grammars*, *HPSG* : Pollard et Sag 1994) tentent à proprement parler une unification des grammaires d'unification. Cela paraîtra éventuellement un peu présomptueux, mais il convient d'admettre que depuis les années 1990, les HPSG ont fait la démonstration de leur productivité descriptive. Les GPSG,

les FUG (pour *Functional Unification Grammars*) et les CG (pour *Categorial Grammars*) se voient donc rassemblées dans le cadre d'une approche méthodologique uniformisée, dans laquelle il est possible de représenter tous types d'unités (qu'il s'agisse de syntagmes ou de phrases, de mots ou de moments discursifs) par des caractéristiques analogues, et surtout dans un même format schématique. Ces grammaires intègrent donc ce qui relève de tous les champs descriptifs supermodulaires, comme le phonologique, le lexical et le sémantique. En tant que telles, celles-ci s'affirment dans la transversalité et la modularité bien sûr, mais aussi dans la multidimensionalité (dans ce sens, elles constituent des grammaires dites de construction). L'autre innovation de ces formalismes renvoie au fait que les traits spécifiés rentrent dans une typologie généraliste limitée : ainsi les traits syntaxiques et les traits sémantiques sont-ils regroupés dans la catégorie SYNSEM (avec, pour les traits SYN, les traits locaux ou LOC), et les constituants sont-ils désignés autour de la tête (HEAD) comme des filles (*daughters*, ou DTRS : voir la séance 1).

Les Grammaires d'Arbres Polychromes (GAP), de leur côté, appartiennent au courant des grammaires syntagmatiques généralisées, et permettent de dégager une représentation monostratale du module syntaxique tout en conservant une compatibilité constante avec une représentation multidimensionnelle des autres modules linguistiques. Comme l'ensemble des grammaires syntagmatiques, qui regroupent les constituants autour d'une tête LEX, les GAP parlent de positions entretenues par rapport à un pivot. Une position est indiquée par une variable généralement binaire (un diagramme) qui contient donc deux caractéristiques principales qui peuvent concerner toutes les dimensions modulaires de l'appareil linguistique. Dans la phrase *Katty boit du thé*, la variable *Katty* occupe ainsi une position que l'on peut caractériser à l'appui de données catégorielle et argumentale (ainsi <nom, *sujet*>), ou bien argumentale et pragmatique (<sujet, *focus*>), ou encore catégorielle et théma-rhématique (<nom, *thème*>), et ainsi de suite. Quand les diagrammes ne donnent pas d'indications sur le rôle argumental du syntagme (c'est le cas dans le dernier exemple), ce type de formalisme reprend la manière dont procèdent les grammaires lexicales et fonctionnelles, lesquelles mettent en place un niveau de description complémentaire (la c-structure) qui permet de spécifier formellement les types de relations qu'entretiennent les constituants entre eux :



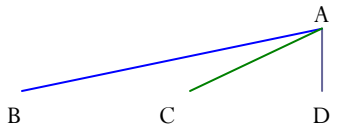
On montre alors dans quelle mesure le sujet (*NP*) a une incidence sur le syntagme verbal (*VP*), en sus de quoi les GAP assignent des caractéristiques de tous ordres dans leur terminologie des positions. Le VP en question dispose ici d'un spécifieur <sujet> (<*fille, sujet*>) et d'une position V <verbe> (<*fille-tête*>), avec d'éventuels compléments (<*fille(s) complément(s)*>). On assite donc à un réseau de relations auquel on assigne des traits qui ne correspondent pas forcément aux catégories grammaticales ou syntagmatiques, comme en LFG, ou à des rôles argumentaux, comme en HPSG, et des traits tels que *thème-rhème* ou encore *personne* ou *accord*. Une position <P, *x*> accueillera ainsi, avec pour rôle argumental sujet par exemple, soit un SN, soit un SV ou autre, notamment dans le cas d'une (pro)nominalisation :

a. Construire un bâtiment ici me paraît difficile.
b. La construction d'un bâtiment ici me paraît difficile.
c. Un bâtiment ici ? Cela me paraît difficile.

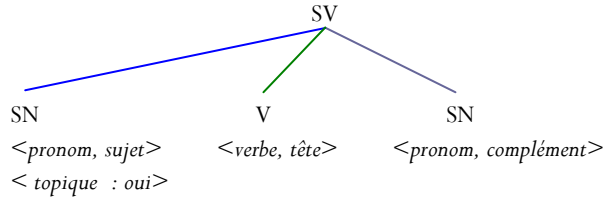
Formellement, les arbres polychromes sont des arbres syntagmatiques classiques (Cori et Marandin 1998), mais ils comprennent une dimension complémentaire, qu'on appelle la *couleur* par analogie. Ces *couleurs* sont là pour distinguer plusieurs types de relations sans solliciter les catégories syntagmatiques ni les rôles argumentaux des constituants. Dans ces termes, la position tête (ou pivot) fusionne dans l'arbre avec les positions plus ou moins proches des constituants qui s'y rapportent. Très schématiquement, mettons un premier arbre, avec deux positions dont l'une (A) est pivot de l'ensemble :

Admettons à présent que le pivot A domine également les positions C et D :

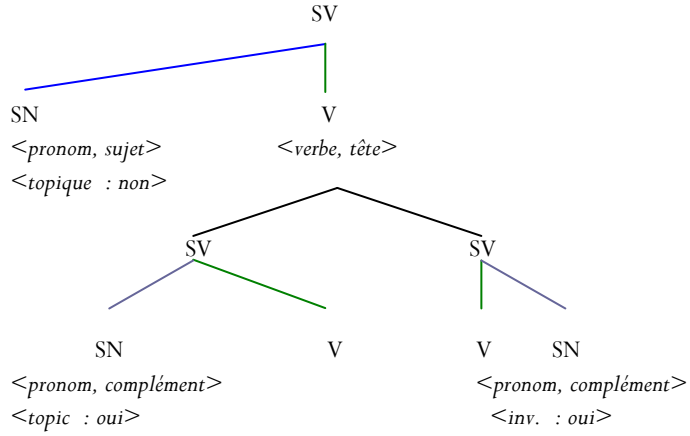
En fusionnant aussi les deux arbres représentés, on compose un arbre polychrome qui compacte toutes les positions :



Dans une phrase qui contiendrait un syntagme "à trou" (*gap*), on peut ainsi assigner une position où aucun constituant ne serait instancié (par exemple, quelqu'un qui, à la question *tu as vu ce saccage ?*, répondrait *j'ai vu* à la place de *je l'ai vu* ou encore *j'ai vu cela*). Une possibilité consisterait à dégager un constituant qui, même à distance, comporterait tous les traits du constituant effacé ou déplacé (on le désigne souvent sous le dénommé *filler*), mais les GAP possèdent d'autres types de notations, comme celle de *focus* quand c'est le cas (ainsi dans *ça, pour voir, j'ai vu*, où le focus porte sur la prédication verbale[14] , ou *Ça, je l'ai vu*, où le focus porte sur une prédication toute entière (*all focus*)). Le fait de noter les positions à la fois des constituants instanciés et non instanciés (*marked* et *unmarked constituents*) permet de dégager les cas d'extraction, d'inversion (dans les « phrases » interrogatives ou les inversions locatives par exemple [15]), et notamment de topicalisation (*cela, je l'ai vu / cela, j'ai vu*) en hiérarchisant les arbres, dans lesquels les relations de dominance sont caractérisées par des *couleurs* qui se répètent et qui permettent de relever les relations établies à distance, chacune des positions étant sous-spécifiée selon ce qu'on entend représenter. Si l'on veut dégager une topicalisation sur la suite *j'ai vu cela*, schématisée de la manière suivante (en spécifiant ici les catégories syntagmatiques et les rôles argumentaux des constituants), cel a donne :



en *cela, j'ai vu*, la composition des arbres donnera (nous simplifions) :



