UNE GRAMMAIRE DE L'ARABE DANS LE FORMALISME GPSG

Hafida TENDJAOUI-MERABTI
Centre de recherche scientifique et technique
pour le développement de la langue arabe

Résumé

Un objectif principal de la linguistique informatique était l'utilisation d'une théorie linguistique pour diriger l'implémentation d'un système efficace de traitement automatique du langage naturel. La GPSG (Generalized Phrase Structure Grammar) qui est un formalisme basé sur l'unification des structures de traits complexes, offre des mécanismes puissants pour rendre compte de certains phénomènes linguistiques particuliers, tels que l'introduction des traits à valeur catégorielle pour le traitement des dépendances non bornées et le format ID/LP pour traiter les langues où l'ordre des mots est libre. L'arabe étant une langue qui possède cette propriété, nous allons essayer dans cet article de présenter l'application de GPSG pour la formalisation des structures syntaxiques de l'arabe.

Mots clés: Traitement automatique du langage naturel.(TAL), grammaires basées sur l'unification, linguistique informatique, GPSG.

الملّخص:

يتمثل الهدف الأساسي للّسانيات الحاسوبية في استغلال نظرية لّسانية لوضع نظـــام فعّـال للعلاج الآلي للغات الطبيعية. يمنح نظام GPSG (القواعد البنوية المعممّة) الصـــوري-المرتكـز على عملية توحيد للبنى ذات التراكيب المعقدة- آليات قوية كفيلة بشرح بعض الظواهر اللغويـة الخاصة كإدخال الصفات ذات القيمة التصنيفية لعلاج بعض التراكيب التي تحوي مثلا الضمـير العائد، والشكل ID/LP لعلاج اللّغات التي تتميّز بترتيب متغيّر للكلمات. وبما أن اللغة العربيـــة تتسم بهذه الميزة، سنحاول في هذا المقال تطبيق GPSG لإعطاء شكل صوري للبنى التركيبيـــة للّغة العربية.

الكلمات المفاتيح: العلاج الآلي للغة - القواعد المرتكـزة علـى توحيـد البنـى - اللسـانيات المعلوماتية - GPSG.

Abstract

An important goal of computational linguistics has been the use of a linguistic theory to direct the implementation of efficient language processing systems. GPSG (Generalized Phrase Structure Grammar) which is a unification based grammar formalism, provides powerful mechanisms to handle many linguistic aspects, as the category valued features that handle the unbounded dependencies and the ID/LP which is the most attractive feature of GPSG for parsing variable word order languages In this paper we will try to employ GPSG to formalize the syntactic structures of Arabic, since has a high degree of word order freedom.

Keys words: Natural Language Processing (NLP), Unification based grammar, computational linguistics, GPSG.

I-Introduction

La GPSG est un formalisme syntaxique récemment mis au point [GAZ 85]. Il est né d'une tentative visant à définir un système formellement restrictif qui soit capable de traiter une grande variété de phénomènes syntaxiques et surtout sémantiques considérés auparavant comme incompatibles avec ces restrictions. GPSG a été utilisée pour décrire un important fragment de l'anglais et par la suite des travaux substantiels utilisant cette théorie ont porté particulièrement sur l'allemand et le japonais [USZ 87], [BUS 87], [VOL 88], [GUN87]. Dans cet article nous allons essayer de présenter une application de GPSG à la modélisation des structures de la langue arabe telles que définies dans le modèle linguistique Néo-Khalilien. Avant d'aborder la formalisation des structures syntaxiques de l'arabe nous allons présenter brièvement la terminologie et le fonctionnement de GPSG ainsi qu'un petit aperçu sur la syntaxe de l'arabe selon l'école Néo-Khalilienne.

II-Terminologie de GPSG

GPSG contient cinq composantes particulières à un langage :
- les règles de dominance immédiate ID (Immediate Dominance)
- les règles de précédence linéaire LP (Linear Précédence)
- les métarègles
- les restrictions sur la cooccurrence des traits FCR (Feature Co-occurrence Restrictions)
- les spécifications des traits par défaut FSD (Feature Specification Defaults) qui interagissent avec trois principes universels :
- le principe des traits de tête HFC (Head Feature Convention)
- le principe des traits de pied FFP (Foot Feature Principle)
- le principe de contrôle et d'accord CAP (Contrôle Agreement Principle) constituant ainsi des conditions de bonne formation d'une structure syntaxique.

Avant d'aborder ces composantes nous devons définir une catégorie syntaxique en GPSG.

II. 1 Conception des catégories syntaxiques comme structures de traits

En GPSG les règles de grammaire ne manipulent pas des éléments non analysables traditionnellement désignés par V, N, SN, etc. mais des catégories syntaxiques définies comme des ensembles de traits syntaxiques. Formellement une catégorie syntaxique est une fonction partielle de l'ensemble des noms de traits vers l'ensemble des valeurs. Une catégorie est donc un ensemble de pairs du type : **trait syntaxique = <attribut, valeur>.**

Pour chaque grammaire il existe un ensemble F = {f1,f2,...,fin}de traits syntaxiques avec |F| = n. Le nombre «n» de traits syntaxiques peut varier d'une grammaire à une autre mais il est constant pour une grammaire particulière, et à chaque trait on associe un domaine D (fi). Une catégorie complexe C est un n-uplet avec C ∈ D (f1) x D (f2)x....x D (fn).

