HAL
open science

Analyse des dépendances à distance à l'aide de graphes de contraintes

Philippe Blache

To cite this version:

Philippe Blache. Analyse des dépendances à distance à l'aide de graphes de contraintes. Revue TAL : traitement automatique des langues, 2003, 43 (3), pp.63-83. hal-00134204

HAL Id: hal-00134204

https://hal.science/hal-00134204

Submitted on 1 Mar 2007

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.

Analyse des dépendances à distance à l'aide de graphes de contraintes

Le cas des disloquées

Philippe Blache[1]

Laboratoire Parole et Langage (CNRS - Université de Provence)
29 Avenue Robert Schuman
F-13621 Aix-en-Provence
pb@lpl.univ-aix.fr

RÉSUMÉ. *Cet article propose une description des dépendances à distances s'appuyant sur une approche totalement déclarative, les* grammaires de propriétés, *décrivant l'information linguistique sous la forme de contraintes. L'approche décrite ici consiste à introduire de façon dynamique en cours d'analyse de nouvelles contraintes, appelées* propriétés distantes. *Cette notion est illustrée par la description du phénomène des disloquées en français.*

ABSTRACT. *We propose in this paper an account of long-distance dependencies relying on a fully declarative approach, called* property grammars, *describing linguistic information by means of constraints. The technique presented here consists in generating dynamically new constraints during the parse. Such constraints are called* distant properties. *We illustrate this notion with an analysis of dislocated constructions in French.*

MOTS-CLÉS : *dépendances à distance, disloquées, grammaires de propriétés, contraintes, analyse syntaxique.*

KEYWORDS: *long-distance dependencies, dislocated, property grammars, constraints, parsing.*

1. Cet article est une version étendue de [Blache01b] présenté lors de la conférence TALN-01 et sélectionné par le comité de programme pour publication dans la revue TAL.

1. Introduction

Les dépendances à distance posent un problème important pour l'analyse syntaxique non seulement du point de vue théorique, mais également concernant les mécanismes nécessaires pour leur implantation. Plusieurs théories ont proposé des solutions, en s'appuyant en particulier sur la propagation d'informations à travers la structure syntaxique. C'est le cas par exemple de GPSG qui propose une analyse précise des constructions clivées (cf. [Gazdar85]) ou encore HPSG (cf. [Pollard94], [Sag99]) dans lesquelles ce mécanisme est traité par la propagation d'un trait particulier. Ce trait est introduit par certaines règles et sa propagation est contrôlée par un principe spécifique, jusqu'à rencontrer le lieu de la dépendance. Dans ces approches, la relation n'est pas spécifiée directement entre deux catégories, mais par la transmission de ces valeurs. C'est la raison pour laquelle l'utilisation d'une catégorie vide a souvent été proposée comme solution naturelle à la représentation de ces constructions (c'était le cas en GPSG et dans les premières versions de HPSG).

Sans entrer dans les détails d'une présentation comparative qui dépasse le cadre de cet article, on peut toutefois rappeler les principes fondamentaux qui gèrent l'analyse de ces dépendances non bornées. En GPSG, l'analyse repose sur un trait particulier (le trait SLASH) dont l'introduction est assurée par une DI-règle, la propagation par une principe et le contrôle par une métarègle. Par exemple, la topicalisation utilisera la DI-règle : $P \rightarrow X^2, H/X^2$. Le *principe des traits de pied* assure quant à lui la propagation du trait en indiquant que l'ensemble des traits de pied d'une projection est l'unification des traits de pied de ses constituants. La *métarègle de fin de slash* permettant de contrôler l'arrêt de la propagation se décrit enfin comme suit : $X \rightarrow W, X^2 \Rightarrow X \rightarrow W, X^2$[+NULL]. Pour HPSG (dans sa version de 1994), le mécanisme est similaire dans l'idée et consiste également à introduire un trait particulier, à assurer sa propagation et contrôler son arrêt. Le sens de la propagation est cependant différent puisque le trait est introduit par la sous-catégorisation qui instancie le trait SLASH (éventuellement *via* une *trace*) qui sera propagé de façon ascendante par le *principe des traits non locaux*. Le liage de la dépendance se fait grâce à la distinction de deux traits spécifiques réglant l'héritage de ce trait non local.

L'analyse de ce phénomène est considérablement simplifié dans une évolution récente de HPSG (cf. [Bouma01]) qui propose une représentation explicite de la structure argumentale et de la structure de dépendance. On est ainsi en mesure d'indiquer la possibilité d'une catégorie présente dans la structure de dépendance, mais pas dans la structure argumentale qui spécifie la structure effectivement réalisée. Cette représentation permet de plus de simplifier la propagation des traits qui est dans ce cas uniquement assurée par le principe des traits de tête.

Dans le cadre des grammaires de dépendance, l'analyse des dépendances à distance pose notamment le problème de la projectivité. Une solution proposée dans [Kahane00] consiste à spécifier la possibilité de dépendances à distance dans des structures élémentaires (à la TAG) et utiliser la notion de bulle (représentant un nucléus)

permettant d'établir cette relation non pas entre des mots (ou des catégories lexicales), mais entre des bulles (éventuellement entre une bulle et un mot).

Nous proposons dans cet article une approche permettant de spécifier directement des dépendances entre deux catégories distantes. Nous nous situons pour cela dans le cadre des grammaires de propriétés (cf. [Blache01a]) qui présentent la particularité de décrire l'information linguistique sous la forme d'un ensemble de relations (des contraintes) entre les catégories. Cette approche permet en particulier d'abandonner la notion de constituance problématique à bien des égards et notamment pour le traitement de ce type de phénomène. Un des aspects intéressants de ce formalisme pour le traitement du problème des dépendances à distance réside dans la possibilité de représenter la structure linguistique sous la forme d'un graphe plutôt que d'un arbre. Il devient ainsi possible de représenter des relations entre deux nœuds quelconques (donc deux catégories), quelle que soit leur position dans le graphe. D'une façon générale, l'utilisation d'une représentation de l'information sous la forme de graphes offre des perspectives très intéressantes, non seulement pour la description de phénomènes de dépendance complexe, mais également le traitement des non-attendus dans l'analyse de textes tout-venant. Nous verrons dans la suite comment tirer parti de cet aspect.

La première partie de cet article présente rapidement le cadre des grammaires de propriété et décrit leur interprétation sous la forme de *graphes de contraintes* : chaque catégorie est spécifiée par un ensemble de contraintes, celui-ci correspondant à un graphe. La deuxième partie est consacrée à la description d'un phénomène de dépendance à distance complexe (et très fréquent tant à l'écrit qu'à l'oral) : la dislocation. La troisième partie est consacrée à l'analyse de ce dispositif syntaxique dans le cadre des grammaires de propriétés en proposant l'introduction d'un type de propriété particulier : les *propriétés distantes*. Nous montrons dans cette même section comment ce mécanisme peut être utilisé pour l'analyse de dépendances non bornées comme les clivées.

2. Les grammaires de propriétés : un réseau de contraintes

Les *grammaires de propriétés* reposent sur la représentation de toutes les informations linguistiques sous la forme de propriétés qui sont également considérées comme des contraintes sur la structure linguistique. On trouvera une présentation de cette approche dans [Bès99], [Blache00] ou [Blache01a]. Nous en rappelons ici rapidement les grandes lignes et proposons une interprétation particulière de cette théorie : une grammaire de propriété peut en effet être considérée comme un réseau de contraintes et représentée sous la forme d'un graphe. Nous exploiterons cette caractéristique pour traiter le problème abordé ici des dépendances distantes : la représentation de l'information syntaxique sous forme de graphe plutôt que d'arbre permet de spécifier directement des relations distantes sans avoir recours à des mécanismes procéduraux de propagation d'information tout en éliminant totalement et naturellement le recours à des catégories vides.

