

Modélisation et analyse des coordinations elliptiques par l'exploitation dynamique des forêts de dérivation

Djamé Seddah, Benoît Sagot

► **To cite this version:**

Djamé Seddah, Benoît Sagot. Modélisation et analyse des coordinations elliptiques par l'exploitation dynamique des forêts de dérivation. Proceedings of TALN 2006 : Traitement Automatique des Langues Naturelles, 2006, Leuven, Belgium. pp.609-618, 2006. <inria-00545437>

HAL Id: inria-00545437

<https://hal.inria.fr/inria-00545437>

Submitted on 10 Dec 2010

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.

Modélisation et analyse des coordinations elliptiques par l'exploitation dynamique des forêts de dérivation

Djamé Seddah (1) et Benoît Sagot (2)

(1) National Centre for Language Technology - Dublin City University
Glasnevin, Dublin 9, Irlande

`djame.seddah@computing.dcu.ie`

(2) Projet ATOLL - INRIA

Domaine de Voluceau, B.P. 105, 78153 Le Chesnay Cedex, France

`benoit.sagot@inria.fr`

Mots-clefs : Analyse syntaxique, TAG, Coordination, Ellipses, forêt partagée, forêt de dérivation

Keywords: Parsing, TAG, Coordination, Ellipsis, shared forest, derivation forest

Résumé Nous présentons dans cet article une approche générale pour la modélisation et l'analyse syntaxique des coordinations elliptiques. Nous montrons que les lexèmes élidés peuvent être remplacés, au cours de l'analyse, par des informations qui proviennent de l'autre membre de la coordination, utilisé comme guide au niveau des dérivations. De plus, nous montrons comment cette approche peut être effectivement mise en œuvre par une légère extension des Grammaires d'Arbres Adjoints Lexicalisées (LTAG) à travers une opération dite de *fusion*. Nous décrivons les algorithmes de dérivation nécessaires pour l'analyse de constructions coordonnées pouvant comporter un nombre quelconque d'ellipses.

Abstract In this paper, we introduce a generic approach to elliptic coordination modeling and parsing. We show that the erased lexical items can be replaced, during parsing, by information gathered in the other member of the coordination, used as a guide at the derivation level. Moreover, we show how this approach can be indeed implemented as a light extension of the LTAG formalism through a so-called *fusion* operation. We provide the derivation algorithms needed to parse coordination constructions which can have an arbitrary number of elisions.

1 Introduction

Le but de cet article est de présenter un modèle du phénomène des coordinations à ellipses fondé sur la volonté de représenter toutes les dépendances argumentales au sein d'un graphe de dépendance. Notre approche repose sur le postulat qu'une coordination met toujours en jeu deux structures sémantiques parallèles, celles-ci pouvant n'être que partiellement réalisées d'un point de vue syntaxique. Bien sur, la littérature abonde de propositions allant dans ce sens et nous nous inscrivons dans la lignée de (Dalrymple *et al.*, 1991) et de (Steedman, 1990). Dans un premier temps, nous présenterons notre analyse du phénomène, puis nous en proposons une instanciation qui repose sur l'ajout au formalisme LTAG (Abeillé, 1991) d'une opération appelée *fusion* mettant en oeuvre un parallélisme des dérivations, puis nous montrons comment notre approche tranche avec celle de (Sarkar & Joshi, 1996) et (Sarkar, 1997). Enfin nous présentons notre implémentation de cette opération, qui repose sur une exploitation des dérivations partielles.

1.1 Un parallélisme des dérivations

Nous considérons que tout lexème élidé dans une coordination elliptique peut être modélisé par un lexème vide qui occupe la même position syntaxique et qui peut être mis en parallèle avec un lexème non-vide, son lexème de référence, dans l'autre membre de la coordination. Son rôle dans la structure de dérivation ne pouvant être induit par un lexème vide, il sera induit par le rôle de son lexème de référence dans la sous-structure de dérivation de l'autre membre de la coordination. Ce processus se fait par une co-indiciation partielle ou totale entre les deux unités sémantiques correspondantes.

