



# Un schéma d'annotation en dépendances syntaxiques profondes pour le français

Guy Perrier, Marie Candito, Bruno Guillaume, Corentin Ribeyre, Karën Fort, Djamé Seddah

## ► To cite this version:

Guy Perrier, Marie Candito, Bruno Guillaume, Corentin Ribeyre, Karën Fort, et al.. Un schéma d'annotation en dépendances syntaxiques profondes pour le français. TALN - Traitement Automatique des Langues Naturelles, Jul 2014, Marseille, France. pp.574-579, 2014. <hal-01054407>

**HAL Id: hal-01054407**

**<https://hal.inria.fr/hal-01054407>**

Submitted on 6 Aug 2014

**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.

## Un schéma d'annotation en dépendances syntaxiques profondes pour le français \*

Guy Perrier<sup>1</sup> Marie Candito<sup>2</sup> Bruno Guillaume<sup>3</sup>

Corentin Ribeyre<sup>2</sup> Karën Fort<sup>1</sup> Djamé Seddah<sup>4</sup>

(1) Université de Lorraine/LORIA (2) Université Paris Diderot/INRIA

(3) Inria Nancy Grand-Est/LORIA (4) Université Paris Sorbonne/INRIA

**Résumé.** À partir du schéma d'annotation en dépendances syntaxiques de surface du corpus Sequoia, nous proposons un schéma en dépendances syntaxiques profondes qui en est une abstraction exprimant les relations grammaticales entre mots sémantiquement pleins. Quand ces relations grammaticales sont partie prenante de diathèses verbales, ces diathèses sont vues comme le résultat de redistributions à partir d'une diathèse canonique et c'est cette dernière qui est retenue dans notre schéma d'annotation syntaxique profonde.

**Abstract.** We describe in this article an annotation scheme for deep dependency syntax, built from the surface annotation scheme of the Sequoia corpus, abstracting away from it and expressing the grammatical relations between content words. When these grammatical relations take part into verbal diatheses, we consider the diatheses as resulting from redistributions from the canonical diathesis, which we retain in our annotation scheme.

**Mots-clés :** schéma d'annotation, syntaxe profonde, grammaires de dépendance.

**Keywords:** annotation scheme, deep syntax, dependency grammar.

### 1 Introduction

Les corpus annotés en dépendances syntaxiques présentent un intérêt croissant par rapport aux corpus annotés en syntagmes dans la mesure où ils permettent plus directement d'extraire les relations prédicat-argument constitutives d'une représentation sémantique. Cette extraction reste cependant non triviale, la syntaxe offrant une grande variabilité dans la façon d'exprimer ces relations. Une solution consiste à définir un niveau intermédiaire, la syntaxe profonde, qui est une abstraction de la syntaxe de surface et qui vise à se rapprocher du niveau sémantique.

La Théorie Sens-Texte (TST)(Mel'čuk, 1988), parmi les différents niveaux de la langue qu'elle définit, propose un niveau de syntaxe profonde, où seuls les mots sémantiquement pleins sont présents. Ce niveau de syntaxe profonde est lié à la sémantique dans la mesure où les mots présents renvoient aux entrées d'un lexique sémantique.

Comme le pointe Zabokrtsky (2005), la TST est très proche du modèle de la Description Générative Fonctionnelle ou Functional Generative Description (FGD) (Sgall *et al.*, 1986), sur lequel s'appuie l'annotation du Prague Dependency Treebank (Hajic *et al.*, 2006). La FGD définit un niveau tectogrammatical qui s'apparente au niveau de syntaxe profonde de la TST avec une différence dans son rapport à la sémantique : l'absence de lien avec un lexique sémantique est compensée par un étiquetage des dépendances prédicat-argument avec des rôles thématiques.