Les informations syntaxiques reposent en grammaire de propriétés sur différents types de propriétés. Chacune d'entre elles représente un type d'information homogène et met en relation des catégories sans préjuger de leur réalisation. En d'autres termes, à la manière des contraintes de préséance linéaire en GPSG, ces contraintes indiquent une propriété particulière au cas où les catégories concernées se rencontrent dans la structure. Il s'agit d'une conception ouverte de la représentation de l'information syntaxique qui spécifie des propriétés mais n'empêche pas la réalisation de structures en dehors du cadre décrit : une structure donnée (en fait un ensemble de catégories) peut être décrit par les contraintes satisfaites ou non satisfaites mais il est également possible d'imaginer une situation dans laquelle aucune contrainte de la grammaire n'est pertinente. En d'autres termes, à la différence des approches génératives classiques, la construction d'une structure syntaxique n'est pas une étape préliminaire indispensable pour la vérification des propriétés linguistiques.

La représentation des informations séparément pour chacune des propriétés répond donc à la même logique que celle ayant présidé à l'introduction du formalisme DI/PL (cf. [Gazdar85]). L'idée consiste à représenter de façon distincte chaque type d'information syntaxique. Les grammaires de propriétés systématisent ainsi la démarche adoptée par le formalisme DI/PL qui avait proposé de représenter séparément seulement les informations linéaires et hiérarchiques. Mais un certain nombre d'informations implicites sont toujours contenues dans les règles syntagmatiques (ou les schémas de règle) : constituants obligatoires, facultatifs, uniques, relations de dépendance, etc. Nous proposons ici un abandon définitif des règles syntagmatiques, toutes les informations étant contenues dans les six types de propriétés suivantes (leur notation est indiquée entre parenthèses) :

– **Obligation** (*Oblig*) : indique un ensemble de catégories obligatoires dont la présence est indispensable et ne dépend pas d'autres catégories[1]. Ces catégories sont généralement en exclusion mutuelle. Cette information est alors spécifiée dans la propriété d'exclusion.

– **Unicité** (*Unic*) : ensemble de catégories uniques dans un syntagme. Comme dans le cas de l'obligation, cette propriété porte sur un ensemble de catégories.

– **Exigence** ($\Rightarrow$) : cooccurrence obligatoire entre ensembles de catégories. Le fait de pouvoir spécifier des ensembles dans cette propriété permet une description contextuelle des relations.

– **Exclusion** ($\not\Rightarrow$) : impossibilité de cooccurrence entre ensembles de catégories.

– **Linéarité** ($\prec$) : contraintes de précédence linéaire.

– **Dépendance** ($\rightsquigarrow$) : relations de dépendance entre catégories. Il s'agit d'une relation sémantique permettant d'indiquer le sens de la composition des contenus sémantiques des catégories. Au sein d'un syntagme, ces relations de dépendance ont généralement pour cible la même catégorie, considérée comme la tête sémantique.

1. Ces catégories correspondent dans les grammaires syntagmatiques à la notion de tête (ou noyau).

Mais il est également possible de représenter des relations de dépendance entre deux catégories non obligatoires.

Plusieurs remarques peuvent être faites sur ces propriétés. Tout d'abord, à la manière des contraintes en programmation, elles peuvent jouer différents rôles. Elles ont ainsi un pouvoir filtrant permettant d'empêcher la construction de certaines structures. C'est typiquement le cas des propriétés d'exclusion. Ce rôle est caractéristique des contraintes, c'est d'ailleurs l'usage exclusif qui est fait de cette notion dans d'autres approches comme la *théorie de l'optimalité* (cf. [Prince93], [Kager99]). Celle-ci considère en effet que les contraintes jouent un rôle de filtrage permettant de sélectionner une structure optimale parmi un ensemble de candidats. On retrouve également cette utilisation filtrante des contraintes dans le cadre des grammaires de dépendance par contraintes (cf. [Maruyama90]) qui consistent à générer un ensemble maximal de relations de dépendance puis à le filtrer en éliminant les relations non valides. Plusieurs heuristiques sont d'ailleurs généralisables, y compris dans le cadre des grammaires syntagmatiques (cf. [Blache99]). Les grammaires de propriétés font cependant un usage plus général des contraintes en tirant également parti du rôle positif qu'elle peuvent jouer, notamment dans leur capacité d'instanciation : certaines contraintes peuvent également ajouter de nouvelles informations (comme les contraintes de dépendance qui jouent les deux rôles décrits).

Par ailleurs, toutes les informations représentées par les propriétés sont situées au même niveau. Ceci est rendu possible en particulier grâce à l'abandon des règles syntagmatiques et des schémas de règles. Là où les autres approches font appel à des mécanismes particuliers complétant ou vérifiant la construction d'un arbre (par exemple les restrictions de cooccurrence de traits en GPSG ou les règles lexicales en HPSG), les propriétés prennent ici en charge la spécification de toutes les informations. De ce point de vue, l'approche n'est pas lexicalisée au sens où HPSG l'entend. Bien entendu, les entrées lexicales contiennent un grand nombre d'informations. Mais toutes les informations de niveau syntaxique, concernant en particulier les relations entre les catégories, sont exprimées par des propriétés.

De plus, on peut remarquer qu'aucune de ces propriétés ne rend compte de la notion de *constituance*. Même si une organisation hiérarchisée de l'information syntaxique est préservée dans cette approche (à la différence des grammaires de dépendance), il n'existe pas de relation directe entre une catégorie de niveau syntagmatique et un ensemble de catégories participant à sa construction. En d'autres termes, la notion de syntagme reste pertinente pour les grammaires de propriétés, en revanche la notion de règle syntagmatique est totalement abandonnée. La seule relation de hiérarchie préservée est celle d'*obligation* qui relie une catégorie avec la catégorie syntagmatique dont elle est la tête. Mais auncune autre catégorie (typiquement celles sélectionnées par cette tête) n'entretiennent de lien avec le niveau syntagmatique, les relations étant exprimées directement entre catégories. La notion de constituant à proprement parler n'est donc plus pertinente ni représentée. Cette particularité introduit une grande souplesse dans la description : il devient en effet possible de rendre compte de l'apparition de certains éléments n'entretenant pas de relation directe avec le reste

de la structure. Ce type d'élément est très fréquent, en particulier à l'oral : phatiques, incises, bribes, associés, etc. (cf. [Blanche-Benveniste97]). Il s'agit d'ailleurs d'un argument supplémentaire en faveur d'une représentation des informations sous la forme de graphe qui, par définition, n'est pas nécessairement connexe.