Exemple :

soit F = {N,V,BAR,PER,PLU,CAS}l'ensemble des traits Fi et leurs domaines respectifs

Traits Fi	Domaines D(fi)
N	{+,-}
V	{+,-}
BAR	{0,1,2}
PER	{1,2,3}
PLU	{+,-}
CAS	{Nominatif, Accusatif, Génétif}

Considérons la catégorie C donnée en (1)

(1) C = {< N,+>,<V,->,<BAR,2>,<PER,3>,<CAS,Nom>}

C'est la catégorie appelée traditionnellement SN nominatif.

- Les traits < N,+> et <V,-> désignent un élément nominal non verbal d'après la décomposition de Chomsky 1970,
- le trait <BAR,2> est le niveau de barre selon la théorie X-barre et
- l'interprétation des traits <PLU,-> et <PER,3> est transparente : il s'agit d'un élément singulier à la troisième personne.

En français le syntagme qui domine (2) satisfait la description donnée (1).

(2)Le petit garçon.

En GPSG les traits se distinguent par leurs valeurs :

-Les traits à valeur atomique qui comprennent les traits à valeur booléenne, ou ayant un autre type de valeur atomique.

Exemple: N ∈ {+,-}ou PER ∈{1,2,3}

-Les traits à valeur catégorielle qui constituent une contribution majeure et originale de GPSG.

II.2 Les règles de la grammaire

La GPSG fait la distinction entre les règles de Dominance immédiate ID (Immediate Dominance) et les règles de précédence linéaire LP (Linear Precedence). Cette distinction constitue le format ID/LP de la grammaire.

II.2.1 Les règles ID

Ce sont des règles non contextuelles qui expriment la relation de dominance qui lie un syntagme à ses constituants immédiats indépendamment de l'ordre de ces derniers. Une ID est de la forme :

Cat0 (Cat1, Cat2, Catn,...). Le membre gauche est la catégorie mère, le membre droit est un ensemble à répétition non ordonné. Une telle règle permet donc n ! règles syntagmatiques traditionnelles. Parmi les règles ID on distingue les règles ID lexicales qui introduisent une tête lexicale. Elles sont de la forme Cat0 (<wordform>).

II.2.2 Les règles LP

Elles contraignent l'ordre des catégories du membre droit d'une règle ID. Une règle LP s'applique à toutes les règles ID, et son champ d'application est constitué des caté-

gories sœurs. Une LP est de la forme Cat1< Cat2, ce qui signifie que la catégorie Cat1 doit précéder la catégorie Cat2 lorsque Cat1 et Cat2 sont des constituants sœurs.

Remarque :

Le format ID/LP permet de formuler explicitement les généralisations concernant l'ordre partiel des constituants filles d'une règles ID via les LP. Pour une interprétation formelle du format ID/LP nous avons besoin de la notion d'arbre local qui est un arbre de profondeur1. un arbre local T est admis par une grammaire ID/LP G si et seulement si il vérifie une certaine ID et toutes les LP dans G.

II.2.3. Les métarègles

Une métarègle est un mécanisme qui s'applique aux règles de la grammaire pour produire de nouvelles règles. Formellement une métarègle est une fonction de l'ensemble des ID lexicales vers lui-même. Une métarègle comporte deux parties : un modèle et une cible. Elle s'applique à toute ID lexicale qui satisfait le modèle et définit à partir d'elle une règle modifiée d'après la description de la cible.

II.3. L'instanciation des traits

Avant de présenter l'instanciation des traits deux définitions sont nécessaires :

Définition1: Une catégorie E est une extension d'un catégorie $C \subseteq E$ ssi :

i)Tous les traits à valeur atomique dans C sont également présents dans E avec la même valeur

ii) Pour tout trait f ayant une valeur catégorielle, la valeur de f dans E est une extension de la valeur de f dans C.

Définition 2: L'unification de deux catégories A et B notée $A \cup B$ est la catégorie minimale qui est l'extension de A et de B. Si une telle catégorie n'existe pas l'unification n'est pas définie.

L'approche adoptée par GPSG fait que les catégories présentes dans un arbre local peuvent être des extensions des catégories de la règle ID responsable de la génération de cet arbre. Ceci signifie que les traits qui apparaissent dans un arbre local ne sont pas obligatoirement spécifiés par une règle ID. Cet ajout de trait c'est **l'instanciation** des traits. Les traits ajoutés de cette façon sont dits **librement instanciés**, ceux dont la présence est requise par une règle ID sont dits **hérités.**

Par ailleurs, GPSG introduit un certain nombre de principes généraux qui gouvernent la distribution des traits syntaxiques dans une structure.