De façon plus précise, et en ne faisant pas appel à un formalisme en particulier, nous proposons le mécanisme suivant. Les nœuds de dépendance non-lexicalisés sont reliés par duplication et/ou partage à leurs nœuds de référence dans la structure de dépendance de référence. Ces dérivations *fantômes* correspondent à des sous-structures en constituants *fantômes*, c'est-à-dire à des sous-structures qui sont (au moins partiellement) ancrées par des lexèmes vides, eux-mêmes en correspondance avec leur lexème de référence. Cette correspondance est le résultat direct du lien entre les nœuds de dépendance correspondants. Comme on peut le voir sur les exemples suivants, la référence sémantique peut être soit co-indicée soit dupliquée (nous indiquons par i' le résultat de la duplication d'une structure indicée par i), et les traits morphosyntaxiques sont répartis entre traits partagés et traits non-partagés¹ :

- (1) Paul _{i} achète des fleurs et ε_i rentre chez lui
- (2) Paul mange _{i} une pomme et Marie $\varepsilon_{i'}$ des cerises
- (3) Marie est belle _{i} ^{fem,sing} et Pierre $\varepsilon_{i'}$ ^{masc,sing} grand

Un tel mécanisme est indépendant du formalisme (pour peu que le formalisme utilisé comporte des notions telles que structures de dépendances et structures de constituants, comme c'est le cas en (L)TAG, en LFG, et dans d'autres formalismes). Cela permet de modéliser un nombre quelconque d'ellipses.

¹Cette information ne peut qu'être spécifiée explicitement pour chaque attribut. Par exemple, quand il y a ellipse d'un verbe, son mode, son temps et sa diathèse sont partagées, mais son genre, son nombre et sa personne sont dupliquées, ce qui permet un changement de la valeur (voir phrase 3). Par exemple, pour illustrer le fait que la diathèse est changée, la phrase suivante est incorrecte : **Pierre tue Marie et Jean par Paul*.

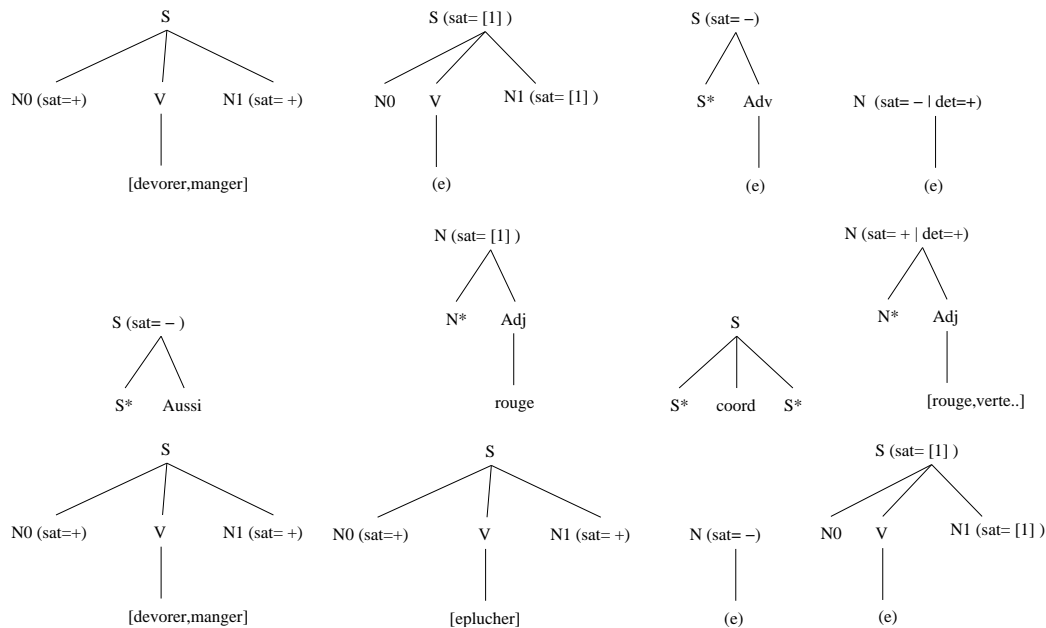


FIG. 1 – Grammaire jouet pour analyser les phrases 4 à 6 et rejeter la phrase 7.

1.2 Sur un exemple

Une implémentation cohérente de ce mécanisme doit donc être capable de construire une structure de dépendance pour le syntagme elliptique avec les liens adéquats vers la structure de référence correspondante, ainsi que de construire parallèlement une structure de constituency qui contienne des éléments vides co-indicés, là encore avec les liens adéquats vers la structure de référence correspondante. Une telle implémentation, légère extension de LTAG, fait l'objet de la section 3.

Pour illustrer le mécanisme que nous proposons, nous l'illustrons dès à présent par une grammaire-jouet (cf. figure 1) qui traite le jeu de phrases ci-dessous, dont seule la dernière phrase est incorrecte. On notera que la grammaire-jouet proposée sert d'illustration au mécanisme général que nous présentons pour le traitement de la coordination elliptique, mais que ses caractéristiques (formalisme utilisé, traits mis en œuvre) ne font pas partie de ce mécanisme général.