Toutes les propriétés sont exprimées à l'aide de contraintes, donc de relations, entre plusieurs catégories. Il est ainsi possible de représenter la description d'une catégorie par un ensemble de relations : chaque catégorie syntagmatique est décrite par un ensemble de propriétés qui se contentent de mettre en relation d'autres catégories, sans aucune référence au syntagme qu'elles décrivent. Plus généralement, l'ensemble des propriétés décrivant les syntagmes forme un système de contraintes (qui n'est rien d'autre que la grammaire). En d'autres termes, le système de contraintes forme un tout, il n'est pas organisé en sous-systèmes indépendants (correspondant à des catégories) et évaluables séparément. La vérification de la satisfaisabilité d'un ensemble de catégories donné se fait toujours sur la totalité du système de contraintes. Cependant, dans la mesure où, dans cette approche, l'information syntaxique est organisée de façon hiérarchisée, il est nécessaire d'indiquer les conditions d'instanciation des catégories syntagmatiques. Pour cela, chaque catégorie syntagmatique est attachée à un ensemble de contraintes qui la décrit. L'évaluation du système de contraintes pour un ensemble de catégories donné permet de spécifier pour chaque propriété de la grammaire si elle est satisfaite, non satisfaite ou indéterminée (la propriété n'est pas pertinente pour l'ensemble de catégories analysé). L'état du système de contraintes permet alors de voir quels sous-ensembles de contraintes sont activés (*i.e.* contiennent des propriétés évaluées). Dans ce cas, il sera possible d'instancier la catégorie décrite par ce sous-ensemble de contraintes.

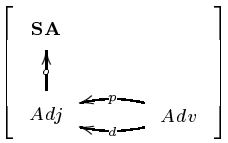
Un sous-ensemble de contraintes décrivant une catégorie peut être vu comme un graphe dont les nœuds sont composés par des catégories et les arcs par les propriétés qui les relient. Une grammaire peut ainsi être vue comme un ensemble de sous-graphes, chacun constituant un système de contraintes associé à une catégorie. Ces graphes, appelés *graphes de contraintes*, peuvent être connectés entre eux (par exemple par une relation de dépendance). L'analyse consiste à trouver une affectation (*i.e.* un ensemble de catégories) pouvant être associée à une phrase à analyser et dont le statut est donné par les graphes de contraintes associés.

La figure [1] présente un ensemble de propriétés décrivant une partie du syntagme adjectival en français et sa représentation sous forme de graphe de contraintes. Chaque propriété est représentée par un arc orienté, le type de la propriété étant indiquée par l'étiquette de l'arc[2]. Par exemple, l'arc $Adj \xleftarrow{p} Adv$ représente la contrainte de précédence linéaire *Adv* $\prec$ *Adj*. La contrainte d'obligation relie quant à elle la catégorie pouvant être la tête avec le syntagme décrit. Comme indiqué précédemment, c'est la seule relation de hiérarchie spécifiée dans les grammaires de propriétés. On indique pour des raisons de lisibilité la catégorie syntagmatique en gras, mais celle-

2. Les étiquettes suivantes sont utilisées : o pour obligation, u pour unicité, e pour exigence, x pour exclusion, p pour précédence linéaire et d pour dépendance.

ci n'a pas de statut particulier autre que celui de pouvoir être la cible d'une relation d'obligation.

Sous graphe du SA :



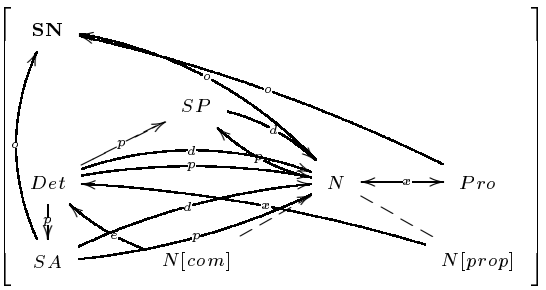
[1]

Une catégorie est donc décrite par un graphe dont une des relations exprime une propriété hiérarchique qui permet d'introduire une représentation arborescente au sein du graphe. La figure [2] propose un exemple plus complet de graphe décrivant le *SN*. Il représente l'ensemble de propriétés suivant :

- *Linéarité :* $Det \prec N$; $Det \prec SA$; $N \prec SP$
- *Exigence :* $N[com] \Rightarrow Det$
- *Exclusion :* $N \not\Leftrightarrow Pro$; $N[prop] \not\Leftrightarrow Det$
- *Dépendance :* $Det \rightsquigarrow N$; $SA \rightsquigarrow N$; $SP \rightsquigarrow N$
- *Obligation :* $Oblig(SN) = \{N, Pro, SA\}$

On indique dans cet ensemble de propriétés la possibilité pour *SA* de faire partie des catégories obligatoires du *SN*. On peut ainsi rendre compte de tournures du type *« je veux la rouge »* sans avoir recours à la notion d'ellipse ou à la modification de la catégorie de l'adjectif.

Sous graphe du SN :



[2]

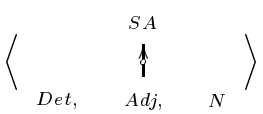
On remarque dans ce schéma l'utilisation d'une relation de sous-typage entre les catégories N, *N[com]* et *N[prop]*. Cette relation est indiquée par un lien en pointillés. De façon classique, toutes les propriétés portant sur un type sont héritées par les sous-types. Cette caractéristique permet de décrire de façon précise et compacte les relations syntaxiques.

Dans une grammaire de propriétes, chaque description d'unité syntaxique correspond à un *graphes de contraintes* (cf. [Blache02]). L'unité syntaxique concernée, qui est la cible de la relation d'obligation dans ce graphe, forme le label du graphe. Une grammaire est donc constituée par un ensemble de graphes de contraintes.

L'analyse syntaxique dans le cadre des grammaires de propriétés s'appuie donc sur un processus de satisfaction de contraintes. En termes de programmation, un tel processus consiste à trouver une *affectation* pour un ensemble de variables qui satisfait un système de contraintes donné. Dans notre cas, une affectation est formée en cours d'analyse par un sous-ensemble des catégories (ou plus exactement une suite de catégories) associées aux mots formant la phrase à analyser. Pour tout ensemble de catégories, il est possible de déterminer sa description en évaluant le système complet de contraintes (*i.e.* la satisfaisablitié de cette affectation dans la grammaire). Dans toute la grammaire, seules certaines contraintes sont effectivement évaluables (on dira qu'elles sont *pertinentes* pour l'affectation concernée). L'identification des graphes de contraintes décrivant une catégorie syntagmatique se fait simplement grâce aux contraintes évaluables : tout graphe de contrainte contenant une ou plusieurs contraintes évaluables est considéré comme *actif*. Dans la plupart des cas, en cours d'analyse, un seul graphe de contraintes est actif pour une affectation. L'état d'un graphe de contraintes actif (indiquant ses contraintes satisfaites, celles qui ne le sont pas ainsi que les contraintes non évaluables) constitue une description précise des propriétés syntaxique de l'affectation concernée. Nous appelons cette description une *caractérisation*[3], le graphe de contraintes évalué formant quant à lui un *graphe de description*. On parlera également de catégorie syntagmatique *caractérisée* par une affectation pour indiquer le label du graphe de description. En cours d'analyse, les catégories caractérisées par une affectation seront instanciées pour faire à leur tour partie de nouvelles affectations.

L'exemple suivant considère l'analyse du *SN* « *le vieux livre* ». L'ensemble de catégories considéré est formé par $\mathcal{E}$={*Det, Adj, N*}. Parmi toutes les affectations vérifiées (tous les sous-ensembles $\mathcal{E}$), nous illustrons pour commencer le cas de l'affectation $\mathcal{A}1 = \langle Adj \rangle$. Le graphe de description décrivant *SA* correspondant à cette affectation est illustré dans [3]. La catégorie *SA* étant caractérisée par un graphe de description, elle est instanciée et fait alors partie de l'ensemble des catégories pouvant être utilisé dans une nouvelle affectation.