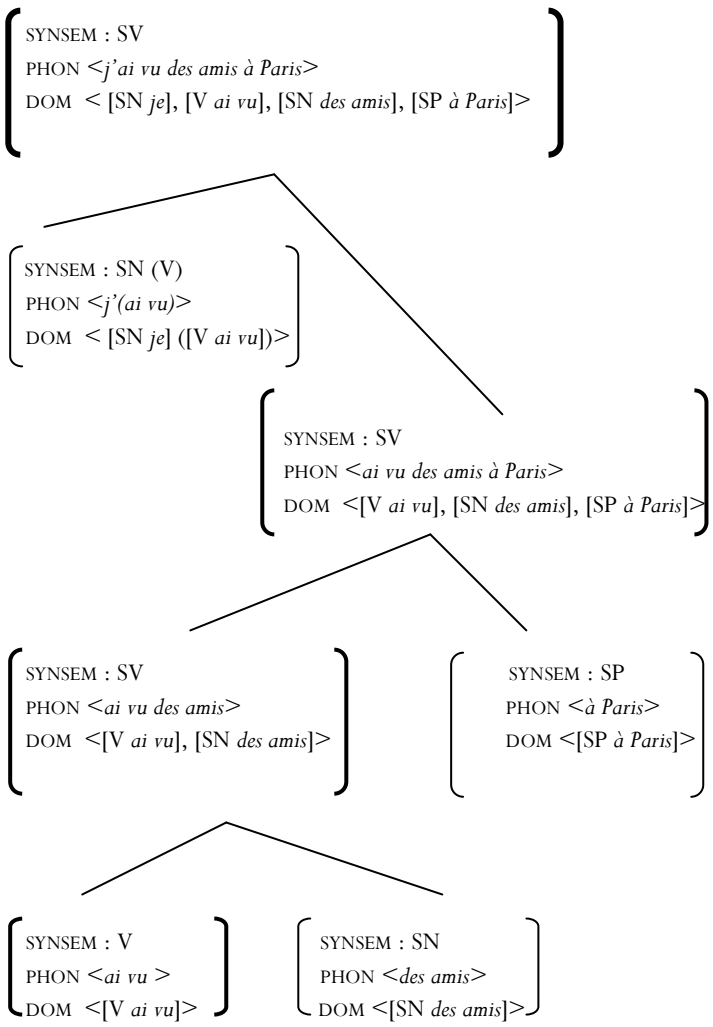
Ce schéma est – insistons là-dessus – tout à fait sommaire, mais il donne l'esprit général d'une représentation en plusieurs temps qui, à travers une fusion des arbres, nous invite à chacun des niveaux de sous-spécifier les traits caractéristiques de la construction tout en conservant les mêmes traits de dominance argumentale. Certains linguistes linéarisent volontiers ces spécifications (on parle ainsi par moments de grammaires de linéarisation

[14] Il s'agit alors d'un *predicate focus.*

[15] Avec respectivement pour ill. *qu'as-tu vu à Paris ?* et *A Paris, tu as vu quoi ?*

directe par exemple), tout en combinant éventuellement ce type de représentation avec une composition d'arbres. Ainsi les grammaires syntagmatiques classiques posent que la composition même influence au moins, si ce n'est détermine complètement la linéarisation des constituants. Ce présupposé descriptif a une incidence directe sur la problématique de l'ordre des mots. On peut de cette manière noter plusieurs ordres possibles dans la linéarisation (ainsi *j'ai vu des amis à Paris*, soit *SV : SN V SN SP / A Paris, j'ai vu des amis*, soit *SV : SP SN V SN*), alors qu'une seule composition s'avère envisageable : l'ordre est plus ou moins contraint, notamment par les règles de dominance immédiate (DI) et de précédence linéaire (PL)[16], mais une et une seule composition est admissible. Dans le formalisme des GAP (GAPH pour les grammaires d'arbres polychromes hiérarchisées), l'analyse syntaxique (*parsing*, ou parsage) est donc fondée sur le fait que les syntagmes sont constitués d'une série de positions ordonnées entre elles. Les GAP réagencent ainsi les branches de l'arbre de manière à montrer que l'ordre des constituants dans le syntagme demeure libre, et que les positions ne coïncident pas forcément avec des catégories syntagmatiques ou des rôles argumentaux. De même, ces positions ne renvoient pas toujours à des unités effectivement instanciées.

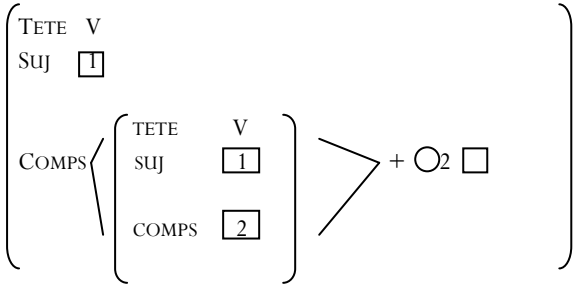
En HPSG, en revanche, le formalisme des grammaires de linéarisation directe occasionne deux représentations distinctes du syntagme : en plus d'une structure en constituants éventuellement hiérarchisée, on note un *domaine d'ordre* (Kathol 1995) qui spécifie la représentation linéaire des unités. Par exemple, une suite telle que *j'ai vu des amis à Paris* donnera ceci :



On comprend, à partir du deuxième niveau, et plus explicitement encore des niveaux suivants de l'arbre, que l'ordre des constituants est plus ou moins libre, et que, dans ce sens, il est possible de sous-spécifier des contraintes qui

[16] Les grammaires dites DI/PL fonctionnent principalement sur celles-ci.

pèsent sur lui. Ce peut être justement le cas d'arguments contraints du verbe (Cf. les verbes transitifs directs), qu'il convient alors de noter par une numérotation intégrée dans le cadre de plusieurs notations qui vont combiner la composition des arguments (*argument composition*) avec les contraintes qui s'exercent sur l'ordre même des arguments (lesquels seront spécifiés en 1, 2, … *n*). Prenons un SV assez simple, comme *je veux qu'elle vienne*, les arguments seront numérotés comme suit :

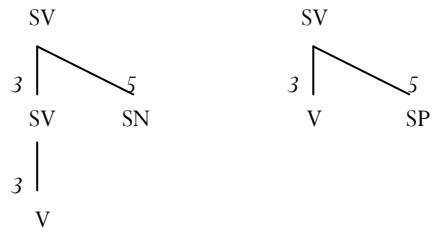


Ce schéma de dominance de la tête à son spécifieur et à ses compléments, lesquels peuvent être concaténés à travers la notation + entourée, indique l'ordre des constituants contraints par le biais de la numérotation. Or, là où les grammaires DI/PL peinent à formaliser les contraintes qui pèsent sur l'ordre des constituants, les GAP et les GLD, l'une par l'intermédiaire de la représentation de pivots, l'autre par celui des spécifications numérotées, sont beaucoup plus explicites sur ce thème. Ainsi, certains cas d'extraposition (par exemple *ceux-là sont venus qui sont mes amis*) seront dégagés suivant les deux approches, dont on indiquera que les GAP ramèneront cette problématique sur le terrain du pivot, laissant par là-même moins de marge descriptive à cette éventualité de modification de linéarisation des constituants.