- (4) Pierre mange une pomme rouge et Marie deux vertes
- (5) Pierre mange une pomme rouge et Marie dévore deux vertes
- (6) Pierre mange une pomme rouge et Marie deux
- (7) *Pierre mange une pomme rouge et Marie dévore deux

En utilisant les arbres élémentaires de la figure 1, et grâce à l'algorithme de dérivation décrit plus bas, qui implémente en LTAG notre mécanisme général, nous sommes en mesure d'analyser les phrases 4 à 6 et de rejeter la phrase 7.

L'idée principale sous-jacente à cette grammaire-jouet, et qui induit les bonnes contraintes sur la forêt, est l'introduction d'un trait de saturation en plus du mécanisme général qui traite de la coordination elliptique. Un sous-arbre est saturé s'il peut se substituer dans un arbre élémentaire ancré par une unité lexicale standard (mais pas par un lexème vide). Afin de générer les phrases 4 à 6 mais pas la phrase 7, nous encodons dans les arbres élémentaires les faits suivants :

1. un déterminant peut s'adjoindre sur un arbre élémentaire de racine N (y compris s'il est ancré par un lexème vide),

2. tout arbre élémentaire ancré par un lexème standard (non vide) est saturé,
3. l'adjonction d'un adjectif sur un arbre élémentaire ancré par N sature cet arbre,
4. les nœuds de substitution d'un arbre élémentaire ancré par V accepte les arbres non-saturés si et seulement s'il est ancré par un lexème vide.

Ainsi, 4 et 5 sont correctes, puisque *deux vertes* est saturé (grâce à l'adjectif). La phrase 6 est correcte bien que *deux* ne soit pas saturé, car il se substitue dans un arbre élémentaire ancré par un lexème vide (le verbe est élidé). Mais puisque dans 7 le verbe n'est *pas* élidé, ses nœuds de substitution ne peuvent accepter d'arbre non-saturé. Cette phrase est donc rejetée.

1.3 L'opération de fusion

Dans le cadre de notre implémentation en LTAG, nous considérons que tout coordonnant, dans le cas de phrase avec ellipse, ancre un arbre initial α_{conj} de racine P avec deux noeuds de substitutions P à droite et à gauche de l'ancre (cf. fig 3). L'opération de fusion consiste à substituer les dérivations manquantes d'une partie quelconque de la coordination par celles qui y correspondent dans l'autre partie. Notre implémentation repose sur l'utilisation d'arbres non lexicalisés (correspondant à des schèmes) que nous appelons *arbres fantômes*. Leur rôle est de servir de support à des dérivations partielles, les *dérivations fantômes*, qui nous servent à reconstruire les parcours de dérivation dans la partie élidée. Nous illustrons les dérivations incomplètes de l'arbre γ dans la figure 2 par un arbre incomplet. Les dérivations fantômes sont induites par l'inclusion de l'arbre fantôme α' qui est nécessairement lié à la partie gauche (donc à l'arbre α). En superposant les deux structures, γ et α' , la fusion reconstruit les dérivations attendues et donc propose une structure de dérivation complète.

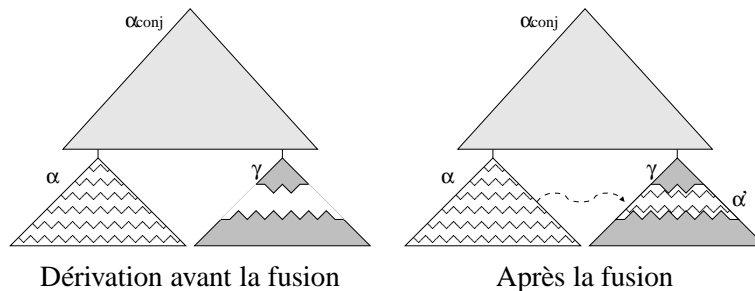


FIG. 2 – Vision abstraite de l'opération de fusion

2 Comparaison de l'approche

Si l'on souhaite pouvoir coordonner des constituants pouvant être syntaxiquement incomplets tout en obtenant une structure logique finie dominée par le coordonnant, la question primordiale est de savoir comment retrouver l'information manquante et surtout où la répartir.