Contrairement à ces deux modèles, le schéma de syntaxe profonde que nous proposons ne marque aucun engagement sémantique, la sémantique ne servant qu'à lever les ambiguïtés de rattachement. Nous ne retenons que les mots sémantiquement pleins mais les relations entre ceux-ci expriment des fonctions grammaticales. Les diathèses verbales présentes en surface sont considérées comme le résultat de redistributions à partir de diathèses canoniques et ce sont ces diathèses canoniques qui sont représentées au niveau profond.

Nous sommes partis de l'annotation de surface du corpus Sequoia (Candito & Seddah, 2012b) et nous avons réalisé son annotation en syntaxe profonde, en nous inspirant des méthodes de Bonfante *et al.* (2011). Le corpus Sequoia annoté

---

\*. Ce travail a été partiellement financé par l'Agence Nationale de la Recherche au titre du programme Investissements d'Avenir, au sein du Labex EFL (ANR-10-LABX-0083). Nous avons bénéficié de discussions avec Alain Polguère et Sylvain Kahane, que nous remercions.

en syntaxe profonde est librement disponible<sup>1</sup>. Si nous décrivons, dans (Candito *et al.*, 2014), l'ensemble du travail d'annotation du corpus, nous nous concentrons ici sur la présentation du schéma d'annotation. Une description plus précise des phénomènes est fournie dans le guide d'annotation sur le site du corpus.

## 2 Schéma d'annotation en syntaxe profonde

Nous avons défini notre schéma d'annotation en syntaxe profonde en partant du schéma d'annotation en dépendances de surface du corpus Sequoia, ce qui a eu une influence certaine sur le résultat. Il est donc important de détailler comment le schéma de surface a été lui-même défini : le corpus Sequoia a été annoté d'abord en arbres de constituants, en suivant le schéma d'annotation du corpus arboré de Paris 7 (Abeillé & Barrier, 2004) ou French Treebank (FTB). Le corpus en constituants a ensuite été converti automatiquement en dépendances de surface, en suivant la procédure décrite dans (Candito *et al.*, 2010). Les arbres de dépendances résultants suivent ainsi très largement les choix linguistiques du FTB (Abeillé *et al.*, 2004; Abeillé, 2004), dans la mesure où la majorité des phénomènes syntaxiques sont mécaniquement traduits en dépendances<sup>2</sup>. Enfin, les dépendances longue distance ont été corrigées manuellement (Candito & Seddah, 2012a) dans les arbres de dépendance obtenus par conversion automatique, ce qui a introduit quelques arcs non projectifs<sup>3</sup>.

Le schéma d'annotation de surface résultant est notre point de départ. Nous avons cherché, pour des raisons pragmatiques, à minimiser les divergences entre les niveaux surfacique et profond, pour nous concentrer sur les phénomènes non directement représentables dans les arbres de surface. Aussi, par exemple, avons-nous conservé la représentation des coordinations avec le premier conjoint comme tête, bien que cela ne permette pas de distinguer les dépendants partagés par plusieurs conjoints des dépendants du seul premier conjoint.

### 2.1 Choix théoriques

L'objectif principal de nos représentations syntaxiques profondes (REPRSYNTPROF dans la suite) est de généraliser sur la variation syntaxique autant que possible sans faire de distinctions ni de généralisations purement sémantiques. Nous utilisons pour cela la notion de *sous-catégorisation canonique* et représentons les changements de diathèse comme des redistributions des fonctions grammaticales sous-catégorisées par un lexème. Nous inspirant de la Grammaire Relationnelle (Perlmutter, 1983), nous distinguons fonction grammaticale *canonique* et fonction grammaticale *finale* d'une part, et *cadre de sous-catégorisation canonique* (CS canonique) et *cadre de sous-catégorisation finale* (CS final)<sup>4</sup>.