II.3.1. Les restrictions sur la co-occurrence des traits FCR (Feature co-occurrence restriction)

Les catégories syntaxiques étant conçues comme des structures de traits, les FCR sont utilisées pour formaliser les contraintes sur la cooccurrence des traits, pour imposer ou exclure certaines combinaisons de traits syntaxiques. Une FCR est une implication de la forme $A \supset B$ où A est la catégorie condition pour l'application de la FCR et B la catégorie conséquence. Une FCR est applicable à une catégorie C si C est une extension de A, dans ce cas C doit être unifiable avec B. Si C et B ne sont pas unifiables la catégorie C est dite non légale. Une catégorie légale doit vérifier toutes les FCRs. Les

spécifications de traits qui apparaissent dans les extensions légales d'une catégorie sont dits libres (non exclus par les FCRs et d'autres principes).

II.3.2. Les spécifications des traits par défaut FSD (Feature Specifications Default)

Il s'agit d'un mécanisme qui permet d'exprimer le fait que n est la valeur par défaut d'un trait f. n doit apparaître dans tous les cas où la valeur de ce trait n'est pas requise ou spécifiée par une règle.

II.3.3. Le principe des traits de tête HFC (Head Feature Convention)

Ce principe exige que dans tout arbre local la catégorie mère et la ou les catégorie(s) tête(s) qu'elle domine partagent leurs traits de tête libres. Ce principe est indirectionnel et donc neutre pour une interprétation ascendante ou descendante.

II.3.4. Le principe des traits de pied FFP (Foot Feature Principle)

Ce principe exige que dans un arbre local les traits de pied instanciés sur la catégorie mère soient identiques à l'unification des traits de pied instanciés sur la catégorie qu'elle domine.

II.3.5. Le principe d'accord CAP (Control Agreement Principle)

Ce principe exige que la valeur des traits de contrôle d'une catégorie contrôlée soit identique à la valeur d'un sous ensemble de traits de la catégorie contrôleur. Contrôleur et contrôlé sont définis sémantiquement.

II.4. Admissibilité des arbres

Pour générer des structures syntaxiques (arbres) durant l'analyse ou la synthèse, les règles ID sont traduites par des arbres locaux où toutes les catégories sont légales. Ce schéma est appelé projection.

Définition 3 : Soit une règle de ID r = (CO, {C1,C2,...,Cn }) et un arbre local t = C'0

Une projection f de r telle que f(r) = t est une bijection dont le domaine est {CO, C1, ..., Cn} et l'image est{C'0,C'1,...,C'n } qui remplit les conditions suivantes :

i) f (C0) = C0

ii) $\forall$ i, O $\leq$ i $\leq$ n ϕ (Ci) est une extension légale de Ci .

L'ensemble des projections de r est noté Φr.

II.4.1. Projection admissible

Lorsqu'un arbre local est admis comme une projection d'une règle ID, les différents principes ainsi que les règles LP lui sont appliquées pour vérifier son admissibilité. Une projection admissible est définie comme étant une projection qui satisfait les règles LP et tous les principes d'instanciation des traits.

Définition 4: Etant donné un ensemble R de règles ID, un arbre t est admissible pour R si

i) t est terminé

ii) tout sous arbre local dans t est soit terminé, soit une projection admissible d'une règle r $\in$ R.

Un arbre est terminé si toutes ses feuilles sont des éléments du vocabulaire terminal.

III. Propriétés de l'arabe

Selon le modèle linguistique Néo-Khalilien [HAD 79], toute séquence syntaxique bien formée de l'arabe est analysable en un ensemble d'éléments notés R, Ti, D qui ont les rôles respectifs d'élément régissant et termes régis et déterminant. Ces éléments constituant l'alphabet de la combinatoire syntaxique relèvent de trois niveaux différents à savoir :

1- le niveau lexical dont les éléments sont des *Kalims* (كلم) (nom, verbe ou mot outil),

2- le niveau intra-lexical ou niveau de la lexie (nominale ou verbale) qui est définie comme étant toute séquence isolable minimale qui admet des ajouts par simple concaténation sans que cela lui fasse perdre son caractère de séquence insécable,

3- le niveau syntaxique.

Concernant les différents types de constructions syntaxiques en arabe, deux grandes classes se distinguent selon la nature du régissant:

- les structures basées sur un verbe non exponentiel tels que : خرج، ضرَبَ، أعْطى
- les structures basées sur l'*ibtida*' (الابتداء) parmi lesquelles :
  i) celles dont le régissant est un verbe exponentiel tel que : أعلَم، حَسبَ، كان
  ii) celles dont le régissant est l'élément vide (ou un exposant tel que *Inna* (إنّ) et ses homologues.

Dans ce qui suit nous appellerons :

1- construction syntaxique verbale une structure basée sur un verbe non exponentiel

2- construction syntaxique nominale une structure basée sur l'élément vide Ø.