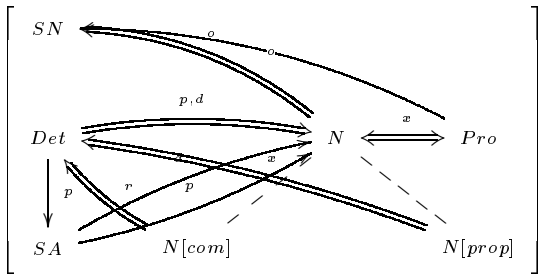
[3]



La seconde étape repose sur la caractérisation de l'affectation $\mathcal{A}2 = \langle Det, SA, N \rangle$. Le graphe de contraintes pertinent décrit un *SN*. Dans ce cas, comme pour l'étape précédente, toutes les contraintes pertinentes sont satisfaites. On obtient ainsi le graphe de description décrit en [4].

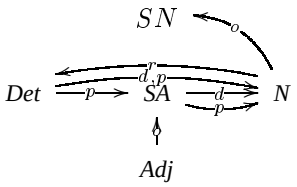
3. La caractérisation correspond à l'ensemble des propriétés satisfaites (noté $\mathcal{P}^+$) et des propriétés non satisfaites (noté $\mathcal{P}^-$) d'une catégorie pour un ensemble de constituants donné.

Graphe de description du SN :



[4]

On représente dans ce graphe par une flèche double pleine les contraintes satisfaites et par une flèche simple les contraintes non évaluables. Pour des besoins de lisibilité, nous ne représentons dans ce qui suit que les contraintes évaluées. De plus, nous représentons une catégorie syntagmatique instanciée par un arc non orienté couvrant l'affectation correspondante et étiqueté par la catégorie concernée. Ainsi, le graphe de description [4] est représenté par le schéma [5].



[5]

Il est intéressant de remarquer qu'un tel graphe contient de façon explicite toutes les informations linguistiques représentées dans la grammaire. Il est donc possible d'extraire si nécessaire les sous-graphes représentant chaque propriété.

3. Les dispositifs disloqués

Les constructions disloquées font partie de ces dispositifs syntaxiques fréquents tant à l'oral qu'à l'écrit (cf. [Blasco99]) et qui posent un certain nombre de problèmes pour l'analyse syntaxique (qu'elle soit ou non automatique). En effet, les relations (syntaxiques et sémantiques) entre les différentes parties d'un énoncé contenant une partie disloquée peuvent être difficiles à établir. Nous donnons dans cette section une rapide description de cette tournure.

De façon informelle, le phénomène de dislocation permet de réaliser avant ou après une phrase[4] une unité syntaxique (le plus souvent un syntagme nominal). Les

4. Il est également possible de réaliser une dislocation à l'intérieur d'une phrase, comme dans « **Il** *voudrait bien,* **mon père**, *que j'aille là* ». A la différence d'une incise, il existe comme dans les autres cas de dislocation une reprise anaphorique.

exemples [1a-e] présentent des cas de dislocation à gauche, tandis que la dislocation à droite est illustrée par les cas [2a-c].

(1) a. **Les livres**, elles **les** avaient envoyés.
b. **L'Etat**, **c**'est moi.
c. **Moi**, **je** n'en veux pas.
d. **Le chocolat**, **moi**, **j**'adore **ça**.
e. La moto, **le guidon**, **il** est cassé.

On reconnaît parmi cette liste une déclinaison de l'exemple célèbre de Culioli (exemple [1.e], dont on dit qu'il n'était pas vraiment attesté).

(2) a. **Il** est parti, **Jean**.
b. **Ça** ne m'intéresse pas, **la polémique**.
c. Elle **l**'est assurément, **ravissante**.

Parmi les propriétés générales de cette tournure, on constate une reprise anaphorique par un clitique qui en précise généralement la fonction (cf. [LeGoffic93]). Ainsi, le clitique en [1.a] indique une fonction objet pour le *SN* disloqué, celui de [1.c] indique un sujet. La même remarque s'applique pour les deux types de construction. Par ailleurs, il est important de souligner que si la grande majorité des éléments en position détachée est de type *SN*, on peut toutefois rencontrer d'autres types de constituants : adjectifs (cf. [2.c]), adverbes ou complétives par exemple. D'une façon générale, le dispositif est donc assez régulier et consiste à ajouter à une phrase une unité syntaxique ayant généralement une reprise anaphorique dans la phrase. La relation syntaxique entre l'élément détaché et la phrase est donc ténue pour ne pas dire inexistante en dehors de la reprise.

Une distinction plus fine est proposée, à la suite de [Blanche-Benveniste84], dans [Blasco99]. Celle-ci introduit en quelque sorte une gradation dans la notion de référence. Dans certains cas, la relation de coréférence entre l'élément disloqué et le clitique est forte. Dans la mesure où les deux éléments marquent la même fonction syntaxique dans une seule construction verbale, on parle alors de *double marquage* [5]. Dans d'autres cas, la relation référentielle peut alors être considérée comme minimale, l'attribution de la même fonction syntaxique aux deux éléments n'est pas claire, et il peut y avoir différence d'accord entre le *SN* et le clitique comme dans l'exemple : « **Ils** *vous l'ont interdit*, **la mairie** *(Blasco, 139)*[6] ». Il est alors possible de considérer que le *SN* n'est pas dans la rection du verbe, mais adjoint au pronom clitique.

Une telle distinction, pour intéressante qu'elle soit, n'induit pas de différence d'analyse dans l'approche proposée dans cet article. Nous ne la conservons donc pas dans la suite en assimilant toutes les constructions disloquées à un phénomène de

5. Une relation de ce type serait par exemple exprimée en HPSG par un partage de structure.
6. On indique entre parenthèses la localisation de l'exemple dans [Blasco99].

double marquage et en excluant de cette construction les éléments simplement associés, sans relation de coréférence comme dans « *Son travail, elle est quand même tranquille* (Blasco, 139) ».

En entrant un peu plus dans le détail de ce dispositif, des phénomènes plus complexes apparaissent. C'est le cas en particulier lorsque plusieurs éléments sont détachés. On peut alors avoir une reprise par un clitique de chacun des éléments détachés, comme en [1.d]. Dans cet exemple, le premier *SN* « *le chocolat* » est repris par le clitique objet « *ça*" et le pronom « *moi* » est repris par le clitique sujet « *je* ». On remarque par ailleurs qu'il y a peu de contraintes d'ordre linéaire entre les différents éléments détachés dans le cas où tous les éléments ont une reprise dans la phrase[7]. En revanche, la situation est différente dans les cas où la partie détachée se présente également sous la forme de plusieurs termes apparemment de même niveau, mais qui constituent en fait une seule unité sans lien syntaxique marqué comme dans l'exemple [3.a].

(3) a. Ta moto, le guidon, il est cassé.

b. *Le guidon, ta moto, il est cassé.

Dans ce cas, la partie détachée forme un thème unique, la relation existant entre les deux *SN* étant simplement indiquée par l'ordre, le référent se trouvant dans la position la plus proche de la phrase, l'interprétation de cette tournure étant ‘‘*le guidon de ta moto* ». Dans ce type de construction avec un seul clitique référent et une seule unité thématique détachée, l'ordre des *SN* ne peut être modifié comme indiqué en [3.b].

Cependant, cette caractéristique ne peut être généralisée : dans de nombreux cas, on peut rencontrer un constituant détaché sans clitique référent dans la phrase mais sans pour autant que les *SN* détachés entretiennent un rapport direct. On se trouve alors dans le cas d'un détachement multiple mais pour lequel le référent d'un des éléments détachés n'est pas réalisé. Dans cette perspective, on pourra rapprocher les tournures de [4.a] et [4.b] pour lesquelles on peut considérer que les deux *SN* ont un lien sémantique de référence avec la valence du verbe recteur. En revanche, ces exemples se distinguent de [4.c] dans lequel le premier *SN* est construit par le second qui, lui, a un référent (clitique sujet).