Dans le formalisme GAP, il convient donc de distinguer explicitement une position de ce qui occupe effectivement cette position. Représentée par une couleur, la position dégage une relation syntagmatique en unifiant (ce que les grammaires dites d'unification ne sont pas toutes en mesure de faire) les constituants de catégories et de constructions distinctes les unes des autres dans un réseau de relations typées, nous l'avons vu, mais en posant que les structures élémentaires d'une langue forment des classes (c'est en ceci que ces grammaires sont hiérarchisées entre autres), dans le cadre donc d'une approche qui procède notamment par généralisation de traits (Lablanche 2003). Autour de la position noyau, qui tient le rôle de pivot de la construction, les constituants de la structure occupent d'autres positions qui, au-delà même des catégories grammaticales auxquelles on peut les apparier, se placent dans un ordre organisé parmi cinq couleurs (la couleur 3 correspond à la position de pivot, la 5 à celle de complément, la 1 à celle de sujet en général, et les 2 et 4 à des positions plus spécifiques, comme celle des constituants incidents). Ainsi par exemple, un SV contenant notamment un verbe et deux compléments verbaux (*j'ai donné la balle à Martin*) donnera ceci sous forme linéaire :

SV : 3 V
SV : 3 SV 5 SN
SV : 3 SV 5 SP

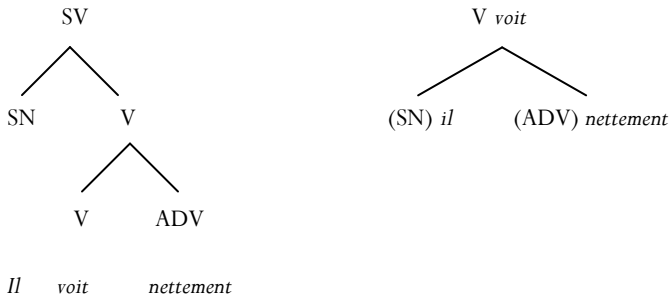
Et, sous forme d'arbres :



L'organisation hiérarchique consiste à spécifier ensuite des traits caractéristiques des positions relevées, après avoir indiqué si celles-ci sont occupées effectivement par un constituant instancié, ou non. La relation d'héritage permet de transmettre les traits spécifiés d'arbre en arbre (ceux qui sont dominés héritent donc des propriétés auparavant dégagées).

Dans le cadre d'un rapprochement évident entre les GAP d'un côté, et les LFG de l'autre, on notera qu'il existe une possibilité de combiner les deux

formalismes dans une méthodologie multiparamétrée et dans une approche modulaire (Candito 1999). Cette grammaire d'arbres adjoints (TAG) lexicalisée reçoit ainsi des spécifications lexicales tout en fusionnant les arbres par adjonction ou substitution. Le principal enjeu de ce type de formalisme est qu'il combine les instructions paradigmatiques et les instructions syntagmatiques, en regroupant les constructions suivant qu'elles renvoient à des opérations syntagmatiques prédéfinies, comme la passivation ou la négativation. D'autre part, une notation minime est laissée aux mots dépourvus de prédicativité (non prédicatifs), comme la plupart des prépositions ou des complémenteurs (ils deviennent alors des « co-têtes » de la tête lexicale qui les implique dans l'arbre). Par exemple pour les noms, on spécifiera s'ils sont déterminés, quels sont leur genre, personne et nombre, et pour les verbes, on distinguera ceux qui transitent vers un régime (prépositionnel ou non) et ceux dont ce n'est pas le cas. Sommairement, voici deux matérialisations possibles, dont la deuxième seulement coïncide avec cette approche généraliste (pour le syntagme *il voit nettement*) :



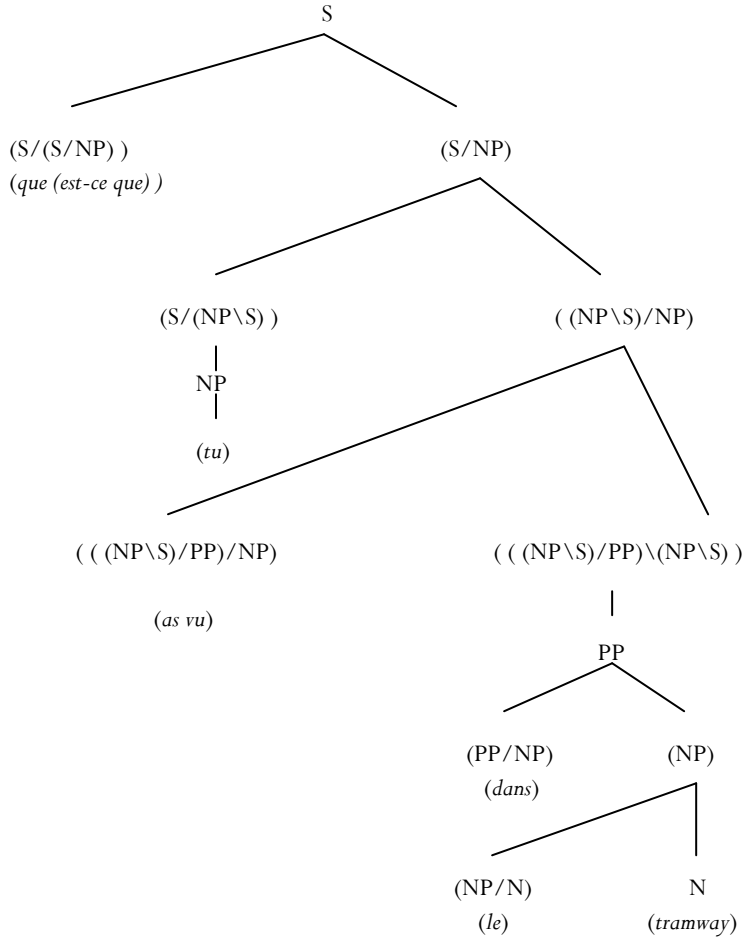
La première représentation correspond à un arbre dérivé, la seconde à un arbre de dérivation. Dans ce dernier cas, l'arbre qui s'ensuit se substitue au noeud verbal et dominerait éventuellement par adjonction les autres arbres qui serait sous cette relation de dominance. Ce type de schématisation implique qu'il existe des mots *pleins* et, par ce biais, des mots *vides*, mais cela demeure discutable, pour ne pas dire contestable.

Les grammaires catégorielles, pour leur part, assignent les catégories de foncteurs (*functors*) et celles d'arguments. Dans ces termes, il s'agit de fusionner les foncteurs avec leurs arguments respectifs autour de quelques règles générales (uniformisées notamment par Lambek 1961, et reformulées opportunément par Moortgat 1988) : par exemple, la caractérisation d'un verbe intransitif ou d'un mot-phrase se notera par un foncteur *s/n* (*sentence/name*), lequel indique que la dénomination catégorielle est aussi celle qui forme phrase, alors qu'un verbe transitif sera noté *(s/n) / n*, indiquant ainsi qu'une notation catégorielle débouche elle-même sur une phrase qui comporte un argument en plus d'un argument formant phrase. Si l'on prédique *je passe*, l'argument sujet *je* est désigné par *n* qui, avec le verbe intransitif, forme phrase (*s/n*), alors que dans *je passe la main*, le verbe se voit assigner le foncteur *(s/n) / n*, qui spécifie par là qu'un autre argument est instancié. Nous ne nous attarderons pas plus sur la verbalisation schématique de ce formalisme, dont nous retiendrons surtout que la combinaison des foncteurs et des arguments occasionne un certain nombre de généralisations (*general rules*). La formulation des règles répertoriées vient directement de la compréhension que l'on peut faire du trait slash (/), lequel peut présenter des versions différentes selon les auteurs[17]. Par exemple, certains linguistes choisissent de représenter l'argument à droite du slash, et par exemple le syntagme relevé à gauche. Le trait *n*, quoi qu'il en soit, renvoie à la catégorie à laquelle appartient l'argument, alors que le syntagme (ou l'expression, pour ne fâcher personne) apparaît sous la notation *s*. Le calcul de Lambek consiste dans ces termes à une reformulation des grammaires indépendantes du contexte. Partee et Borschev 2000 déterminent ainsi des généralisations en foncteurs pour des catégories aussi disparates que les mots-phrases (*t*), les arguments seuls (*e*), les verbes intransitifs (<*e,t*>), les verbes transitifs (<*e,*<*e,t*>>), voire les phrases nominales avec des quantifieurs généralisants (<<*e,t*>*,t*>). L'enjeu principal de ces règles générales est de permettre une plus grande simplicité dans les représentations (notamment translinguistiques), mais aussi une uniformisation invariable des notations.