IV. Certaines structures syntaxiques de l'arabe dans le formalisme GPSG

Pour la formalisation des structures nous utiliserons quatre valeurs du trait BAR {0,1,2,3}correspondant respectivement au niveau lexical, niveau transitoire, niveau de la lexie [HAD 79] et le niveau syntaxique.

IV.1. La lexie nominale

La lexie nominale est construite à partir d'un noyau nominal et des ajouts à droite et à gauche de ce noyau. Cette définition couvre le syntagme nominal et syntagme prépositionnel. Sa structure est schématisée comme suit :

Préposition	Déterminant	Noyau	flexion	Complément/ Tanwin	Caractérisant

avec : Préposition ∈ {على ، إلى، من،...}, Déterminant ∈ { آل }noyau : nom,
Flexion ∈{الضمّة، الفتحة، الكسرة }, Caractérisant : peut être un objectif ou une phrase.
complément/*tanwin* : peut être soit le *tanwin* (التنوين) soit une lexie nominale ou une phrase.
Exemple : ... ، ولَدٌ،الولدُ، الولدُ الصّغيرُ، نافذةُ الغُرْفَة، نافذةُ غُرْفة المَنْزِل، ابن الذي أصطادَ السمكة

IV.1.1. Particularités de la lexie nominale

-La lexie nominale possède une structure récursive. Cette récursivité apparaît au niveau des positions du complément adnominal et du caractérisant.

-En arabe il faut distinguer une lexie nominale définie (معرفة) d'une lexie nominale indéfinie (نكـــرة) car cette différence est majeure lors de la reconnaissance de certaines constructions. Pour ce faire nous utilisons le trait DEF ∈{ +,- }.

-En plus du trait DEF la lexie nominale possède d'autres caractéristiques données par les traits tels que Cas {Nominatif, Accusatif, Génitif }, le genre et le nombre qui sont des traits de tête.

-Certains éléments de la lexie s'excluent mutuellement. Par exemple :

-un nom marqué par le déterminent (ال) ne peut contenir un complément adnominal,

-un nom qui possède le tanwin ne peut être marqué par le déterminant ni contenir un complément,

-un nom défini ne peut être précédé par le déterminant.

Pour rendre compte de ces phénomènes nous avons besoin de distinguer différentes classes de noms décrits par un ensemble de traits. Par exemple le nom commun et le nom propre correspondent respectivement aux entrée lexicales (1.a) et (1.b)

(1) a-{[+N],[-V],[[BAR 0],[-Cad],[-P],[-A],[-DEF]}

b-{[+N],[-V], [BAR 0],[+Cad],[-P],[-A],[+DEF]} avec

Cad ∈ {+,- } signifie qu'un nom peut admettre ou non un complément adnominal .

P ∈ {+,- } signifie qu'il s'agit ou non d'un pronom.

A ∈ {+,- } signifie qu'il s'agit ou non d'un pronom affixe.

DEF ∈{+,- } signifie qu'il s'agit ou non d'un nom défini.

IV.1.2 La structure de la lexie nominale

La lexie nominale sans préposition est considérée comme étant la catégorie {[+N], [-V], [BAR 2]}, en abrégé N2, qui signifie qu'il s'agit d'un élément nominal non verbal de niveau de barre 2. Nous donnons ci-dessus un ensemble de règles dans le format ID/ LP, qui couvre un sous-ensemble des lexies nominales.

(2) a - N 2 → N1, (A2)

b - N1 → N0,[-Cad], N1[Gen]

c - N1 → N2[+P,-A]

d - N1 → N0[+Cad], (A2)

e - N1 [+DEF] → ART, N0, [-Cad], (A2)

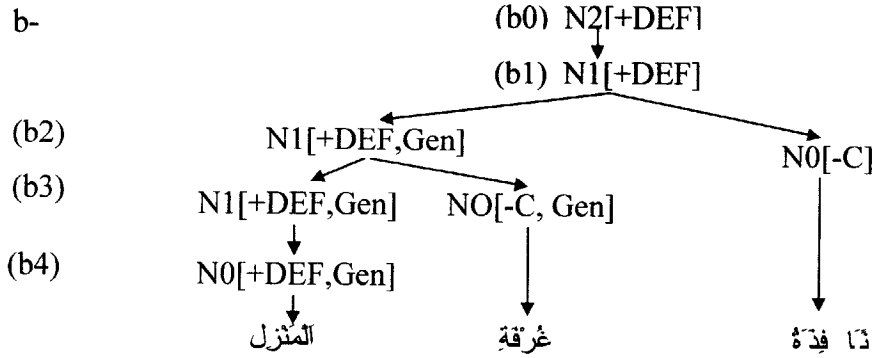
(3) a - N2< A2 b - N0< N1 c - ART < N0

Exemple : soit la séquence (5.a)

(5) a- نـــافِذَة غرفَةِ المَنزِل

L'arbre (5.b) correspondant à la structure de (5.a) est généré par l'application des ID (2.a),(2.b) et (2.e).

b-



Le trait DEF est un trait de tête et de pied, le FFP et le HFC sont responsables de la transmission du trait DEF dans les arbres locaux (b1),(b2), et (b3).