2.1 Unification d'ordre supérieur et ellipses

Dans (Dalrymple *et al.*, 1991), les auteurs présentent la problématique des ellipses comme un problème relevant de la résolution d'une équation logique simple. Les auteurs considèrent en ef-

fet que la phrase *Paul aime le golf et Georges aussi* s'interprète comme *Paul et Georges aiment tout le deux le golf*. Georges partage donc une même propriété P avec Paul. Cette phrase peut donc être recevoir l'interprétation $P(\text{Paul}) \wedge P(\text{Georges})$. Si l'on considère que le syntagme *Paul aime le golf* est associé au λ -terme $\text{aime}(\text{Paul}, \text{golf})$, alors il existe une propriété d'ordre supérieure P telle que $P(\text{Paul}) = \text{aime}(\text{Paul}, \text{golf})$. Une résolution de second ordre de cette équation aboutit, parmi d'autre, à la propriété $P(X) \rightarrow \lambda x. \text{aime}(x, \text{golf})$. Ainsi en utilisant l'unification d'ordre supérieur et le λ -calcul typé comme support au mécanisme de résolution, on obtient finalement l'interprétation sémantique $\text{aime}(\text{Paul}, \text{golf}) \wedge \text{aime}(\text{Georges}, \text{golf})$. L'élégance de cette approche consiste à lier à travers des variables d'ordres supérieures les interprétations sémantiques non réalisés syntaxiquement, ce qui permet de reposer sur un mécanisme de dérivation simple pour construire l'interprétation finale.

2.2 Type raising et coordination

L'optique très élégante de (Steedman, 1990) consiste à profiter des caractéristiques des grammaires catégorielles combinatoires (CCG, (Steedman & Baldridge, 2003)) pour utiliser les catégories syntaxiques comme des types pouvant varier (et coexister) selon la fonction d'application qui les consomme. Ceci aboutit à des règles agissant comme des foncteurs qui ont la possibilité d'abstraire la fonction qui consomme un argument à partir de ce même argument. La règle de *Forward Type Raising* sur un argument quelconque a proposée est $X : a \rightarrow T/_i(T \setminus_i X) : \lambda f. fa$, où T est une méta-variable instanciant n'importe quelle catégorie de base. Ainsi si $T = S$ et si l'on applique cette règle à la catégorie NP ($X = NP$, avec $a : \text{Marie}'$) la règle d'élévation du type devient $NP : \text{Marie}' \rightarrow S/_i(S \setminus_i NP) : \lambda f. f(\text{Marie}')$. La nouvelle fonction peut ensuite être appliquée à tel ou tel argument qui aurait besoin de ce λ -terme pour donner un constituant complet, lequel pourra alors être coordonné par la conjonction, qui suit le schéma classique de coordination ($X \text{ conj } X \Rightarrow X$). L'ensemble réduit de règles nécessaires au traitement complet de la coordination en CCG constitue sans aucun doute la *carte de visite* de ce formalisme. Cependant, des questions demeurent sur le statut à donner à la coordination de catégories différentes (*Jean est républicain et fier de l'être*).

2.3 Arbres fantômes et abstractions logiques

Les approches de (Dalrymple *et al.*, 1991) et de (Steedman, 1990) offrent donc finalement un appareillage un peu similaire : dans les deux cas on cherche à abstraire l'élément présent dans un membre de la coordination pour l'instancier dans l'autre partie en utilisant le coordonnant comme pivot de l'opération. Dans notre analyse, c'est très exactement le rôle des arbres fantômes que d'être le support d'une telle abstraction (que l'on parle de variables d'ordre supérieur ou de λ -abstractions). Dans cette optique, le rôle de l'opération de *fusion* n'est que de vérifier que les dérivations induites par l'arbre fantôme se superposent à celles de la partie réalisée. En somme, le rôle de cette opération est de vérifier que les types des arguments abstraits correspondent à ceux des arguments réalisés. En cela nous différons de l'approche de (Sarkar & Joshi, 1996) car l'opération ajoutée aux TAG (l'adjonction conjointe) nécessite en réalité la présence d'un arbre auxiliaire ancré par un coordonnant qui agit comme ayant deux noeuds pieds. C'est pourquoi les auteurs étendent ensuite cette méthode dans un formalisme de grammaire synchronisé beaucoup plus orthodoxe (Sarkar, 1997). Pourtant leur approche s'affiche clairement comme une implémentation en TAG de l'approche promue par les CCG, mais à notre sens

le fait de ne pas utiliser de productions vides complique considérablement la notion même de dérivation, et produit des arbres dérivés pas entièrement satisfaisants.