Le problème dans ce cas vient du fait qu'aucun critère apparent ne permet de distinguer les constructions en dehors d'une analyse de la valence du verbe recteur : le *SN « ta sœur »* ne peut être ici complément de *« déchirer »*, l'interprétation ne peut être que celle du référent unique « *la robe de ta sœur* ». De même, dans [4.a], l'interprétation de deux *SN* disloqués est retenue (« *la robe ne va pas à ta sœur* »). En revanche, dans [4.b], il y a ambiguïté et les deux interprétations sont possibles (*« la robe ne va pas à ta sœur »* vs. *« la robe de ta sœur ne va pas »*).

(4) a. Ta sœur, la robe, elle lui va pas.

7. Cette caractéristique est moins vraie pour les détachements à droite : on a *« j'adore ça, moi, le chocolat »*, mais pas *« *j'adore ça, le chocolat, moi »*.

b. Ta sœur, la robe, elle va pas.

c. Ta sœur, la robe, elle est déchirée.

d. Paul, la maîtresse, il ne l'écoute pas.

e. *Paul, la maîtresse, il écoute pas.

f. Paul, la maîtresse, elle a une belle robe.

Dans la plupart des cas cependant, l'interprétation de la dépendance entre *SN* est favorisée si un seul clitique référent est réalisé. Cette interprétation est même impérative dans certains cas comme [4.e] dans lequel la réalisation du clitique coréférent avec *« la maîtresse »* est obligatoire. L'analyse de la valence et des propriétés syntactico-semantiques du verbe joue bien entendu un rôle prépondérant pour distinguer les cas [4.e] et [4.f].

Il faut enfin signaler une tournure proche de celles décrites, mais dans laquelle le lien sémantique avec la phrase est très indirect. C'est le cas des exemples suivants[8] :

(5) a. Le violon, il faut être doué.

b. Le piano, les doigts, ça a beaucoup d'importance.

Dans ces exemples, on a un dispositif de type *thème/propos* dans lequel on annonce tout d'abord un thème général, puis on spécifie une caractéristique. On peut, dans celle-ci, retrouver un détachement plus classique avec un référent clitique comme en [5.b]. Mais le premier constituant ne peut être analysé comme appartenant, de façon directe ou non, à la rection d'un autre élément de l'énoncé. On considère qu'il s'agit d'éléments adjoints à la totalité de la chaîne verbale.

4. Analyse en GP : les propriétés distantes

Nous proposons dans cette section une représentation des principales caractéristiques des disloquées dans le cadre des grammaires de propriétés. Nous considérons, à la différence de ce qui avait été proposé dans [Blache01b], que la construction disloquée ne forme pas une catégorie à part entière, mais que les catégories pouvant apparaître en position de dislocation sont ajoutées directement en tant que constituants possibles de la phrase. Ces catégories sont essentiellement des *SN* ainsi, plus marginalement, que d'autres catégories pouvant dans une certaine mesure avoir un fonctionnement nominal, au moins pour ce qui concerne la possibilité de reprise par un clitique. On trouve ainsi la possibilité de construire comme élément disloqué : des *SN*, des *SA*, des complétives et des infinitives. Toutes ces catégories peuvent être répétées, on ne précisera donc pas de contrainte d'unicité.

Dans la description de la catégorie correspondant à la phrase (notée P), nous indiquerons qu'en plus d'une construction de base formée d'un *SN* sujet et d'un *SV*, il

8. Ces exemples sont donnés par José Deulofeu.

est possible de construire un P emboîté accompagné d'autres catégories. En d'autres termes, la description de P est récursive car contenant une catégorie de même type, mais cette possibilité est contrôlée par certaines retrictions.

L'ensemble de propriétés suivantes constitue une première approximation de la description de P. Dans un premier temps, par souci de lisibilité, nous ne prenons en compte que la possibilité de dislocation du *SN* et du *SV*. Dans ce dernier cas, il ne peut s'agir que d'un *SV* à l'infinitif[9].

- *Obligation :* $Oblig(P) = \{SV, P\}$
- *Linéarité :* $SN \prec SV$; $SV \prec P$;
- *Exigence :* $P \Rightarrow \{SN \vee SV[inf]\}$
- *Exclusion :* $P \not\Rightarrow SV[tens]$; $SN \not\Rightarrow SV[inf]$
- *Dépendance :* $SN \rightsquigarrow SV[tens]$; $SN \rightsquigarrow P$; $SV[inf] \rightsquigarrow P$

Cet ensemble de propriétés décrit un certain nombre de phénomènes. Tout d'abord, la construction de base du P constituée de *SN* et *SV* est caractérisée par une contrainte d'obligation (le *SV* fait partie des constituants obligatoirement réalisés), de linéarité et de dépendance entre ces deux constituants. La possibilité de dislocation de *SN* et de *SV[inf]* est quant à elle prise en compte par la contrainte d'obligation qui prévoit la possibilité pour un P d'être une des catégories obligatoires d'un autre P. Dans ce cas, une contrainte d'exigence entraîne la nécessité de réalisation d'un élément disloqué [10]. Les contraintes d'exclusion interdisent la réalisation d'un P en même temps qu'un *SV* tensé ainsi qu'une cooccurrence d'un *SN* et d'un *SV* disloqué. La propriété de dépendance précise quant à elle la relation entre le sujet et le *SV* recteur du cas classique. Enfin, les propriétés de dépendances concernant les éléments disloqués établissent une relation entre ceux-ci et le P.

Ainsi que nous l'avons vu dans la section précédente, un des éléments essentiels de notre analyse de la dislocation réside dans la possibilité de coréférence entre l'élément disloqué et un clitique à l'intérieur de la phrase. Une telle relation s'exprime, dans le cadre des grammaires de propriétés, par une relation de dépendance. Mais il existe également d'autres relations entre l'élément disloqué et le clitique. Il est par exemple nécessaire de représenter les restrictions suivantes : les complétives ou les infinitives ne peuvent être référents que des clitiques *ce* ou *ça*, le *SA* n'apparaît qu'avec un clitique *le*, etc. De la même façon, l'analyse classique (quelquefois remise en cause, cf. [Blasco99]) consiste à dire que la fonction des *SN* disloqués est donnée par la forme du clitique. C'est le cas de l'exemple suivant dans lequel la fonction syntaxique est effectivement donnée par la forme du clitique et qui la transmet, par voie de conséquence, au *SN disloqué*.

(6) a. **La voiture**, il **la** trouve très bien.

b. **La voiture**, **elle** est très bien.

9. Par exemple : « *Chanter, c'est ça que je veux.* »
10. Cette contrainte introduit au passage la possibilité pour une propriété de mettre en relation des conjonctions (cas classique) mais également des disjonctions de constituants.

Ces caractéristiques (parmi d'autres) font donc intervenir des relations entre plusieurs catégories, ce qui correspond à la présentation des propriétés telle que nous l'avons donnée dans la première partie de cet article. Cependant, ces catégories (par exemple un *SN* disloqué et un clitique sujet) ne sont pas constituants de la même catégorie syntagmatique. Or, les propriétés ont été jusqu'à présent définies comme l'expression de relations entre des catégories appartenant au même ensemble de constituants (et formant un graphe de contraintes). Ce n'est pas le cas du type d'informations que nous voulons représenter ici. Il convient donc de proposer un mécanisme spécifique permettant de représenter ce nouveau type de relation.