Dans l'arbre local (b1) DEF est instancié sur le constituant tête N1 par application du HFC, dans l'arbre local (b2) et (b3) le HCE ne peut être appliqué et donc DEF ne sera pas instancié sur le NO qui est la tête de N1 à cause de la FCR (4), mais puisque DEF est un trait de pied il sera donc instancié sur N1 en vertu du FFP.

(4) FCR : [-C ad]⊃[-DEF

IV.2 La lexie verbale

Selon [HADJ79] la lexie verbale est l'ensemble formé d'un noyau verbal, des pronoms affixes compléments ainsi que de certains éléments qui sont appelés convertisseurs ou exposants tels que {قَدْ، مَا، لَنْ، لَمْ ، ...}. Ce noyau contient toujours une kalima dérivable et un pronom sujet explicite dans certains cas par *exemple*: لَمْ أخْـرُجْ. Ces éléments qui apportent des informations sémantiques telle que la négation, ont un effet de rection sur le verbe c'est-à-dire (الرّفع،النصب، الجزم) pour rendre compte de ce fait nous utilisons le trait de tête VFORM qui indique en plus de la forme rectionnelle du verbe, son mode à savoir : {الماضي،المضارع،الأمـر}.De plus, pour distinguer le verbe infinitif du verbe conjugué nous utilisons le trait à valeur booléenne SUJ qui indique si le verbe contient dans sa forme morphologique l'information concernant son sujet ou non. Par exemple les deux verbes (خرج، خرجوا) satisfont respectivement les descriptions V0[-SUJ] et V0[+SUJ]. Par ailleurs, les verbes se distinguent par la nature et le nombre de compléments qu'ils sous catégorisent. Pour formaliser ce fait nous utilisons le trait SUBCAT introduit par GPSG et qui fonctionne comme un pointeur pour classe de verbes. Ce trait qui apparaît dans les entrées lexicales ainsi que dans les règles ID permet de relier une entité à son contexte d'insertion tout en restant dans un système de règles non contextuelles. Par exemple, un verbe transitif est décrit comme: VO [SUBCAT2]. Nous utilisons le trait SUBCAT pour désigner aussi d'autres éléments tels que les convertisseurs et les exposants : les catégories [SUBCAT α] et [SUBCAT β] désignent respectivement un exposant et un convertisseur, avec $\alpha \in$ {قَدْ، ما، ...}et $\beta \in$ { لَمّا، لَن...}. La structure d'une lexie verbale dans le format ID/LP est comme suit :

(1) V2 [exp α,Conv β] [› [SUBCAT α], [SUBCAT β],

(2) [SUBCAT α] < [SUBCAT β] < V0

Il faut noter que certains convertisseurs apparaissent avec des exposants particuliers, ceci est assuré par des FCR de la forme (3)

(3)FCR : ~ [Exp A, Conv B] avec A et B appartenant à des ensembles bien définis.

La forme rectionnelle du verbe due aux convertisseurs peut être imposée par des FCR de la forme :

(4) FCR :[Conv B] ⊃ [VFORM X] qui signifie que si le convertisseur B précède un verbe alors VFORM doit avoir la valeur X..

IV.3 Les constructions syntaxiques verbales

En arabe la distinction traditionnelle entre un SN et un SV n'est pas applicable dans une construction syntaxique verbale à cause de l'ordre relativement libre des constituants. On note une seule restriction : l'élément au nominatif ne peut jamais être antéposé au verbe. Nous proposons donc pour une phrase telle que (1) la structure donnée par la règle ID (2).

(1) السّمكة الوَلدُ اصطاد

(2) V3 › V2 [-SUJ], N2 [Nom], N2 [Acc]

Cette règle permet six permutations possibles dont trois seulement sont acceptables en arabe :

(3) a- اصطَادَ السَّمكَةَ الأَوْلادُ b- اصطَادَ السَّمكَةَ الأَوْلادُ c- الأَوْلادُ السَّمكَةَ اصطَادَ

Pour rendre compte de ce fait nous introduisons la règle LP (4)

(4) V2 < N2 [Nom]

Il faut noter que bien que l'ordre des constituants soit libre dans une construction syntaxique verbale, dans certaines situations cet ordre est figé, c'est le cas où les arguments du verbe sont des pronoms et donc le pronom doit immédiatement suivre le verbe. Ainsi la séquence (5) est inacceptable.

(5) *اصطاد الولدُ ها

Pour rendre compte de ce fait nous pouvons exploiter le trait PRO pour exprimer les différentes règles LP que doit vérifier une phrase verbale. Ces LP sont les suivantes :

(6) a- V2< N2 [Acc, +PRO]

b- N2 [+P, Acc] <N2 [-PRO, Nom]

Dans ce type de constructions nous voyons clairement l'intérêt d'utiliser le format ID/LP.