Les particularités du calcul lambékien apparaissent plus explicitement encore à l'appui des changements de types (ce dernier terme désignant les

[17] Concernant le trait slash, c'est Bar-Hillel qui a introduit la notion de directionalité, faisant ainsi la distinction entre *s/n* et *s\n*. On s'y reportera éventuellement.

catégories), la position méthodologique consistant tout de même à assigner la catégorie la plus courante et la plus fondée au membre du lexique. Dans le cas d'une montée d'un argument par exemple (notamment dans celui des phrases interrogatives, comme *qu'est-ce que tu as vu dans le tramway ?*), la direction (à droite) du foncteur permet d'indiquer qu'un argument normalement placé après lui est instancié à sa gauche. Schématiquement, cela donne :



On assite ici à une polycatégorisation en somme très simple à fonder (Cf. Godart-Wendling 2002) : les mots et les syntagmes étant généralement appariés à des catégories conformes à leurs principaux emplois, ils peuvent se voir assigner d'autres catégorisations selon que leur emploi en contexte leur attribue d'autres caractéristiques, et fait en sorte qu'ils se comportent comme des membres d'autres catégories simultanément.

Concernant la Grammaire Applicative Universelle enfin, que nous n'avons fait qu'entrevoir, nous avons deux niveaux de relation des constituants : d'une part, un inventaire de constructions syntagmatiques propres à une langue en particulier (phénotype), et, d'autre part, la revendication d'opérations universelles, pour ainsi dire d'invariants translinguistiques, tels que la prédication, la détermination, la quantification et la thématisation (le génotype). L'enjeu majeur de cette formalisation est de répartir ce qui relève de la déformabilité et de l'indéformabilité, de l'invariance et de la variabilité, avec, pour cette dernière, une entrée par des grammaires catégorielles. Par ailleurs, elle permet de ne pas organiser la catégorie à partir d'une ressemblance d'unités ou d'un quelconque air de famille, mais de catégorisations plus élémentaires, de manière à ce que les propriétés communes des opérateurs soient effectivement les éléments constitutifs de la catégorie, et non l'inverse. Par exemple, un adjectif substantivé comme dans *il fait le beau* voit sa catégorisation ramenée à son comportement syntagmatique dans la phrase donnée : le type de l'unité construite est donc l'application d'un opérateur à un opérande. Pour reprendre ici Jean-Pierre Desclés, nous dirons que l'analyse syntaxique (*parsing*) devient alors un véritable « calcul » sur les types des unités agencées par l'application dans l'énoncé. En outre, autant d'invariants linguistiques comme les opérateurs statiques, cinématiques, topologiques, se rapportent effectivement, à un autre niveau, à des primitives cognitives et dans une perspective actionnelle.

Bibliographie sélective :

Abeillé A., 2002 : *Une Grammaire électronique du français*, Paris, CNRS éditions (coll. *Sciences du Langage*).

Candito M.H., 1999 : *Une Organisation modulaire et paramétrable de grammaires lexicalisées : application au français et à l'italien*, Université de Paris VII, thèse de doctorat nouveau régime.

Cori M., Marandin J.M., 1998 : « Héritage de propriétés dans les grammaires d'arbres poilychromes », *Linx 39, 2*, Université de Paris X : 13-39.

Godart-Wendling B., 2002 *(dir.)* : *Les Grammaires Catégorielles*, *Langages* 128.

Grootjen F, Kamphuis V. et Sarbo J., 1998 : « Coordination and multi-relational modelling : X and X' revisited », in *6e Conférence annuelle sur le Traitement automatique des langues naturelles (TALN)* : 345-351.

Kathol A., 1995 : *Linearization-Based German Syntax*, PhD Dissertation, Ohio State University.

Lablanche A., 2003 : « Les Constructions infinitives dans les grammaires d'arbres polychromes hiérarchisées », *Linx* 48, Université de Paris X : 29-42.

Lambek J., 1961 : « On the Calculus of syntactic types », in *Structure of Language and its mathematical Aspects* (*éd*. Roman Jackobson), Providence, Americain Mathematical Society.

Moortgat M., 1988 : *Categorial Investigations : Logical and Linguistic Aspects of the Lambek Calculus*, Dordrecht, Foris.

Partee B.H., Borschev V.B., 2000 : *Integrating formal and lexical Semantics*, Kolding.

Pollard C., Sag I.A., 1994 : *Head-Driven Phrase Structure Grammar*, Chicago, Université of Chicago Press.

Sleator D., Temperley D., 1993 : « Parsing english with a link grammar », in *Proceedings of the Third International Workshop on Parsing Technologies*.

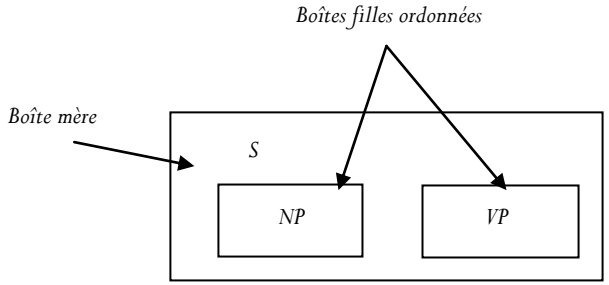
Evaluations possibles :

- confronter les Grammaires d'Arbres Adjoints (*TAG*) et les Grammaires d'Arbres Polychromes (GAP) avec les mêmes exemples ;
- traiter la question de la coordination et subordination dans un ou plusieurs formalismes résumés.

SEANCE 6 :
LES METHODOLOGIES EXPLICATIVES
ET LEUR FORMULATION DIDACTIQUE

Dégager une construction grammaticale, montrer comment s'organise un discours, une textualité, tout cela exige un traitement plus ou moins unifié [18]. Mettons des constructions verbales qui peuvent paraître problématiques, comme *te konpra(n)n Potoprens se paradi* ou *Tobi ansanm ak Frank gen dwa gen opinyon e viv lavi* : sur quoi insister ? comment aborder les phénomènes d'extraction, de coordination asymétrique, de réduplication ? Pour ce faire, il convient d'admettre que les supports pédagogiques ne peuvent être envisagés que dans une problématique téléonomique. Si l'on prend une phrase pour exemplifier tel ou tel phénomène cognitif, on est contraint de le (re)contextualiser. De même, les supports en question réclament d'être didactisés. En outre, une problématique téléonomique renvoie tout simplement à une contrainte incompressible de l'enseignement : il s'agit de savoir ce qu'on entend démontrer, et pourquoi pas à l'appui de textes littéraires (Vernet 2001) ou journalistiques (D'Aboville 2004).