Prenons le cas des restrictions existant entre les types de l'élément disloqué et celui des clitiques référents telles que nous venons de les évoquer. Cette information est représentée intuitivement par des propriétés d'exigence de la forme :

$$\begin{aligned} &\mathit{SA} \Rightarrow \{\mathit{le}\} \\ &\mathit{Compl} \Rightarrow \{\mathit{ce}, \mathit{ça}\} \\ &\mathit{SV[inf]} \Rightarrow \{\mathit{ce}, \mathit{ça}\} \end{aligned} \qquad [6]$$

Cependant, de telles propriétés ne peuvent figurer directement dans l'ensemble des contraintes décrivant la catégorie *P*. En effet, les clitiques indiqués ne font pas partie de l'ensemble de catégories entrant en jeu dans la description de P. Il convient donc de préciser la catégorie dans laquelle ces clitiques peuvent apparaître. En fonction des exemples donnés dans la section précédente, on remarque que ces clitiques peuvent être construits à n'importe quelle position à l'intérieur du *P* situé au même niveau que la disloquée. La première indication que nous pouvons donner contient donc l'information selon laquelle le clitique cible de la relation est un constituant (direct ou indirect) de ce P.

Afin de spécifier le clitique en relation avec l'élément disloqué, il est nécessaire de disposer d'une relation permettant de spécifier l'appartenance d'une catégorie à une affectation caractérisant un objet de niveau syntagmatique ou à une affectation caractérisant l'un de ses constituants. En d'autres termes, cette relation doit indiquer la possibilité pour un élément d'appartenir à un ensemble de catégories ou un ensemble caractérisant une de celles-ci. Il s'agit en d'autres termes de la *fermeture transitive* de la relation d'appartenance à une affectation. Par exemple, si le clitique référent est un complément du verbe, il appartient donc à un *SV* appartenant à son tour au *P* (cf. exemple [1-a]). Or, dans cet exemple, l'élément disloqué appartient à l'affectation $\mathcal{A}_i = \langle \mathit{SN}, \mathit{P} \rangle$. La seule possibilité d'indiquer une relation entre le *SN* disloqué et le clitique référent est donc l'utilisation de la fermeture transitive de la relation d'appartenance : $\mathit{Clit} \in \mathit{SV} \in \mathit{P}$[11].

Nous notons $\in^*$ la fermeture transitive de la relation d'appartenance, ce qui donne pour l'exemple précédent : $\mathit{Clit} \in^* \mathit{P}$. L'indication de la localisation du clitique réfé-

11. On note par abus de langage l'appartenance à la catégorie plutôt qu'à une affectation caractérisant cette catégorie. Ainsi, la notation $x \in P$ indique $\exists \mathcal{A}$ tq $\mathcal{A}$ caractérise P et $x \in \mathcal{A}$.

rent repose ainsi sur des informations contenues dans la caractérisation de P construite pendant l'analyse. Plus précisément, il s'agit de vérifier qu'une catégorie précise appartient à une affectation caractérisée (*i.e.* pour lesquelles il existe des propriétés pertinentes, donc évaluées) telle que celle-ci participe, directement ou non, à la caractérisation de *P*. Cette information est immédiatement accessible, il suffit pour cela de parcourir l'ensemble des affectations caractérisées en cours d'analyse telles que celles-ci sont élément de *P* ou d'un de ses constituants.

Plus généralement, la difficulté dans l'expression d'une relation de ce type vient avant tout de la difficulté de spécifier les éléments entrant en jeu dans cette relation. Le premier type d'information à représenter porte donc sur ce point qui consiste à préciser le type d'élément recherché. Nous appelons « *description* » cet aspect de caractérisation des catégories de la relation distante. Elle consiste à préciser, soient deux catégories c_1 et c_2, la possibilité pour c_2 d'appartenir à une affectation caractérisant une catégorie *SX* de même niveau que c_1. On appellera c_1 la *catégorie locale* et c_2 la *catégorie distante*. Plus formellement, on a :

> Soient les catégories c_1, c_2, SX_j, soit $\mathcal{A}_i$ une affectation telle que $\{c_1, SX_j\} \subset \mathcal{A}_i$, alors $\exists \mathcal{A}_j$ caractérisant SX_j telle que $c_2 \in^* \mathcal{A}_j$.

La relation distante représentée ici met en jeu une catégorie locale et une éloignée. La catégorie locale est décrite comme faisant partie d'une affectation, sa présence entraîne la recherche de la catégorie distante spécifiée comme nous venons de l'indiquer.

Reprenons l'exemple de restriction de sélection entre un *SA* disloqué et le type de clitique référent construit dans le P. Cette relation nécessite tout d'abord la spécification de la catégorie distante (*i.e.* le clitique) ainsi que la définition de la restriction elle-même (représentée par une contrainte d'exigence). Au total, la relation représentée est prise en charge par une propriété complexe, que nous appelons *propriété distante*. Il s'agit plus précisément d'une propriété d'implication dont la partie gauche contient la catégorie déclenchant la relation distante et la partie droite se décompose à son tour en deux parties : la première précisant la catégorie distante et la seconde contenant la propriété spécifiée entre la catégorie locale et la catégorie distante. On représente la partie droite de cette propriété distante à l'aide d'une structure de traits contenant deux éléments (appelés DESCRIPTIONS et PROPRIÉTÉS) correspondant à ces deux types d'information comme indiqué dans le schéma 7 :

$$\left\{SA, P\right\} \Rightarrow \begin{bmatrix} \textsc{Descriptions} \left[le \in^* P \right] \\ \textsc{Propriétés} \begin{bmatrix} SA \Rightarrow le \\ SA \rightsquigarrow le \end{bmatrix} \end{bmatrix} \qquad [7]$$

On indique donc dans cette propriété distante, appartenant à l'ensemble des propriétés décrivant *P*, que si un *SA* est réalisé dans une affectation contenant également un *P*, alors si un clitique de type *le* est réalisé dans une des affectations caractérisant *P* on a une relation d'implication et de dépendance entre *SA* et *le*.

Les relations exprimées entre les éléments disloqués et les clitiques sont en fait plus précises. Elles contiennent une partie syntaxique (l'accord et la fonction) mais également sémantique (la référence elle-même). Ces informations sont prises en charge par des propriétés de dépendance qui expriment la partie sémantique et permettent également de spécifier certains échanges de traits. Ainsi, les références entre plusieurs *SN* et plusieurs clitiques de P sont exprimées par la propriété distante suivante (dans laquelle le trait ACC représente l'accord et SEM les traits sémantiques) :

$$\left\{ \mathrm{SN}_1 \begin{bmatrix} \text{ACC } x_1 \\ \text{SEM } y_1 \end{bmatrix}, \mathrm{SN}_2 \begin{bmatrix} \text{ACC } x_2 \\ \text{SEM } y_2 \end{bmatrix}, \mathrm{P} \right\} \Rightarrow \begin{bmatrix} \textsc{Descriptions} \left[\left\{ \mathit{Clit}_1 \begin{bmatrix} \text{ACC } x_1 \\ \text{SEM } y_1 \end{bmatrix}, \mathit{Clit}_2 \begin{bmatrix} \text{ACC } x_2 \\ \text{SEM } y_2 \end{bmatrix} \right\} \in^* P \right] \\ \textsc{Propriétés} \begin{bmatrix} \mathit{SN}_1 \rightsquigarrow \mathit{Clit}_1 \\ \mathit{SN}_2 \rightsquigarrow \mathit{Clit}_2 \end{bmatrix} \end{bmatrix} \quad [8]$$

Cette propriété spécifie les relations entre deux *SN* appartenant à la même affectation que *P* et des clitiques réalisés dans une affectation caractérisant *P* ou l'un de ses constituants. Si l'accord et les traits sémantiques entre les *SN* et les clitiques sont compatibles, alors, on précise la relation de dépendance correspondante. Dans cette propriété, la vérification de l'unification des traits d'accord constitue un contrôle pour la sélection du clitique référent. La compatibilité des valeurs des traits sémantique joue un rôle fondamental. Elle est simplement ici décrite, comme pour l'accord, par une unification des traits. Ce mécanisme n'est bien entendu pas suffisant, il convient en effet de tenir compte de phénomènes de sémantique lexicale plus précis non décrits dans cet article.