3 Une implémentation en LTAG

3.1 La forêt partagée

Dans la suite, nous appelons *forêt partagée* une structure décrivant les arbres dérivés et les arbres de dérivation construits par l'analyse d'une phrase donnée selon un ensemble de règles de réécriture. À Chaque arbre correspond un ensemble de règles qui le définissent (Vijay-Shanker & Weir, 1993; Seddah & Gaiffe, 2005). Chaque règle validant une dérivation réussie produit un item de dérivation et a accès à toutes les structures de traits locales du nœud qu'elle décrit. L'objectif est d'utiliser cette forêt comme un guide servant à synchroniser les structures de dérivation de part et d'autre d'une coordination. Nous représentons la forêt via une grammaire non-contextuelle augmentée d'une pile contenant les adjonctions en cours (Seddah & Gaiffe, 2005), ce qui est proche d'une Grammaire d'indexation linéaire (LIG, (Aho, 1968)).

Chaque membre d'une règle, qui correspond à un item de type CKY (voir (Shieber *et al.*, 1995)), est de la forme $\langle N, POS, I, J, STACK \rangle$. N est un nœud d'un arbre élémentaire, POS marque la situation face à une décision d'adjonction (notée \top si une adjonction est toujours possible, \perp sinon), I et J sont les indices de début et de fin d'une chaîne dominée par N et $STACK$ est la pile contenant tous les retours d'adjonction en cours sur le nœud N (ou sur le *spine* d'un arbre auxiliaire). On note S l'axiome de la grammaire et n la longueur de la chaîne à analyser. Une règle décrit le passage d'un arbre à un autre témoin d'une dérivation.

La forme d'un item de dérivation est $\boxed{Nom : \langle N, \gamma_{from}, \gamma_{to}, Type, \gamma_{ghost} \rangle}$. Nom désigne la dérivation par l'opération $Type^2$ de l'arbre γ_{from} sur le nœud N de l'arbre γ_{to} . γ_{ghost} désigne le nom de l'arbre fantôme, s'il existe, auquel le nœud N appartient (il vaut $-$ sinon).

3.2 Vue d'ensemble

Nommons *dérivation fantôme* une dérivation mettant un jeu un arbre ancré par un élément vide et *arbre fantôme* un tel arbre. Comme le suggère la figure 4, on fait l'hypothèse que le bon arbre fantôme a été sélectionné. Dès lors, le principal problème est de savoir quelle structure utiliser pour synchroniser le processus de dérivation.

Admettons qu'un arbre α_{conj} (figure 3) soit ancré par un coordonnant et qu'un arbre initial α_1 de racine P se substitue sur le nœud P de la partie gauche de α_{conj} .

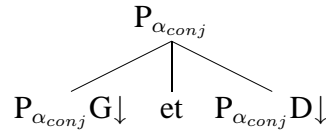
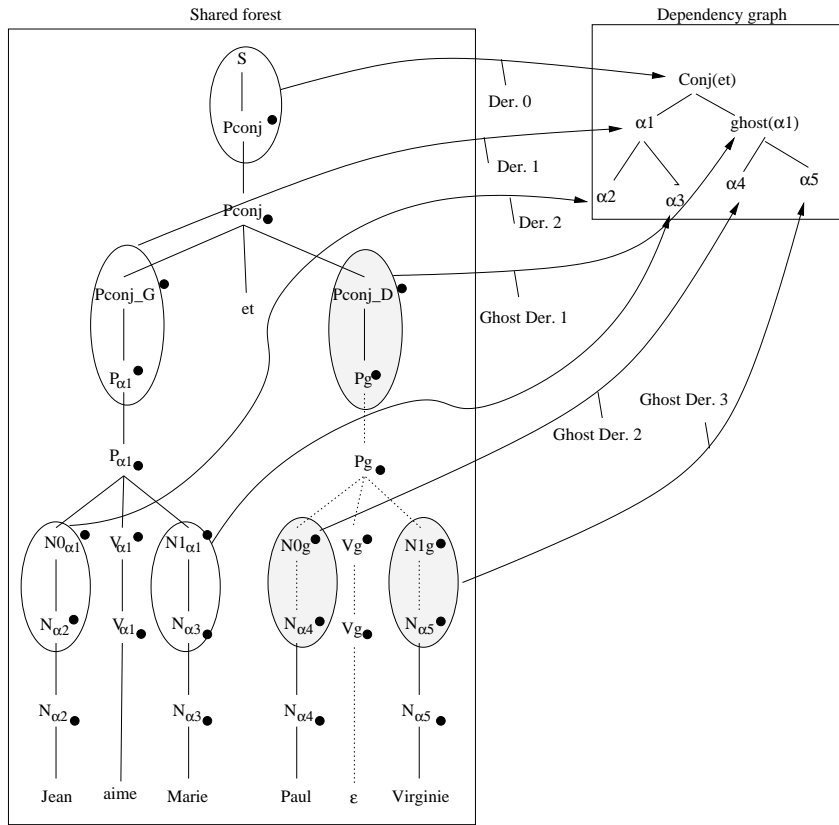
La règle correspondant au parcours de ce nœud P est alors :

$$\boxed{P_{\alpha_{conj}G}(\top, i, j, -, -) \longrightarrow P_{\alpha_1}(\top, i, j, -, -)}.$$