L'une des applications de type génératif dans le domaine général des grammaires d'unification consiste dans les grammaires dites de construction (Kay et Fillmore 1999). Dans ce formalisme encore une fois, les modules descriptifs tels que la syntaxe, la phonologie et la sémantique sont rassemblés dans un supermodule où les constuctions par exemple locutionnelles sont envisagées dans des termes analogues à celles des formulations non locutionnelles, et dans le cadre d'une même normalisation. Dans cette vue, une construction renvoie à un schéma sommaire qui, comme un gabarit en somme, permet de reproduire tous les composants linguistiques comme le mot, le syntagme ou la phrase. Par ailleurs, les unités dégagées sont entièrement spécifiées (des *constructs*). La grammaire formalise ainsi un ensemble de constructions hiérarchisées qui peuvent être unifiées en vue de générer tous les syntagmes et toutes les phrases admissibles dans les langues. La phrase comporte dans ce sens des constituants syntagmatiques regroupés autour de la tête lexicale sélectionnée, dans une construction que schématisent des boîtes qui les contiennent. Chacun des constituants implique une notation binaire (déjà vue avec les GAP) d'attribut-valeur (*AVM*, pour *Attribute-Value Matrix*) et intègre donc une structure de traits, lesquels sont répartis en quatre classes : le trait ROLE, les traits SYNSEM, les traits VAL et les traits PHON[19]. Une phrase, notée *S* pour *sentence*, comprendra par exemple un NP sujet et un VP avec deux compléments (NP, PP, par exemple), répartis en boîte mère et en boîtes filles (ordonnées) :



Les constituants *NP* et *VP* de la phrase *S* impliquent tous deux des traits AVMs qui sont autant de variables qui se combinent dans la boîte mère dans laquelle ils sont inclus. Leur trait ROLE renvoie au rôle argumental qu'ils ont dans l'ensemble, comme celui de tête (HEAD), de spécifieur (SPEC), de complément (COMP), de modifieur (MOD), alors que les traits SYNSEM permettent de leur assigner un attribut catégoriel (CAT) et un attribut lexical

[18] On rappellera ici les explications générales de Moortgat 1987, Shieber 1989, Miller et Torris 1990, Van Benthem 1991, Lecomte et Rétoré 1997 ou Bickel 2000, qui établissent tous à leur manière un lien entre le positionnement méthodologique de la recherche et son contexte, à noter encore que le web laisse à disposition et à ce sujet de nombreux racourcis (ainsi Livia Polanyi (de Berkeley et Chicago), Robert Franck (de Delaware), mais également Dani Byrd (USC), James P. Blevins (de l'Université de Cambridge), Elliot Saltzman (Boston / lab. Haskins) et Robert D. Van Valin (State University of New York at Buffalo)).

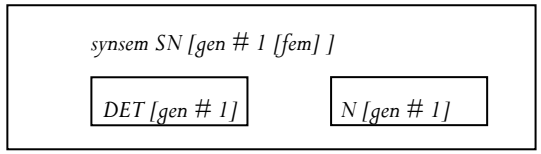
[19] On notera que la négation, dans le formalisme des grammaires de construction (schématisée en ¬), appartient en général à la classe des traits PHON.

(LEXH). Les traits de tête sont communs aux syntagmes dominés par la tête et la fille-tête : ils caractérisent donc à la fois les têtes et les syntagmes qui leur sont rattachés. Les grammaires de construction, dont l'enjeu descriptif réside surtout dans l'explicitation des structures relevées, adjoint à cette notation celle de niveaux (*levels*) de structure, à l'occasion desquels on indique si une unité lexicale est instanciée (+) ou non (–) avec le trait LEX ; et d'autre part si elle constitue l'argument principal ou non de la tête à travers le trait MAX. Si l'on prédique la suite *je l'ai connu*, le verbe *connaître* se voit attribuer les traits LEX +, MAX +, ce qui ne sera pas le cas, pour MAX, du pronom élidé complément. On aura donc, schématiquement :

$$\begin{bmatrix} synsem & \begin{bmatrix} head & \begin{bmatrix} cat & [V] \\ lexh & [ai\ connu] \end{bmatrix} \\ level & \begin{bmatrix} lex & + \\ max & + \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

Parmi les traits SEM, intégrés dans SYNSEM, les notations de genre et de nombre sont les plus courantes, mais ces spécifications incluent aussi les caractéristiques du temps, ainsi que des propriétés plus discursives résumées dans les FRAMES sémantiques (ou scènes), lesquelles renvoient aux schèmes énonciatifs en termes de participants à l'action de prédication (dans ces termes par exemple, un *part 1* parle à un *part 2*, et ainsi de suite)[20]. D'autres traits, d'ordre relationnel (REL), sont également admis dans ce type de formalisme : les rôles grammaticaux des constituants dans le syntagme, notés GF (sujet, objet par exemple), mais aussi les rôles thématiques (THETA, ou θ), comme l'agent ou le patient, le thème ou le rhème, le contenu (*content*), l'instrument, la source ou le but (*goal*).

Les traits VAL, de leur côté, correspondent à des traits de valence : les arguments compléments par exemple sont sous-catégorisés par la tête qui leur assigne un rôle dans le cadre valentiel, dans lequel des règles d'appariement (*linking rules*) combinent les modules PHON, SYNSEM et ROLE. Tout ceci renvoie au mécanisme d'unification, qui permet d'établir un héritage de propriétés dans les constructions envisagées. Deux valeurs étant unifiées par exemple, elles sont indiquées par la notation # et numérotée. Si l'on prédique *une femme*, le trait du genre (ici féminin) est commun aux deux composantes du SN, où elles s'emboîteront de la manière suivante :

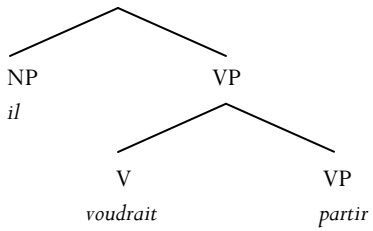


Que l'on prenne pour cibles descriptives les catégories *sujet-verbe* à travers la tête, ou *thème-rhème* à travers le *filler*, le mécanisme d'unification demeure le même à partir du moment où les propriétés indiquées sont rassemblées dans une construction emboîtante où différents niveaux de représentations héritent les uns les autres de ces traits qui les caractérisent (les boîtes mères sont ainsi notées INHERIT).

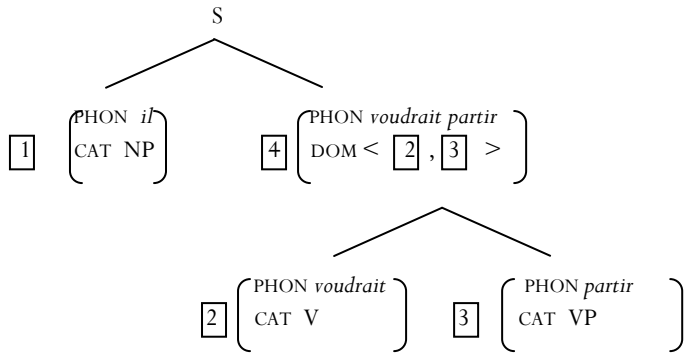
Le formalisme HPSG possède aussi une dimension explicative dans ce sens où il permet de sous-catégoriser (SUBCAT) les constituants filles de la tête, tout en exposant les contraintes qui s'exercent sur leur emploi. Ainsi sont convoquées les listes des structures argumentales (ARG-S), qui voient leur contrepartie lexicale indiquée dans une liste SUBCAT (voir *supra*). Partant de là, il est envisageable de linéariser les structures à l'appui de plusieurs schématisations (ici pour *il voudrait partir*) :

S

[20] Certains auteurs représentent les participants avec la notation ARG.