Les informations proposées par cette propriété distante ne sont cependant pas suffisantes pour contrôler précisément la relation distante entre les différentes catégories. De plus, dans le cas des exemples [4] décrits en section précédente, la relation est différente selon que les *SN* disloqués peuvent être tous deux construits par le verbe recteur de P (exemple [4-a]) ou non (exemples [4-c], [4-f]). Dans le premier cas, il faut ajouter une information spécifant toutes les propriétés permettant de préciser cette contrainte. En première approximation (nous y reviendrons plus loin), nous devrons donc vérifier la conjonction de propriétés suivante : $SV_1 \in Oblig(P_1) \wedge V_1 \in Oblig(SV_1) \wedge Clit_1 \rightsquigarrow SV_1 \wedge Clit_2 \rightsquigarrow V_1$. Cette formule précise la situation du verbe recteur et la relation de dépendance entre le $Clit_1$ et le *SV* projection de ce verbe (relation sujet) ainsi que la relation de complément existant entre $Clit_2$ et ce verbe. Dans le second cas, il y a un effet d'emboîtement entre les *SN*, l'un étant construit par l'autre (avec un effet de complément de nom comme dans l'exemple [4-c]). Cet effet n'est pas présent si, comme dans l'exemple [4-b], les deux *SN* de la disloquée peuvent être construits par le verbe (on aurait dans ce cas une ellipse du second clitique référent). La distinction entre ces différents cas vient donc de l'analyse de la valence verbale : l'effet d'emboîtement entre les *SN* est obtenu (1) si un seul clitique référentiel est présent dans P et (2) si le verbe recteur de P ne peut contenir dans sa rection un *SN* du type de celui contenu dans la disloquée.

La description de la propriété distante doit donc préciser, en plus de la *description* de la caractérisation du syntagme contenant la catégorie distante, un certain nombre de propriétés des éléments de ce syntagme. Une propriété distante est donc finalement formée de trois parties :

– la description des caractérisations : permet de spécifier et de localiser les catégories distantes utilisées dans la description de la propriété. Il s'agit en quelque sorte d'une rubrique de déclaration [12] ;

– les conditions sur les catégories : précisent les préconditions d'application de la propriété ;

– les propriétés : ensemble de propriétés ajoutées par la propriété distante. Il est important de noter que ces nouvelles propriétés vont s'ajouter au système de contraintes décrivant la catégorie. Le statut de ces contraintes est cependant particulier en ceci qu'en cas d'incohérence, la propriété distante ajoutée est prioritaire sur la propriété statique.

La description des contructions disloquées à *SN* emboîtés se fait alors grâce à la propriété distante [9]. Cette propriété précise tout d'abord la localisation des différentes catégories utilisées. Comme dans le cas précédent, cette spécification se fait en indiquant les catégories locales concernées par la relation accompagnées de leurs caractéristiques d'accord et sémantique. La partie précisant les conditions supplémentaires d'application de cette propriétés spécifie deux conditions. Tout d'abord le *SN* n'appartient pas à la valence du verbe recteur. On utilise pour cette contrainte négative la notation **Prop* indiquant que la propriété *Prop* ne peut être satisfaite. Dans cet exemple, le fait pour le *SN* de ne pas appartenir à la valence de *V* est précisé par la non-satisfaction des propriétés d'implication et de dépendance [13]. La seconde condition précise la position respective des deux *SN* disloqués. Cette contrainte est importante dans la mesure où elle permet de préciser quel *SN* dépend de l'autre. La dernière partie de la propriété distante précise enfin la propriété à ajouter (ici les propriétés de dépendance). Au final, cette propriété indique que si la disloquée contient deux *SN*, si le P ne contient qu'un clitique référant au second *SN* et si le premier *SN* ne peut pas appartenir à la valence du verbe recteur, alors le premier *SN* dépend du second.

12. Il est possible dans cette rubrique de spécifier la localisation des catégories de manière absolue en spécifiant l'ensemble des propriétés (notamment d'implication et de dépendances) reliant la catégorie décrite avec d'autres ou au contraire de manière relâchée en indiquant simplement une relation éventuellement distante (fermeture transitive) entre cette catégorie et la racine.

13. Il est possible de préciser que la catégorie $Clit_1$ peut être réalisée mais n'appartient pas à la valence de V_1 en précisant, à la manière de l'exemple précédent, d'une part, la localisation du verbe dans le trait description par les propriétés $SV_1 \in Oblig(P_1) \wedge V_1 \in Oblig(SV_1)$ ainsi que la restriction suivante dans les traits conditions : $*\langle Clit_1 \rightsquigarrow SV_1;\ Clit_1 \rightsquigarrow V_1\rangle$.

$$\left\{ \mathrm{SN}_1 \begin{bmatrix} \text{ACC } x_1 \\ \text{SEM } y_1 \end{bmatrix}, \mathrm{SN}_2 \begin{bmatrix} \text{ACC } x_2 \\ \text{SEM } y_2 \end{bmatrix}, \mathrm{P} \right\} \Rightarrow \begin{bmatrix} \textsc{Descriptions} \begin{bmatrix} \left\{ \mathit{Clit}_1 \begin{bmatrix} \text{ACC } x_1 \\ \text{SEM } y_1 \end{bmatrix} \right\} \notin^* P \\ \left\{ \mathit{Clit}_2 \begin{bmatrix} \text{ACC } x_2 \\ \text{SEM } y_2 \end{bmatrix} \right\} \in^* P \end{bmatrix} \\ \textsc{Conditions} \begin{bmatrix} {}^*\langle SN_1 \rightsquigarrow V \,;\, V \Rightarrow SN_1 \rangle \\ SN_1 \prec SN_2 \end{bmatrix} \\ \textsc{Propriétés} \begin{bmatrix} SN_1 \rightsquigarrow SN_2 \\ SN_2 \rightsquigarrow \mathit{Clit}_2 \end{bmatrix} \end{bmatrix} \qquad [9]$$

Cet usage des propriétés distantes est généralisable à d'autres dépendances non bornées. Le problème de ce type de construction consiste en effet à représenter une relation (généralement d'exigence ou de dépendance) entre des éléments n'appartenant pas à la même catégorie. Le cas de dispositifs clivés est emblématique de ce type de construction. Comme pour les disloquées, il s'agit de spécifier une relation de dépendance entre un élément de la clivée et un élément recteur appartenant au P situé au même niveau que la clivée. Cette construction peut présenter un contrôle supplémentaire pour la localisation de la catégorie rectrice du syntagme clivé : la caractérisation de cette catégorie peut en effet contenir une contrainte d'exigence non satisfaite (par exemple dans les constructions de type *filler-gap*). C'est le cas de l'énoncé [7a] dans lequel le *SN* objet a été clivé.

(7) a. C'est un livre que Paul donne à Marie.