Définissons d'abord la notion de sous-catégorisation finale pour un verbe : elle contient d'une part les fonctions finales associées aux arguments *exprimés* du verbe, et d'autre part, dans le cas d'ellipse et/ou de verbes non conjugués, les fonctions des éléments qui seraient des arguments du verbe si celui-ci était utilisé sans ellipse et conjugué. Cette formulation permet de faire entrer dans le CS final par exemple le sujet des infinitifs, le sujet ou l'objet des participes épithètes, le sujet de verbes coordonnés ou plus généralement tout argument partagé par plusieurs prédicats. Par exemple, dans « *Anna veut dormir, mais devra peut-être veiller* », le CS final de *dormir* est [sujet] rempli par *Anna*, et le CS final de *devra* est [sujet, objet], rempli par *Anna* et *veiller*. Entrent également dans le CS final l'élément modifié par un participe épithète : par exemple pour « *les personnes nées en 40* », le nom *personnes* est la tête du sujet final de *nées*.

Passons maintenant à la définition précise de la sous-catégorisation canonique. Afin de neutraliser la variation syntaxique due aux changements de diathèse, nous considérons ceux-ci comme des redistributions des fonction canoniques associées aux arguments syntaxiques. Suivant la Grammaire Relationnelle (Perlmutter, 1983), le CS final est vu comme résultant de l'application de 0 à  $n$  redistributions sur un CS canonique. Etant donnée une occurrence de verbe, le CS canonique peut donc par définition être obtenu par application inverse des redistributions appropriées. Un exemple simple est le cas d'un verbe au passif dont le CS final est [sujet, par-objet] et le CS canonique est [sujet, objet]. Les éléments du CS canonique, appelés *arguments syntaxiques profonds*, sont obligatoirement sémantiquement pleins. Le *il* explétif n'appartient qu'au CS final. Donc dans l'exemple « *Trois personnes arrivent* », à partir du CS canonique [sujet], la redistribution de l'impersonnel rétrograde le sujet en objet direct, et un *il* explétif remplit la fonction sujet final : « *Il arrive trois personnes* » a pour CS canonique [sujet] et pour CS final [sujet, objet].

Pour définir nos REPRSYNTPROF, nous n'avons considéré que les redistributions qui comportent un marquage morpho-syntaxique (typiquement l'auxiliaire pour le passif, ou le clitique sémantiquement vide *se* pour les alternances moyennes

1. <http://deep-sequoia.inria.fr>

2. Des informations supplémentaires sont prédites par la procédure de conversion, en cas de sous-spécification dans la version en constituants (c'est le cas pour les étiquettes de dépendances pour les dépendants de gouverneurs non verbaux).

3. Par exemple, à la conversion de l'arbre syntagmatique de « ... *le succès que la municipalité était en droit d'attendre* », le pronom relatif est mécaniquement rattaché comme dépendant de *était*, et manuellement corrigé pour dépendre de *attendre*.

4. La Grammaire Relationnelle utilise l'adjectif *initial* plutôt que *canonique*.

et neutres). Les alternances syntaxiques sans marquage morpho-syntaxique ne sont pas capturées au sein de nos REPR-SYNTPROF, et donnent lieu à des CS canoniques différents. Les repérer, en l'absence de marquage formel, relève pour nous de l'analyse sémantique. Par exemple, pour le verbe *baisser*, les deux emplois *X baisse Y* et *Y baisse*, typiquement reliés par l'alternance causative/inchoative non marquée, ne sont pas reliés dans nos REPRSYNTPROF, et ont deux CS canoniques distinctes [sujet, objet] et [sujet] respectivement, qui sont aussi ici les CS finaux en l'absence de redistribution. En revanche, pour l'alternance moyenne (par exemple « *On avale facilement ce médicament / Ce médicament s'avale facilement* ») ou l'alternance neutre (qui "efface" l'actant agentif ou causal, comme dans « *Cela dissout le médicament / Le médicament se dissout (de lui-même)* »), le lien entre les deux versions est capturé par redistribution, et pour ces deux alternances, l'objet direct dans la version transitive (*médicament*) est sujet final mais objet canonique dans la version intransitive. Nous retenons comme redistributions : le passif, l'impersonnel, le moyen, le neutre et le causatif (voir les exemples en section 3)<sup>5</sup>, certaines pouvant interagir. Nous renvoyons à (Candito, 1999) pour une étude des interactions entre redistributions pour le français.