IV.4 Les constructions syntaxiques nominales

Nous considérons les constructions syntaxiques nominales à régissant Ø comme étant la catégorie N3. Ces constructions sont constituées d'un sujet (مبتدأ) et d'un attribut (خبر). En général le sujet est une lexie nominale définie au nominatif, et l'attribut est un adjectif indéfini au nominatif. Exemple (1) نائمُ الرَّجل

La structure de N3 est donnée par :

(2) N3→N2 [+DEF], A2[-DEF, Nom]

Dans (2) la valeur de Cas dans N2 n'est pas spécifiée et donc nous permettons, par là aux différentes valeurs de CAS d'apparaître aux mauvais endroits. Pour éviter cela et pour garantir que le N2 soit au nominatif nous utilisons la FSD (3)

(3) FSD :[Nom]

(3) signifie que lorsque la valeur du trait Cas n'est pas spécifié par une règle ID ou par un principe d'instanciation des traits (HFC,ou FFP), il est par défaut au nominatif.

L'intérêt de cette FSD réside dans le fait que nous pouvons utiliser la règle ID (2) pour rendre compte de certaines structures, au lieu d'écrire des règles spécifiques à chaque cas.

En effet, en ajoutant à N3 un élément de l'ensemble (إنّ) et ses homologues on obtient une construction syntaxique nominale où le sujet est à l'accusatif ce qui est illustré en (4).

(4)إنّ الرَّجُلَ نائمٌ

L'élément (إنّ) qui sera considéré comme la catégorie {[SUBCATإنّ]}, et pour distinguer les constructions syntaxiques nominales à régissant Ø de celles dont le régissant est (إنّ) nous utiliserons le trait REG. La règles ID (2) nous donne la structure d'une construction syntaxique nominale à régissant (إنّ).

(5) N3 [REG α] ([SUBCAT α], N3[REG NIL] où α ∈ {***inna*** et ses homologues}

Pour rendre compte du fait que le thème c'est-à-dire le N2 dans N3 [REG( ] est à l'accusatif nous utiliserons la FCR (6).

(6)FCR :[REG α] < [Cas Acc].

La FCR (6) signifie que si le régissant est (le constituant tête doit être à l'accusatif. Ceci est garanti par le HFC puisque Cas est un trait de tête.

Dans ce type de constructions l'ordre des constituants est figé. إن précède le sujet qui précède l'attribut. Ceci est garanti par la LP (7.a).

(7) a- [SUBCAT إنّ]< N2[Nom]

IV.5 Les structures récursives

Dans certaines constructions et pour des raisons purement sémantiques, plusieurs constituants peuvent être déplacés. Dans certains cas ce changement d'ordre n'entraîne aucun changement de structure et le format ID/LP rend bien compte de cela. Dans d'autres situations ce changement d'ordre entraîne bien un changement de structure ce qui correspond à la transformation d'emphase dans certains cas. Les structures obtenues sont récursives.

IV .5.1 Transformations dans le cas des constructions syntaxiques verbales

Dans une construction syntaxique verbale les éléments qui peuvent être déplacés sont le sujet (1.a) ou le complément (1.b).

(1)a-اصْطَادَ الأَوْلاَدُ السَّمَكَةَ ⇒ c- الأَوْلاَدُ اصْطَادُوا السَّمَكَةَ

b- اصْطَادَ الأَوْلاَدُ السَّمَكَةَ ⇒ d- السَّمَكَةُ اصْطَـادَهَا الأَوْلاَدُ

Dans ce type de structures on note la présence d'un pronom affixe dans la séquence qui suit l'élément déplacé et qui s'accorde en genre, en nombre et en cas avec ce dernier. La séquence résultant de cette transformation sera considérée comme :

-une construction syntaxique nominale dont le sujet est l'élément déplacé et l'attribut une construction syntaxique verbale, si l'élément déplacé est une lexie définie.

-une lexie nominale dont le noyau est l'élément déplacé si ce dernier est indéfini.

Nous considérons dans ce qui suit le cas où la séquence résultante est une construction syntaxique nominale.

A-Cas du déplacement du sujet

Dans ce type de constructions, l'information concernant le sujet déplacé est contenue dans la forme morphologique du verbe. Pour rendre compte de ce fait nous utilisons le trait de tête, à valeur booléenne, SUJ qui indique si le verbe est à l'infinitif ou non.

Les constructions relevant de ce cas seront données par la règle ID (2)

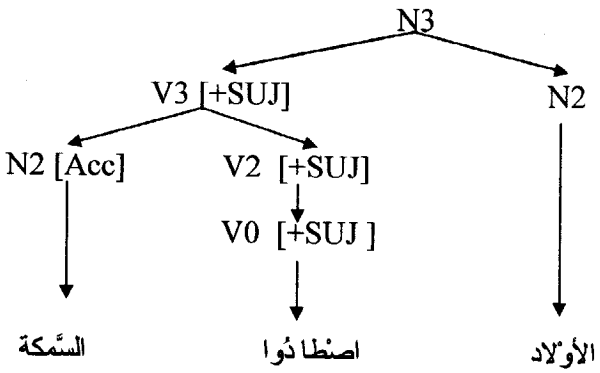
(2)N3 → N2, V3 [+SUJ]