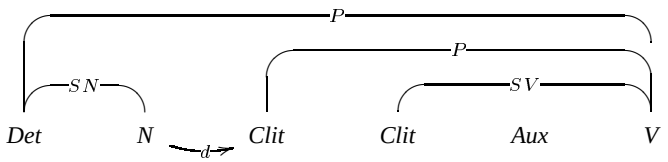
Dans ce cas, la caractérisation du *SV* comporte la propriété non satisfaite $V \Rightarrow SN$. Cette information pourra être utilisée pour la localisation de la catégorie concernée. La propriété distante correspondante, décrite dans [10], indique la relation de dépendance existant entre le syntagme clivé et sa catégorie rectrice dans la phrase, identifiée grâce à la description des caractérisations. Cette propriété fait partie de la liste des propriétés décrivant la clivée (qui constitue une catégorie à part entière). La specification de la dépendance lointaine se fait en identifiant un *SV* dans lequel une catégorie correspondante (avec les mêmes caractéristiques syntaxiques et sémantiques) n'a pas été réalisée.

$$\left\{ \mathrm{SN}_1 \begin{bmatrix} \text{ACC } x_1 \\ \text{SEM } y_1 \end{bmatrix} \right\} \Rightarrow \begin{bmatrix} \textsc{Descriptions} \begin{bmatrix} \left\{ \mathrm{V}_2 \right\} \in^* SV \\ \langle\, \mathrm{V}_2 \Rightarrow \mathrm{SN}_1 \,\rangle \in \mathcal{P}^-(\mathrm{SV}) \end{bmatrix} \\ \textsc{Propriétés} \left[SN_1 \rightsquigarrow V_2 \right] \end{bmatrix} \qquad [10]$$

Il est possible de proposer un schéma plus général pour l'ensemble des constructions clivées de ce type. Cette propriété indique qu'un élément clivé c_1 entretient une relation de dépendance avec une autre catégorie c_2 telle que c_1 est régi par c_2. La propriété distante correspondante peut ainsi être ajoutée à l'ensemble des propriétés décrivant les clivées et se présente comme suit :

$$\left\{ \mathrm{C}_1 \begin{bmatrix} \text{ACC } x_1 \\ \text{SEM } y_1 \end{bmatrix} \right\} \Rightarrow \begin{bmatrix} \textsc{Descriptions} \begin{bmatrix} \left\{ \mathrm{C}_2 \right\} \in^* P \\ \langle\ \mathrm{C}_2 \Rightarrow \mathrm{C}_1\ \rangle \in \mathcal{P}^-(\mathrm{SX}) \end{bmatrix} \\ \textsc{Propriétés} \left[c_1 \rightsquigarrow c_2 \right] \end{bmatrix} \qquad [11]$$

Les caractéristiques exprimées par les propriétés distantes correspondent ainsi à des relations entre des catégories lointaines. Elles sont simplement représentées par une arrête particulière dans le graphe reliant, comme indiqué dans le graphe [12], la catégorie disloquée et le clitique référent.



[12]

On remarque, de plus, qu'il est possible de spécifier des restrictions particulières en fonction de la forme des catégories concernées (*i.e.* aussi bien la catégorie extraite que celle contenant le syntagme extrait) : les phénomènes « d'ilôts » (par exemple de restriction d'extraction sur la conjonction) mais également l'utilisation du marquage morphologique de l'extraction dans certaines langues peuvent ainsi être directement exploités par la spécification de propriétés particulières.

5. Conclusion

Les grammaires de propriétés reposent sur une représentation par graphe plutôt que par arbre. Cette caractéristique offre un grand nombre d'avantages et notamment la possibilité de représenter des relations entre deux éléments quelconques du graphe (donc entre deux catégories quelconques). L'analyse des dépendances à distance s'en trouve ainsi facilitée. L'introduction de la notion de *propriété distante* décrite dans cet article permet dans ce cadre de préciser la forme et la localisation des catégories entrant en jeu dans une relation à distance. Elle permet, de plus, de spécifier si nécessaire un ensemble de caractéristiques formant des préconditions à l'application de la propriété elle-même. Nous disposons donc d'un mécanisme permettant de contrôler de façon très précise l'analyse de ces phénomènes.

Par ailleurs, la représentation de relations entre catégories n'appartenant pas à la même unité syntaxique est d'un grand intérêt non seulement pour le type de constructions décrit ici, mais également pour la représentation de relations entre propriétés de différents niveaux de l'analyse linguistique (par exemple entre la prosodie et la syntaxe). D'une façon générale, la représentation en graphes proposée par les grammaires de propriétés constitue donc un outil extrêmement efficace pour l'analyse linguistique.

6. Bibliographie

[Bès99] Bès G. & Blache P. (1999), « Propriétés et analyse d'un langage », in actes de *TALN'99*.

[Blache99] Blache P. (1999), "Filtering and Fusion : A Technique for Parsing with Properties", in proceedings of *NLPRS'99*.

[Blache00] Blache P. (2000), "Constraints, Linguistic Theories and Natural Language Processing", in *Natural Language Processing*, D. Christodoulakis (ed.), Lecture Notes in Artificial Intelligence, Springer.

[Blache01a] Blache P. (2001), *Les Grammaires de Propriétés : Des contraintes pour le traitement automatique des langues naturelles*, Hermès Science Publications.

[Blache01b] Blache P. (2001), « Dépendances à distance dans les Grammaires de Propriétés : l'exemple des disloquées », in actes de *TALN'01*.

[Blache02] Blache P.& Azulay D. (2002),"Parsing Ill-formed Inputs with Constraints Graphs", in A. Gelbukh (ed), Intelligent Text Processing and Computational Linguistics, LNCS, Springer.

[Blanche-Benveniste84] Blanche-Benveniste C. (1984), *Pronom et syntaxe. L'approche pronoominale et son application au français*, SELAF.

[Blanche-Benveniste97] Blanche-Benveniste C. (1997), *Approches de la langue parlée en français*, Ophrys.

[Blasco99] Blasco-Dulbecco M. (1999), *Les dislocations en français contemporain. Etude syntaxique*, Champion.

[Bouma01] Bouma G., Malouf R. & Sag I. (2001), "Satisfying Constraints on Extraction and Adjunction", in *Natural Language and Linguistic Theory*, 19 :1, Kluwer.

[Gazdar85] Gazdar G., Klein E., Pullum G. & Sag I. (1985), *Generalized Phrase Structure Grammar*, Blackwell.

[Hirst93] Hirst D. (1993), "Detaching Intonational Phrases from Syntactic Structure", in *Linguistic Inquiry*, 24 :4.

[Hirst98] Hirst D. (1998), "Intonation in British English", in Hirst D. & Di Cristo A. (eds) *Intonation Systems*, Cambridge Universtity Press.

[Kager99] Kager R. (1999), *Optimality Theory*, Cambridge University Press.

[Kahane00] Kahane S. (2000), « Extractions dans une grammaire de dépendance lexicalisée à bulles », in revue *TAL*, 41 :1, Hermès Science Publications.

[LeGoffic93] Le Goffic P. (1993), *Grammaire de la phrase française*, Hachette Supérieur.

[Maruyama90] Maruyama H. (1990), "Structural Disambiguation with Constraint Propagation", in proceedings of *ACL'90*.

[Pollard94] Pollard C. & Sag I. (1994), *Head-driven Phrase Structure Grammars*, CSLI Publications.

[Prince93] Prince A. & Smolensky P. (1993), *Optimality Theory : Constraint Interaction in Generative Grammars*, Technical Report RUCCS TR-2, Rutgers Center for Cognitive Science.

[Sag99] Sag I. & Wasow T. (1999), *Syntactic Theory. A Formal Introduction*, CSLI.