Nous avons volontairement fait deux distinctions différentes : représentation profonde *versus* de surface, et fonction grammaticale finale *versus* canonique. Cela nous est utile pour capturer par exemple la régularité concernant le contrôle des sujets des infinitifs : il s'agit du sujet *final* de l'infinitif, quelle que soit la diathèse de celui-ci. Ainsi, pour « *Paul veut être embauché* », dans la représentation de surface, *Paul* est le sujet final de *veut*. Dans la représentation profonde, il est aussi le sujet *final* de *être embauché* et son objet canonique.

Outre les verbes, nous traitons dans nos annotations les adjectifs, auxquels la notion de sous-catégorisation peut être étendue, mais sans distinction entre final et canonique (ce que nous ne détaillons pas ici par manque de place). Le travail reste à faire pour les autres catégories de prédicats, en particulier les noms.

Nous pouvons maintenant lister précisément les informations qui sont explicitées dans nos REPRSYNTPROF par rapport aux représentations de surface :

- le statut sémantique des mots de la phrase (sémantiquement vide ou plein),
- la liste complète des arguments syntaxiques profonds des verbes et des adjectifs,
- et leur fonction grammaticale canonique (et la ou les redistributions amenant à l'emploi observé en surface).

Dans nos REPRSYNTPROF, chaque argument syntaxique profond d'un prédicat lui est directement rattaché. Plus précisément, nous prenons comme tête d'un argument syntaxique profond l'élément sémantiquement plein le plus haut (les complémenteurs vides et les prépositions régies sont court-circuitées, comme le sont aussi les pronoms relatifs dont la référence peut être résolue syntaxiquement, comme illustré dans la phrase (2) de la figure 1).

## 2.2 Caractéristiques formelles

Nous définissons une représentation *complète* comme un graphe de dépendances contenant à la fois la REPRSYNTSURF et la REPRSYNTPROF. La figure 1 fournit des exemples, tirés du corpus Sequoia, de représentations complètes. Les nœuds sont les mots de la phrase (ou des composants de composés réguliers), et sont typés comme sémantiquement vides (en rouge dans la figure) ou pleins (en noir). Les arcs portent :

- une information sur leur appartenance à la REPRSYNTSURF ou pas, et leur appartenance à la REPRSYNTPROF ou pas : un arc peut être surfacique mais non profond (arcs rouges), profond mais non surfacique (arcs bleus), et à la fois profond et surfacique (arcs noirs) ;
- une étiquette qui est constituée soit d'une seule fonction, à la fois finale et canonique, pour les fonctions n'intervenant jamais dans des changements de diathèse (comme Mod), soit de la fonction finale et de la fonction canonique (dans toute la suite, une étiquette notée "*f : c*" correspond à la fonction finale *f* et la fonction canonique *c*).

La REPRSYNTPROF pour une phrase donnée est formée des nœuds sémantiquement pleins et des arcs profonds, c'est-à-dire à la fois les arcs qui sont profonds et surfaciques (noirs) et les arcs profonds non surfaciques (bleus). Il s'agit formellement d'un graphe orienté, qui peut contenir des cycles et des arcs multiples (un même couple gouverneur / dépendant peut être relié par plusieurs arcs de même orientation mais d'étiquettes différentes).

## 3 Illustration sur des exemples de phénomènes relatifs au verbe

Dans cette section, nous montrons comment le schéma qui vient d'être défini s'applique à certains phénomènes linguistiques intéressants. Compte tenu des contraintes de l'article, nous nous limitons à des phénomènes relatifs au verbe. Les exemples présentés dans la figure 1 illustrent ces phénomènes et les interactions complexes qui existent entre l'ajout de relations prédicat-argument liées à la syntaxe profonde, la redistribution entre les fonctions grammaticales canoniques et

5. Le réfléchi, bien que présentant des propriétés « *intransitives* » qui peuvent justifier une représentation *via* redistributions, n'est pas traité dans nos REPRSYNTPROF *via* redistribution (cf. l'exemple (1), dans la figure 1).