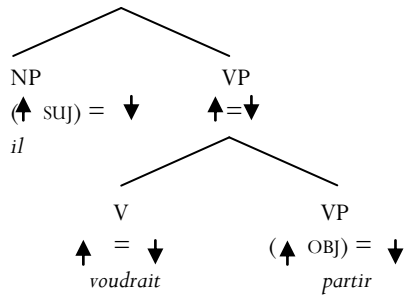


Et avec une série de sous-catégorisations :

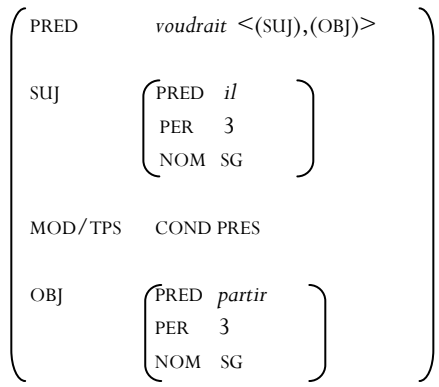


En grammaire lexicale et fonctionnelle en revanche, on sollicitera à la fois les approches basées sur des traits et les approches relationnelles. Les LFG dégagent par ce biais une architecture formelle assez simple à saisir, avec des transitions explicites entre les différents niveaux de représentation schématique. Les structures de constituants (*C-Structures*, pour *constituent-structures*) renvoient surtout aux catégories grammaticales et à l'ordre des mots, alors que les structures fonctionnelles (*F-Structures*, pour *functional-structures*) représentent les caractéristiques valentielles et sémantiques. Or ces deux niveaux de représentations occasionnent deux types de schématisations distinctes, dont voici la première, en C-structure :

S



et ensuite en F-structure :

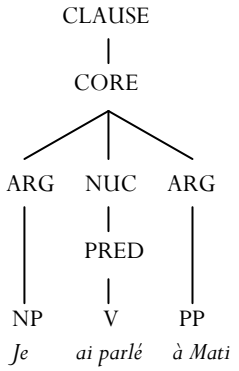


Les notations d'ordre fonctionnel établissent la correspondance entre les nœuds de la c-structure et leur contrepartie en f-structure, tandis que l'annotation = fléchée désigne les renvois qui sont effectués d'une figure à l'autre : la première schématise alors la structure en constituants à laquelle la seconde assigne un réseau de traits.

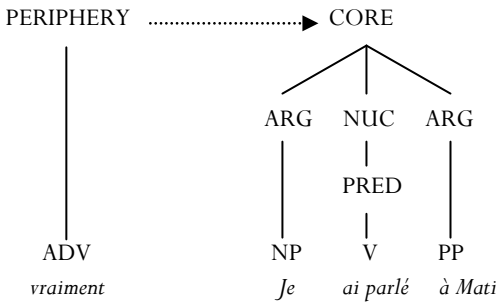
On peut déclarer de ce fait que ce qui attribue une dimension explicative à une problématique, un phénomène linguistique, et pourquoi pas un type de formalisme en particulier, c'est qu'il soit possible ensuite de l'externaliser. Par exemple, si nous pratiquons une recherche sur certaines formes syntagmatiques ou certains faits phonologiques, il est quelquefois très

opportun de rattacher les phénomènes (ou les unités) pris pour support à des phénomènes cognitifs, ou pourquoi pas culturels, sociologiques, psychanalytiques. Cela dit, on aura aussi la possibilité d'extraire le fait traité de toute forme de particularisation ou de généralisation, pour ainsi dire à l'état brut, en marge des données externes. Dans ce sens, il convient de moduler son domaine de recherche, et c'est dans cet esprit que Nølke 1999 préconise une vision entièrement méthodologique de la modularité.

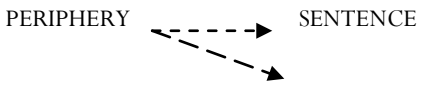
Dans le cadre des grammaires du rôle et de la référence (*RRG*), l'approche en termes de parsage (CF. les *parsers* sélectionnés) débouche sur une représentation d'ordre syntagmatique à laquelle est liée, par l'intermédiaire du lexique, une représentation sémantique. L'ensemble ainsi formalisé nous conduit à aborder le discours dans son entier. Les RRG rejettent les formats schématiques des représentations phrastiques telles que les relations grammaticales, le module X-barre, en leur reprochant notamment leur inappropriété descriptive et explicative dans le domaine des invariants interlinguistiques, lesquels sont notés en RRG à travers des caractéristiques définies (Van Valin et LaPolla 1997). La conception de la construction syntagmatique (*clause structure*), connue sous l'appellatif de *layered structure of the clause (LSC)*, est basée sur le fait que cette structure se résume dans le NUCLEUS, qui contient les PREDICATS ; le CORE, qui contient à la fois le nucleus et les ARGUMENTS des prédicats, et enfin la PERIPHERIE, qui inclut les MODIFIEURS temporels et locatifs du core. Dans la phrase (*sentence*), on compte ainsi un ou plusieurs syntagmes (*clause*) et donc un ou plusieurs core, dans lequel(s) sont distribués les arguments d'un nucleus qui renvoie à un ou plusieurs prédicat(s). Les notations sous-spécifiées, comme l'aspect verbal ou la négation, apparaissent donc à un niveau de nucleus du prédicat tout en renvoyant aux caractéristiques plus discursives en fin de parcours. Encore une fois, les indications (sous-spécifications, caractérisations) sont hiérarchisées dans ce sens où elles interviennent à des niveaux différents de la représentation syntagmatique de la phrase (les *junctures* : on parle de *core juncture, clausal juncture* ou par exemple de *nuclear juncture*). Un syntagme formant phrase par exemple, tel que *j'ai parlé à Mati*, donnerait schématiquement ceci :

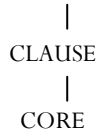


Mettons à présent que nous employions un modifieur comme *vraiment* : quelle est sa portée ? que modifie-t-il ? et donc à quel niveau de la représentation va-t-on l'indiquer ? Si celui-ci porte uniquement sur le verbe (*j'ai vraiment parlé à Mati*), c'est par exemple au niveau (à la *juncture*) du core que l'on va l'apposer en périphérie :

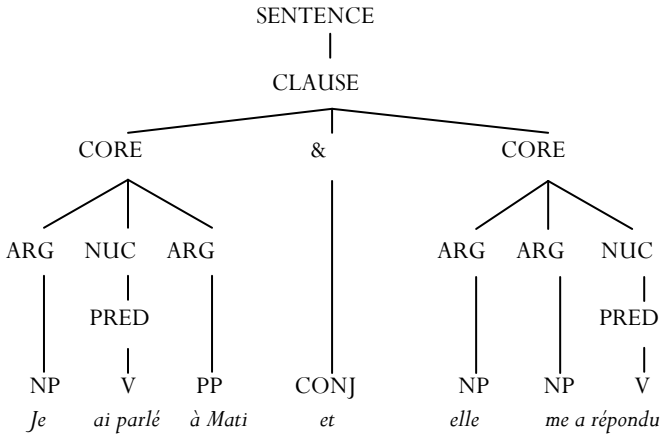


Si, dans un autre cas, c'est sur l'ensemble de l'action de prédiquer que porte le modifieur (ainsi avec les adverbes d'assertion par exemple : *effectivement j'ai parlé à Mati*), on fera remonter la notation au niveau du syntagme verbal dans son ensemble, voire la phrase elle-même :