V3[+SUJ] est la construction syntaxique verbale dont le sujet a été déplacé. Sa structure est donnée par (3)

(3) V3 [+SUJ] → V2,N2 [Acc]

Dans ce cas V2 portera le trait [+SUJ] en vertu du HFC, ce qui garantit que l'information concernant le sujet déplacé soit contenue dans la forme morphologique du verbe. La phrase (1.c) aura la structure :

(4)



B-Cas du déplacement du complément

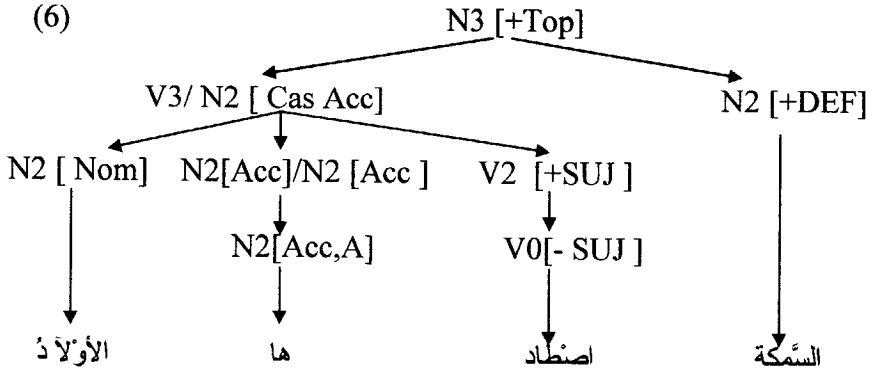
Le cas où le complément est déplacé est donné par la règle ID (5) introduisant le trait SLASH et qui nous donne un arbre tel que (6)

(5) a- N3 [+Top] → N2 [+DEF],V3/N2 [Acc].

Le trait SLASH est un trait à valeur catégorielle introduit par GPSG pour le traitement des dépendances non bornées. Il s'agit d'un trait de tête et de pied.

La notation V3/N2 [Acc]est une abréviation de V3 {SLASH N2 [Cas Acc]}qui signifie que dans la construction syntaxique verbale l'élément dont le Cas est Accusatif est déplacé hors de sa séquence matrice.

(6)



Bien que SLASH soit un trait de tête, il ne peut apparaître sur la tête de V3 qui estV2 à cause de la FCR (7) qui signifie que la catégorie V2 ne peut porter le trait SLASH.

(7) FCR : V2 ⊃ ~ [SLASH]

SLASH étant un trait de pied, le FFP requière que le trait [SLASH N2] instancié sur une catégorie mère doit l'être sur un constituant fille aussi. Il marquera le N2 [Acc]et non l'autre N2 grâce à la FCR (8) qui impose la même valeur du trait Cas pour le trait SLASH et la catégorie qui le porte.

(8) FCR: [SLASH [Acc]] ⊃ N2[Acc]

Le nœud N2 [Acc]/ N2 [Acc] se termine par un pronom affixe ce qui est garanti par (9).

(9) N2 [Acc]/ N2 [Acc] → N2 [+P, Acc]

IV.6 Les relatives

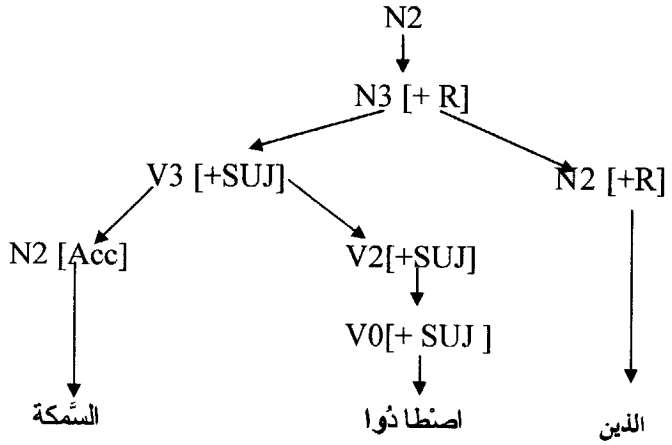
Les relatives ont le statut de lexies nominales définies caractérisées par le trait de tête, à valeur booléenne R. Nous considérons donc un relatif comme la catégorie N2 [+R].Pour les séquences telles que données en (1) nous proposons la règle (2).