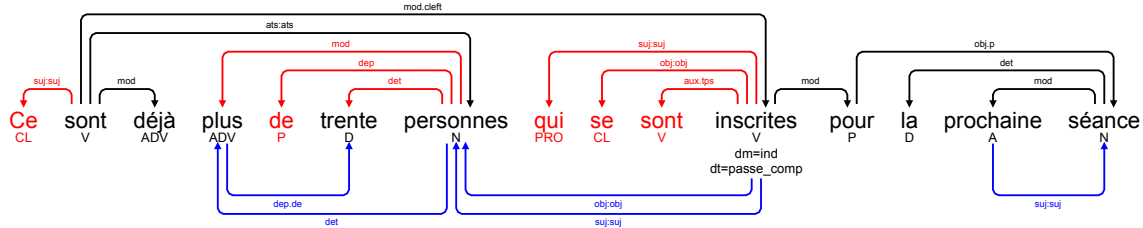
finale, et la suppression de mots sémantiquement vides.<sup>6</sup>

**Transformation de marqueurs grammaticaux en traits** Sont concernés les auxiliaires et les clitiques ne représentant aucun argument mais destinés à modifier la sémantique du verbe auquel ils s’appliquent.

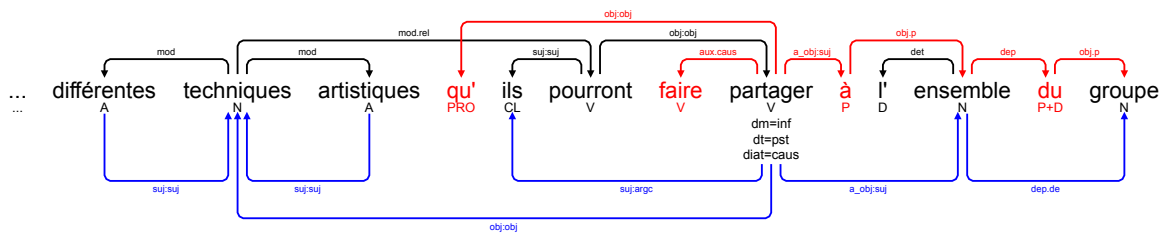
En REPRSYNTSURF, les auxiliaires (pour les temps composés, les passifs et les causatifs) sont systématiquement rattachés au dernier verbe du noyau verbal (le verbe sémantiquement plein) et ils sont supprimés en REPRSYNT-PROF. L’information de mode, de temps et éventuellement de diathèse dont ils sont porteurs est conservée sous forme de traits profonds attachés au verbe plein.

Les clitiques réfléchis intrinsèques *se* sont également traités comme des marqueurs et systématiquement transformés en traits (cf. infra la section spécifique sur les réfléchis).