Dès lors, si cette règle se vérifie, on infère un item de dérivation $D1$ tel que :

$$\boxed{D1 : \langle P_{\alpha_{conj}G}, \alpha_1, \alpha_{conj}, subst, - \rangle}.$$

²Il peut s'agir d'une adjonction ($type = adj$), d'une substitution ($subst$), d'une dérivation axiomatique (ax) d'un ancrage ($anch$) — habituellement implicite dans l'arbre de dérivation —, ou une version fantôme ($ghost$) de ces opérations ($adj_g, subst_g, anch_g$).

FIG. 3 – Arbre α_{conj} FIG. 4 – Forêt partagée et graphe de dépendance pour *Jean aime Marie et Paul Virginie*

Admettons que le nœud $P_{\alpha_{conj}D}$ situé à droite du coordonnant domine un syntagme dont la tête verbale a été élidée (par exemple *et Paul Virginie*). Admettons aussi qu'il existe un arbre fantôme de racine P ayant deux positions argumentales (par exemple un quasi-arbre $N0VN1$). Appelons α_{ghost} cet arbre ancree par un élément vide. Si nous substituons cet arbre sur le nœud $P_{\alpha_{conj}D}$, nous aurons une règle, *Call-subst-ghost*, décrivant cette substitution :

$$\text{Call-subst-ghost : } P_{\alpha_{conj}D}(\top, j+1, n, -, -) \longrightarrow P_{\alpha_{ghost}}(\top, j+1, n, -, -).$$

Comme l'arbre α_{ghost} n'est pas réellement un arbre TAG au sens strict³, appelons $D1'$ la pseudo dérivation correspondant à la règle *Call-subst-ghost* :

$$D1' :< P_{\alpha_{conj}D}, \boxed{?}, \alpha_{conj}, subst, \alpha_{ghost} >$$

La variable non instanciée de cette règle dénote l'information manquante sur l'arbre synchronisé. Si notre hypothèse est correcte, cet arbre doit être co-indicé avec l'ancree d' α_1 . Nous devons donc préparer cette opération en nous synchronisant sur la dérivation de référence dans la partie gauche de l'arbre ancree par le coordonnant.

Cela nous conduit à inférer une dérivation fantôme de substitution de l'arbre α_1 sur le nœud

³Tout simplement parce qu'il n'est pas réellement lexicalisé

$P_{\alpha_{conj}D}$. La règle d'inférence correspondante, nommée $ghost(\alpha_1)$ dans la figure 4 est :

$$\frac{D1' : \langle P_{\alpha_{conj}D}, \boxed{?}, \alpha_{conj}, subst, \alpha_{ghost} \rangle}{D1 : \langle P_{\alpha_{conj}L}, \alpha_1, \alpha_{conj}, subst, - \rangle} \quad \text{Ghost} - D1 : \langle P_{\alpha_{conj}L}, \alpha_1, \alpha_{conj}, subst_g, \alpha_{ghost} \rangle$$

Ce processus pratiquement identique pour les dérivations restantes est décrit section 3.4.

3.3 Construction des dérivations fantômes

Nous avons défini une dérivation fantôme comme étant une dérivation mettant un jeu un arbre ancré par un élément vide. Cela se justifie

- par la nécessité de distinguer dans la forêt partagée les nœuds qui sont réellement parcourus lors de l'analyse et ceux sur lesquels nous voulons nous synchroniser ;
- par le besoin d'avoir accès à toutes les informations (structures de traits,...) associées à l'arbre synchronisé.

Or la seule règle permettant de savoir quel arbre va être synchronisé est celle initiant justement la substitution non-fantôme d'un côté ou de l'autre du coordonnant. La construction des dérivations fantômes (par synchronisation) nécessite donc, quel que soit l'endroit relatif dans le parcours des arbres, d'être capable en permanence de remonter la chaîne des items de dérivation. Par conséquent, la complexité du processus dépend directement du nombre de dérivations nécessaires à cette « remontée » : elle est donc exponentielle, même si la profondeur des ellipses est limitée en pratique. Nous décrivons, dans la section suivante, les règles d'inférences servant aux synchronisations.