(1) a- السمكة الذين اصطادوا b - الأولاد التي اصطادها

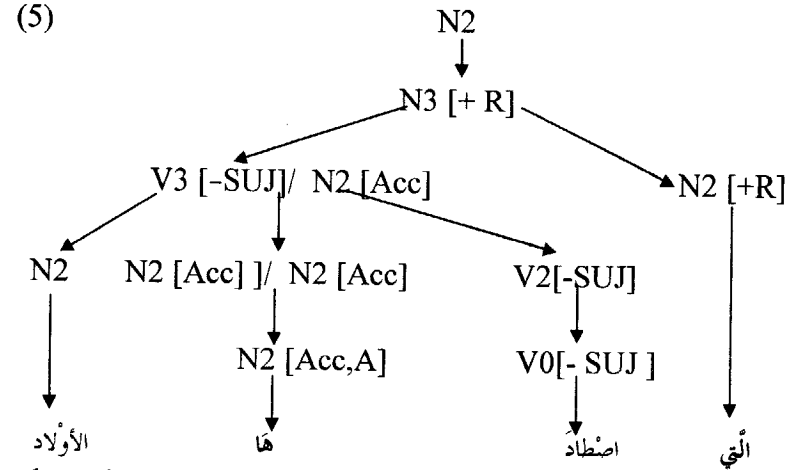
(2) N2 → N3 [+R]

La FCR (3) :[+R] ⊃ [+DEF] garantie qu'une relative est équivalente à une lexie nominale définie. R étant un trait de tête, donc partagé par N3 et sa tête qui est N2 grâce au HFC, ce qui donne les structures (4) et (5) respectivement à (1.a),(1.b)

(4)



(5)



Pour les séquences telles qu'en (7) et (8) nous avons besoin des règles respectives (6.a) et (6.b)

(6) a- N2 → N1 [+DEF], N3 [+R]
b- N2 → N1 [-DEF], N3 [+R]

(7) a- الأوْلاَدُ **الذين** اصْطَادوا السَّمَكَةَ b- السَّمَكَةَ **الّتي** اصْطَادَهَا الأوْلاَدُ

(8) a- السَّمَكَةَ **الذين** اصْطَادوا صَديقُ b- أبُ **الّتي** ضَربها الوَلَدُ

Cette différence de structure qui entraîne aussi une différence sur le plan sémantique, est due au fait que dans le premier cas la relative occupe la position du caractérisant, dans le second elle a le rôle de complément adnominal.

V. Conclusion

La GPSG est un modèle formel de la syntaxe du langage naturel qui a apporté deux innovations par rapport aux autres modèles. La première consiste à faire la distinction entre les règles de dominance immédiate (ID) et les règles de précédence linéaire (LP) la seconde est l'introduction des traits à valeur catégorielle pour le traitement des dépendances non bornées tout en restant dans un système de règles contextuelles. GPSG fournit donc une représentation compacte du langage naturel grâce aux puissants mécanismes qu'elle offre et que nous avons essayé d'exploiter pour la modélisation des structures syntaxiques de la langue arabe. Nous avons pu voir que le format ID/LP s'adapte bien pour les phrases verbales de l'arabe où l'ordre des constituants est assez libre. La puissance des structures de traits ainsi que les principes universels d'instanciation des traits de GPSG permettent d'exprimer des généralisation couvrant diverses structures évitant, ainsi beaucoup de redondances dans la grammaire. De plus, comme ses principes sont indirectionnels, GPSG peut servir lors de l'implantation d'un système TAL, pour la construction d'une représentation syntaxique d'une phrase, aussi bien pour l'analyse que pour la génération.

Références

[BEN 95] Paul BENNET. "A Course in Generalized Phrase Structure Grammar". Studies in Computational Linguistics. First published in 1995 by University College London (UCL) press in association with the Center for Computational Linguistics.

[BUS 87] S.BUSMANN. "Ein parser fur GPSG", rapport de l'université technique de Berlin.

[BUS 88] S.BUSMANN and C. HAUENSCHILD. "A constructive view of GPSG or how to make it work", 12th international conference on computational linguistics Budapest 1988.

[GAZ 85] Gazdar Gerald, KLEIN Ewan, PULLUM Geoffrey and SAG Ivan. "Generalized Phrase Structure Grammar", Oxford : Basl Blackwell, 1985.

[GUN87]: GUNJI Takao. "Japenese Phrase Structure Grammar, Dordrecht": Reidel.

[HAD 79] HADJ-SALAH A. "Linguistique arabe et linguistique générale, Essai de méthodologie et d'épisthomologie du ilm al-Arabiyya", 2 volumes, 1979.

[MIL 90] Philip MILLER et Thérèse TORIS "Formalismes syntaxiques pour le traitement automatique du langage naturel", Editions HERMES, Paris, 1990.

[SAB 89] Gerard SABAH. "L'intelligence artificielle et le langage naturel", Tome1, Editions HERMES, Paris, 1990.

[WEI 88] Wilhelem WEISWEBER. "Using constraints in constructive version of GPSG", in Procs 12 th COLLING-88, Budapest.

[VOL 88] Martin VOLK. "Parsing German with GPSG : The problem of separable prefix verbs". A Thesis submitted to the graduate faculty of university of Georgia in partial fulfillment of the requirement for the degree MASTER OF SCIENCE, ATHENES, GEORGIEA, 1988.

[USZ 87] USZOREIT Hans. "Word Order and Constituent Structure in German", STANDFORD California: CSLI (=CSLI Lecture notes N° 8) distribué par Chicago, University Press, 1987.