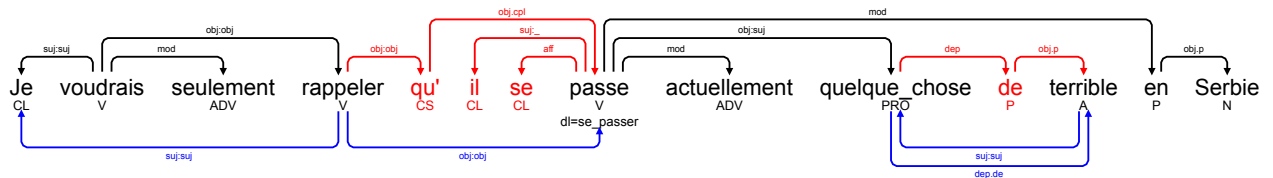
(1) auxiliaire, réfléchi, clivée



(2) causatif, verbe à montée, relative épithète



(3) réfléchi, impersonnel



(4) réfléchi, coordination

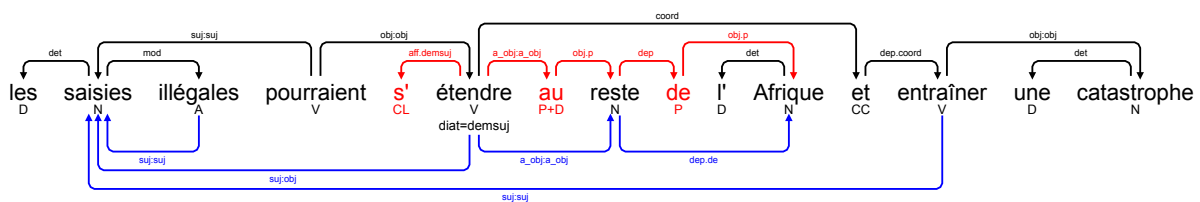


FIGURE 1 – Exemples d’annotation de certains phénomènes. La REPRSYNTPROF est formée des mots en noir, des arcs noirs et des arcs bleus. La REPRSYNTSURF contient tous les mots, et les arcs rouges et noirs.

**Les causatifs** La phrase (2) illustre une construction causative dont le noyau est « faire partager » et le passage de la REPRSYNTSURF à la REPRSYNTPROF montre l’interférence entre plusieurs phénomènes.

Le verbe *pourront* est ici un verbe à montée du sujet, si bien qu’*ils* est sujet profond final de *partager*. Le changement de la diathèse finale causative en diathèse canonique fait qu’*ils* passe de la fonction finale SUJ à la fonction canonique

6. Les schémas sont simplifiés et ne font apparaître que les traits pertinents concernant les phénomènes décrits dans l’article. Dans la version finale, une adresse URL renverra à un tableau complet des traits utilisés.

ARGC, pour « *argument causateur* ». Parallèlement, à passe de AOBJ à SUJ, mais comme en REPRSYNTPROF, la préposition régie est effacée, ces fonctions sont transférées sur *ensemble*.

Enfin, l'auxiliaire causatif *faire* est supprimé et une indication de diathèse causative (*diat=caus*) est ajoutée au verbe *partager*<sup>7</sup>.

**Les clitiques réfléchis** Le clitique réfléchi *se* a été annoté de trois manières différentes. Dans la phrase (1), *se* est un vrai réfléchi, objet direct du verbe *inscrites* en REPRSYNTSURF et il disparaît en REPRSYNTPROF ; sa fonction OBJ est alors reportée sur son antécédent *personnes* (via le pronom relatif *qui* qui disparaît lui aussi en REPRSYNTPROF). Dans la phrase (3), *se* est un réfléchi intrinsèque. Affixe du verbe essentiellement pronominal *se passer*, il disparaît en REPRSYNTPROF, où il est remplacé par un trait de lemme profond *dl=se\_passer* attaché à *passer*. Enfin, dans la phrase (4), *s'* avec le verbe *étendre* s'inscrit dans une diathèse neutre. Cette diathèse provient d'une redistribution dans laquelle l'objet canonique devient le sujet final et le sujet canonique n'a aucun référent clairement identifié. Cela la distingue de la diathèse moyenne où ce référent n'est pas exprimé mais existe. Comme il est difficile de distinguer les deux constructions, nous les marquons toutes deux en REPRSYNTPROF par un trait *diat=demsuj* attaché au verbe, *étendre* dans notre exemple.

**Les sujets impersonnels** Le clitique *il* explétif peut apparaître soit comme sujet final de verbes essentiellement impersonnels, soit dans le cas de diathèses impersonnelles de verbes qui peuvent également apparaître avec des sujets référentiels. La phrase (3), avec « *il se passe* », illustre le second cas. Le clitique *il* disparaît en REPRSYNTPROF et l'objet final *quelque chose* y est sujet canonique.