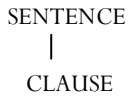


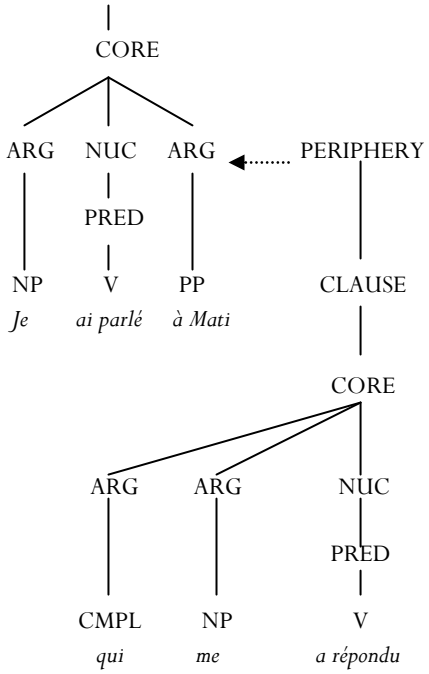


Les notations de coordination et de subordination répondent aux mêmes exigences schématiques. Par exemple dans une phrase où l'on aura deux syntagmes verbaux (*j'ai parlé à Mati / elle m'a répondu*), on aura la possibilité d'indiquer la coordination de l'un à l'autre (*j'ai parlé à Mati, et elle m'a répondu*), à partir du niveau CLAUSE :



Mais on aura aussi la possibilité de subordonner le second syntagme à l'un des arguments du premier core (*j'ai parlé à Mati, qui m'a répondu*) :





On remarque assez facilement dans le schéma que le SV adjoint à *Mati* constitue un modifieur à ce niveau, mais dans le même temps n'intervient pas ailleurs. En outre, un tel formalisme ne refuse pas l'assignation d'autres classes de traits, comme ceux d'agent et de patient (*actor / undergoer*). La portée explicative de ces traits caractéristiques est évidente : ces derniers attribuent des macrorôles sémantiques en complément des fonctions syntaxiques dans un système uniformisé (*RRG Linking System*) et à caractère universel.

Il convient donc de déterminer un ensemble fini non pas de représentations, mais d'opérations schématiques, ce qui se fait déjà en linguistique cognitive bien entendu, mais aussi en linguistique computationnelle (ainsi Stolzenburg,

Höhne, Koch et Volk en appellent, dans leur programmation, à un *finite set operations*). C'est ce que nous avons dans la GAU, mais aussi, de manière plus revendiquée, dans les RRG et le programme minimaliste chomskyen.

A titre d'ouverture, remarquons ensemble que l'approche de Culioli 1989 et 1990 répond aussi à une exigence de minimalisme. Celui-ci préconise de se débarrasser d'abord de tout ce qui concerne les particularités co(n)textuelles pour cerner des phénomènes principalement cognitifs, autour du classement terminologique de *lexis* et de *notion*. La lexis renvoie à l'action de prédiquer (de *dire*, schématiquement), au-delà même des constructions grammaticales et des unités discrétisées en discours. La notion, elle, reporte à la problématique de la généralisation : par exemple, la question de la diathèse subsume celle de l'actif et du passif, celle de l'assertion transcende la typologie grammaticale des phrases (déclaratives ou autres), tout comme il est des faits cognitifs qui permettent de passer outre la distinction entre l'affirmation et la négation, la rhématisation et la concaténation, l'adjectivation et la nominalisation.

L'approche de Culioli, d'un point de vue épistémologique, projette d'épurer le programme minimaliste de Chomsky, et de se placer au-delà même du principe d'unification que préconisent les grammaires qui relèvent de ce courant méthodologique. L'apport général des démarches explicatives qu'elle implique consiste, *primo*, à dégager quelles sont les véritables opérations marquées dans l'énoncé, en termes d'emploi et d'action (par exemple, s'agit-il d'interpeller ? d'asserter ? de reformuler ?) ; *secundo*, à donner les meilleures représentations (épi)linguistiques de ce qui peut être ramené à des phénomènes de cognition et de construction (d'où la dénomination des *poly*opérateurs : plusieurs opérations sont effectuées, lesquelles sont éventuellement hiérarchisées). L'approche culiolienne part donc d'énoncés, lesquels ne sont pas regroupés d'après des catégories fonctionnelles, mais en termes de (quasi)paraphrase. On échappe ainsi à la binarité de l'exemplification chomskyenne des faits linguistiques : Culioli établit des paradigmes répartis en *n*-membres, lesquels présentent des traits communs (à travers la *lexis*), en nombre restreint, et des différences qui relèvent des opérations qui sont accomplies. Mettons cette quasi-paraphrase résumable en DARLINE – TISSER :

Darline tisse
Voici Darline qui tisse
C'est Darline qui tisse
C'est bien tisser que fait Darline
Tisser, c'est ce que fait Darline
Darline fait du tissage
Tisse-t-elle, Darline ?
Darline tisse ?

Ces quasi-paraphrases dénotent toutes la même lexis, mais procèdent d'opérations distinctes, avec des marqueurs (ou des représentants) qui en sont autant de représentations particulières, notamment ici à l'appui du mécanisme d'extraction. Les opérations en question inventorient en un sens tout un dispositif d'invariants (et non d'universaux) : on ne peut pas parler par exemple de présentatifs dans toutes les langues, mais on peut le faire du *fléchage* ou du *parcours* des unités, à savoir que les combinaisons d'invariants sont aussi résumables en termes de combinaisons d'opérations. Voyons cette phrase : *Amélie et Emilie, non, Emilie et Amélie ont été reçues par Monsieur Untel.* Dans cet énoncé, l'opérateur *non* exerce une négativation du syntagme nominal précédent, mais simultanément de la prédication précédente dans son ensemble ; d'autre part, il consiste dans un correctif de ce qui vient d'être instancié, tout en procédant d'une rupture thématique. En fait, il est tout cela à la fois (et même plus), et c'est dans ce sens qu'il peut être envisagé comme un polyopérateur. La réaffirmation de cette combinaison d'opérations dénonce en même temps certains invariants dont les représentations épilinguistiques appartiennent au niveau tant de la lexis que de celui de la notion. Ces opérations sont donc indirectement cognitives : elles concernent une construction de formes qui renvoie plus ou moins explicitement à des opérations cognitives immatérielles. Du point de vue des représentations, celles-ci, ainsi matérialisées, partent d'emplois spécifiques, de constructions formelles, et répondent aussi à des contraintes sur ces emplois respectifs. Mais ce qui les rend très productives en termes explicatifs, c'est qu'elles concernent un ensemble limité de paramètres, en marge d'une possibilité de combinaisons beaucoup moins réduite. Les explications que cela permet de fournir sur les faits linguistiques répondent donc à des traits typiques (de l'ordre de la représentation, du notionnel), dans le même temps qu'à des traits spécifiques (singularités textuelles, phrastiques, syntagmatiques).

Dans le domaine pédagogique, de nombreuses questions sont traitées d'un point de vue ontologique bien entendu, mais surtout cognitif. Les problématiques d'enseignement/apprentissage posent donc d'une autre manière le sens des démarches didactiques dans leur dimension à la fois individuelle et sociétale[21].