3.4 Description des règles d'inférence

Nous ne décrivons ici que les règles d'inférence relatives aux dérivations de la partie droite de la coordination visible figure 4. Rappelons que l'arbre nommé α_{conj} est l'arbre ancré par un coordonnant. Nous ne traitons que ce qu'on peut appeler des dérivations argumentales obligatoires, parce que le statut de l'adjonction n'est pas fixé : en fait le traitement de l'adjonction serait exactement similaire à celui d'une substitution (seuls diffèrent les indices et la pile sur la règle d'inférence, ce qui n'est pas significatif en termes de dérivation). Le problème est de savoir s'il faut synchroniser ou non (voire s'il faut faire les deux, créant ainsi une ambiguïté⁴). Nous choisissons de ne pas trancher et de conserver l'analyse standard pour les adjonctions.

Substitution d'un arbre initial fantôme sur un arbre α_{conj} Si un arbre fantôme γ_g se substitue sur le nœud $P_{\alpha_{conj}D}$ d'un arbre α_{conj} et si un arbre α_1 s'est substitué sur son nœud symétrique gauche, $P_{\alpha_{conj}G}$, alors on infère un item de dérivation fantôme, nommé *Ghost-D1*, témoignant de la synchronisation, c'est-à-dire de la substitution fantôme de l'arbre α_1 sur le nœud $P_{\alpha_{conj}D}$ de l'arbre α_{conj} .

⁴Par exemple, dans la phrase *Jean achète des livres et des cahiers neufs*, devons-nous forcer la synchronisation de *neufs* via une pseudo adjonction sur l'arbre ancré par *livres* ?

$$\frac{D1' : < P_{\alpha_{conj}D}, \boxed{?}, \alpha_{conj}, subst, ghost >}{D1 : < P_{\alpha_{conj}L}, \alpha_1, \alpha_{conj}, subst, - >}}{Ghost-D1 : < P_{\alpha_{conj}D}, \alpha_1, \alpha_{conj}, subst_g, ghost >}$$

Substitution d'un arbre initial α sur un arbre fantôme γ_{ghost} Supposons qu'un arbre α_1 s'est substitué sur un nœud N_i d'un arbre fantôme γ_{ghost} (c'est-à-dire la dérivation Ghost-Der2' sur la figure 4), où i est l'indice traditionnel d'une position argumentale ($N_0, N_1...$). Supposons qu'il existe une dérivation fantôme témoignant d'une substitution d'un arbre γ_{ghost} sur un arbre α_{conj} (c'est-à-dire la dérivation Ghost-Der1, fig 4), et que cette dérivation fantôme pointe sur un arbre α_X ayant une substitution d'un arbre α_2 sur son nœud N_i (Dérivation Der2, fig. fig 4). Alors on infère une dérivation fantôme qui témoigne de la substitution de l'arbre α_1 sur l'arbre fantôme γ_{ghost} lié à α_X (c'est-à-dire Dérivation Ghost-Der2)

$$\frac{Ghost-Der2' : < N_{i_{\alpha_1}}, \alpha_1, \gamma_{ghost}, subst_g, - >}{Ghost-Der1 : < P_{\alpha_{conj}D}, \alpha_X, \alpha_{conj}, subst_g, \gamma_{ghost} >}}{Der2 : < N_{i_{\alpha_X}}, \alpha_2, \alpha_X, subst, - >}}{Ghost-Der2 : < N_{i_{\alpha}}, \alpha_1, \alpha_X, subst_g, \gamma_{ghost} >}$$

La contrainte sur les arguments évite la reconnaissance de **Jean dort et Pierre Virginie* et permet aussi de vérifier les restrictions de sélection afin d'éviter **Jean mange une pomme et Pierre l'avion*. C'est le mécanisme mis en oeuvre pour l'analyse de *Jean aime Marie et Pierre Virginie*.

Substitution d'un arbre initial fantôme α_{ghost} sur un arbre γ substitué sur un arbre α_{conj} Nous sommes ici dans une configuration totalement différente, nous avons un sous-arbre réalisé à qui il manque un argument (par exemple *Jean_i dort puis ε_i mourut*). L'objectif est de retrouver l'argument co-indicé. Contrairement à (Seddah & Gaiffe, 2005), nous ne pouvons nous reposer sur une information lexicale pour extraire cette information de la structure dérivationnelle. Nous devons d'abord laisser une trace dans la forêt, remonter toutes les dérivations aboutissant à cette dérivation fantôme, puis redescendre à gauche du coordonnant afin d'atteindre la dérivation correspondant à l'argument attendu. Contrairement aux autres cas, nous voulons créer un véritable lien via une dérivation effective, car il s'agit d'un réel partage d'argument au sens dépendance sémantique du terme.