Dans une problématique de regroupement, on assiste bel et bien à une disciplinarisation des savoirs linguistiques (l'expression est de Jean-Louis Chiss et Christiant Puech), dont la transposition didactique n'est encore arrivée qu'incomplètement dans les classes du primaire et du secondaire (notamment dans le contexte français, comme l'a bien noté Daniel Véronique). Au-delà des principes qui présentent une invariance à démontrer (et à exemplifier, dans des ensembles bien sûr limités !), les dimensions de la variation interlinguistique, celle des paramètres, n'est sans doute pas à mettre sur un même plan de didactisation. A ce titre, la linguistique fonctionnelle procède d'une approche sensiblement différente de celle de la grammaire générative (Hagège 2004) : les fonctionnalistes partent de langues dont ils décrivent les propriétés à tous les niveaux, de la phonologie à la pragmatique, et dont ils dégagent des constantes sans exceptions. Plutôt que d'affirmer que la phrase *x* viole la règle *y*, ceux-ci déclarent que la structure *x* n'existe pas dans la langue *z*. C'est justement cette dimension profondément normative de la grammaire générative que rejettent notamment les LFG, de sorte que l'on pourrait presque voir d'un côté les fonctionnalistes, qui partent de la forme vers la signification (une perspective sémasiologique), et de l'autre les formalistes, qui procèdent de manière inverse (une perspective onomasiologique, donc). Par ailleurs, la GG envisage surtout des phrases, là où les fonctionnalistes abordent volontiers les interfaces précitées à l'appui par exemple du texte ou encore du paragraphe. Dans une grammaire (linguistique) descriptive d'une langue en particulier, dont on explicite les structures vis-à-vis d'une autre, il convient donc d'énumérer les spécifications les plus accessibles, mais aussi les plus invariablement admises en général, et c'est là qu'interviennent les interfaces (ci-dessous les caractéristiques casuelles combinées avec des caractéristiques morphologiques et pragmatiques, un ex. repris de Creissels et Robert, en wolof) :

Peer	*moo*	*ko*	*lekk*
Pierre	SUJ. FOC.	3SG	manger. PASSE

C'est Pierre qui l'a mangé.

Au-delà des unités du lexique, les catégories grammaticales et les parties du discours comportent comme invariants des notations qui limitent justement le champ de la variabilité. Par exemple, tel ou tel affixe flexionnel appartiendra à la catégorie du temps, là où le nombre présentera une invariance de fait qui demeure en soi incontestable. Cela étant, il est peu vraisemblable d'imaginer l'existence d'une grammaire universelle : les formalismes et les composantes grammaticales sont juste là pour aborder les invariants à travers une classification non péremptoire (leur apport taxinomique), une dénomination particulière (leur apport terminologique et épinlinguistique), et un appareil descriptivo-explicatif fondé en partie sur une histoire et une approche qui lui sont propres (leur apport méthodologique). D'autre part, les coïncidences interlinguistiques entre les formes n'ont rien d'évident ni d'arrêté en soi, et il en est de même pour les significations (la production de sens), même si l'on peut dégager des domaines de grammaticalisation où s'épanouissent les catégorisations grammaticales (Lazard 1992).

Aussitôt qu'il s'agit d'enseigner dans le cadre des cours de premier cycle universitaire, et vu la profusion des approches envisageables aujourd'hui en linguistique, il convient de prendre du recul par rapport à cette dimension pour ainsi dire encyclopédique du savoir. Dans ce sens, tout nous contraint à faire office d'épistémologue, et la tâche en est rendue d'autant plus difficile que l'épistémologie ne peut en aucun cas faire l'économie d'une historiographie patiente et co-construite.

Que l'on s'affirme comme généraliste de la linguistique ou non, on oubliera quoi qu'il en soit d'ambitionner toute forme de linguistique générale, laquelle n'appartient peut-être, comme l'anthropologie, qu'à l'univers des sommes philosophiques. Parlons plutôt plus simplement d'approches, de démarches,

[21] W. Mérant m'a appris par exemple que des expressions telles que *lwa se boujwa wi lwa yè* sont courantes chez Maurice Sixto. Or, le simple fait de reconnaître dans ce genre de formules une réduplication polémique, au-delà même de la tournure elliptique et superlative des SV, nous place d'emblée dans un débat à la fois linguistique et culturel.

de problématiques et, bien entendu, de formalismes, et rappelons encore une fois, si cela se peut, que le linguiste est avant tout un méthodologue aujourd'hui sans doute incontournable, certes, mais un méthodologue quand même.

Bibliographie sélective :

Bickel B., 2000 : « Grammar and social Pratice : the role of « culture » in linguistic relativity », in *Evidence for linguistic relativity* (*éd*. R. Dirven et S. Niemeier), Amsterdam, Benjamin.

Culioli A., 1989 : *Rôle des représentations métalinguistiques en syntaxe*, comm. au XIIIe Congrès des Linguistes, Tokyo, 1982, repris dans la coll. ERA 642 (PARIS VII).

Culioli A., 1990 : *Pour une Linguistique de l'énonciation* (I : *Opérations et Représentations*), Paris, Ophrys.

D'Aboville F., 2004 : *Co-construction de l'image de soi dans la presse écrite haïtienne*, mémoire de master 1, Paris V-Sorbonne.

Hagège C., 2004 : « On Categories, Rules and Interfaces in linguistics », *The Linguistic Review* 21 : 257-276.

Kay P., Fillmore Ch. J., 1999 : « Grammatical Constructions and linguistic Generalizations : the *what's X doing Y ?* construction », Larousse, *Langages* 75 :

Lazard G., 1992 : « Y a-t-il des catégories interlangagières ? », in *Texte, Sätze, Wörter und Moneme, Festschrift K. Heger* (*éd.* S. Anschütz), Heidelberg, Heidelberg Orientverlag : 427-434.

Lecomte A., Rétoré C., 1997 : « Words as modules : a lexicalised grammar in the framework of linear logic proof nets », in *International Conference on Mathematical Linguistics* II (éd. C. Martin-Vide), Amsterdam, Benjamin.

Miller P., Torris T. (*éd*.), 1990 : *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Paris, Hermès.

Moortgat M., 1987 : « Compositionality and the Syntax of words », in *Foundations of Pragmatics and Lexical Semantics* (*éd.* J. Groenendijk, D. de Jongh et M. Stokhof), *Groningen-Amsterdam Studies in Semantics (GRASS)* 7, Dordrecht, Foris : 41-62.

Nølke H., 1999 : « Linguistique modulaire : principes méthodologiques et applications », *in Approches modulaires, de la langue au discours* (*dir*. H. Nølke et J.M. Adam), Lausanne, Delachaux et Niestlé : 17-74.

Shieber S., 1989 : *Parsing and Type Inference for natural and computer languages*, Ph. D. Thesis, Standford University Press.

Van Benthem J., 1991 : *Language in action : Categories, Lambdas and dynamic Logic*, in *Studies in Logic and the Foundations of Mathematics* 130, North Holland, Amsterdam.

Van Valin R. J., LaPolla R., 1997 : *Syntax : Structure, Meaning and Function*, Cambridge, Cambridge University Press.

Vernet P., 2001 : *Analyse du roman / analiz woman* Eritye Vilokan*, de Pierre Michel Chéry*, Port-au-Prince, *FLA, collection Anthropos*.

Evaluations possibles :

- dégager ce qui, en linguistique modulaire, prend une dimension explicative en arrière-plan de ce qui présente une dimension descriptive ;
- exemplifier l'opportunité explicative et descriptive des Grammaires du Rôle et de la Référence (*RRG*).