Supposons qu'un arbre fantôme α_{ghost} s'est substitué sur le nœud N_i d'un arbre α (dérivation Ghost-Der1'), et que cet arbre α s'est substitué sur le nœud $P_{conj}D$ d'un arbre α_{conj} (dérivation Der1). Supposons que le nœud symétrique $P_{conj}G$ de l'arbre α_{conj} a reçu la substitution d'un arbre α_s , (dérivation Der2), et que cet arbre a reçu sur son nœud N_i la substitution d'un arbre α_{final} . Alors on infère un item qui témoigne d'une dérivation de substitution de l'arbre α_{final} sur le nœud N_i de l'arbre α (dérivation Ghost-Der1)⁵.

$$\frac{Ghost-Der1' : < N_{i_{\alpha_{ghost}}}, \alpha_{ghost}, \alpha, subst_g, - >}{Der1 : < P_{\alpha_{conj}D}, \alpha, \alpha_{conj}, subst, - >}}{Der2 : < P_{\alpha_{conj}L}, \alpha_s, \alpha_{conj}, subst, - >}}{Der3 : < N_{i_{\alpha_s}}, \alpha_{final}, \alpha_s, subst, - >}}{Ghost-Der1 : < N_{i_{\alpha}}, \alpha_{final}, \alpha, subst_g, \alpha_{ghost} >}$$

⁵On notera, comme indiqué plus haut, que ce mécanisme, sans restriction et dans le cas général, peut conduire à une complexité exponentielle de l'analyse en fonction de la longueur de la phrase.

4 Conclusion

Nous avons présenté un cadre général pour modéliser et analyser les constructions avec coordination elliptique. Nous avons montré que des mécanismes linguistiques simples, à savoir le partage partiel et la duplication partielle d'informations depuis une unité sémantique lexicalisée vers une unité sémantique non-lexicalisée (élide), permettent de modéliser de façon pertinente différents types de coordinations elliptiques. De plus, nous avons montré que ces mécanismes peuvent être formalisés sous la forme d'une légère extension du formalisme LTAG, et nous avons proposé des algorithmes permettant l'implémentation effective d'un analyseur qui puisse gérer les coordinations elliptiques.

Le mécanisme décrit peut être vu comme un exemple particulier d'une approche plus générale : les forêts partagées peuvent être utilisées non seulement pour accumuler l'information résultant du processus d'analyse, mais également en tant que source d'information pour ce processus. Ceci ouvre des perspectives intéressantes pour le traitement de phénomènes proches comme certains problèmes de résolutions d'anaphores ou de coréférences.

Remerciements

Nous tenons à remercier très chaleureusement nos relecteurs pour la pertinence de leurs commentaires. Ce travail a grandement bénéficié des remarques faites par Éric Villemonte de La Clergerie, Alexandra Mounier et Bertrand Gaiffe. Les propos tenus n'engageant bien sûr que les auteurs.

Références

- ABEILLÉ A. (1991). *Une grammaire lexicalisée d'arbres adjoints pour le français*. PhD thesis, Paris 7.
- AHO A. V. (1968). Indexed grammars—an extension of context-free grammars. *J. ACM*, **15**(4), 647–671.
- DALRYMPLE M., SHIEBER S. M. & PEREIRA F. C. N. (1991). Ellipsis and higher-order unification. *Linguistics and Philosophy*, **14**(4), 399–452.
- SARKAR A. (1997). Separating dependency from constituency in a tree rewriting system. In *MOL'97, Saarbrücken*.
- SARKAR A. & JOSHI A. (1996). Coordination in tree adjoining grammars : Formalization and implementation. In *COLING'96, Copenhagen*, p. 610–615.
- SEDDAH D. & GAIFFE B. (2005). Des arbres de dérivation aux forêts de dépendance : un chemin via les forêts partagées. In *Traitement automatique des langues Naturelles - TALN'05, Dourdan, France*.
- SHIEBER S., SCHABES Y. & PEREIRA F. (1995). Principles and implementation of deductive parsing. *Journal of Logic Programming*, **24**, 3–36.
- STEEDMAN M. (1990). Gapping as constituent coordination. *Linguistic and Philosophy*, **13**, 207–264.
- STEEDMAN M. & BALDRIDGE J. (2003). Combinatory categorial grammar.
- VIJAY-SHANKER K. & WEIR D. (1993). The use of shared forests in tree adjoining grammar parsing. In *EACL '93*, p. 384–393.