**Les coordinations** Selon le schéma d'annotation du FTB, la tête de la structure de coordination est la tête du premier conjoint. La conjonction de coordination est reliée à la tête avec une étiquette COORD et la tête du second conjoint à la conjonction avec une étiquette DEP.COORD. Ces dépendances sont présentes aussi bien en surface qu'en profondeur. Pour des raisons sémantiques, nous traitons au niveau profond les dépendances entrantes différemment des dépendances sortantes. D'un point de vue entrant, une coordination est vue comme un tout et les dépendances ne sont pas distribuées entre les conjoints (autrement dit, nous ne désambiguïsons pas entre lecture collective et distributive de la coordination). En revanche, les dépendants d'un élément coordonné qui sont reliés à son premier conjoint en surface sont distribués aux autres conjoints en profondeur (sauf exception non détaillée ici). Ainsi, dans la phrase (4), le sujet profond *saisies* des conjoints *étendre* et *entraîner* est distribué, alors que le gouverneur de la coordination *pourraient* n'est pas distribué.

**Les clivées** La phrase (1) est une clivée : le syntagme nominal « *plus de trente personnes* » est extrait de la phrase canonique « *plus de trente personnes se sont inscrites pour la prochaine séance* » pour être placé en attribut du sujet de *sont*, afin que l'accent soit mis sur lui. La trace du syntagme extrait est représentée par le pronom relatif *qui* et la phrase canonique devient une proposition relative rattachée à *sont* par une dépendance MOD.CLEFT.

En REPRSYNTPROF, la phrase canonique n'est pas complètement restaurée et la dépendance MOD.CLEFT est conservée afin de permettre au verbe *sont* de recevoir des modificateurs, comme l'adverbe *déjà* dans l'exemple. Seuls *ce* et *qui* sont supprimés et la fonction *sujet* du pronom relatif est transférée sur son antécédent *personnes*.

## 4 Représentation syntaxique profonde versus représentation sémantique

Si nos REPRSYNTPROF partagent certaines caractéristiques avec des représentations sémantiques (qui sont parfois appelées *profondes*), elles s'en différencient sur des points importants. Le premier point est que dans nos REPRSYNTPROF, le sens des nœuds lexicaux n'est pas désambiguïté. La sémantique n'est utilisée que pour désambiguïser des rattachements syntaxiques ou des changements de diathèse. Ensuite, même si les mots sémantiquement vides sont écartés de nos REPRSYNTPROF, les nœuds restants ne forment pas nécessairement une unité sémantique. En effet, les unités polylexicales (expressions figées, constructions à verbe support, composés syntaxiquement réguliers) ne sont pas marquées comme formant une unité, et sont représentées par une structure syntaxique régulière.

En outre, la sous-catégorisation canonique des prédicats explicitée dans les REPRSYNTPROF peut contenir des éléments qui ne sont pas des arguments sémantiques du prédicat (comme par exemple le sujet *ils* du verbe à montée *pourront* de la phrase (2), dans la figure 1). Enfin, alors que les représentations sémantiques typent en général les actants sémantiques des prédicats au moyen de rôles sémantiques ou de simples numéros, nos REPRSYNTPROF utilisent seulement les fonctions canoniques, ce qui est cohérent avec le fait que les lemmes ne sont pas désambiguïtés. Les fonctions demeurent alors des indices importants pour la (future) désambiguïté des prédicats.

Pour résumer, pour obtenir des structures prédicat-arguments (sémantiques) à partir de nos REPRSYNTPROF, il serait nécessaire de repérer les unités polylexicales, désambiguïser les prédicats, ne retenir que les arguments sémantiques de

7. Par manque de place, nous ne détaillons pas ici les différents types de causatifs.

ceux-ci parmi les arguments syntaxiques, et les associer à un rôle sémantique ou une simple numérotation. Des numéros d'argument peuvent être obtenus en utilisant un ordre d'oblicité des fonctions grammaticales canoniques (typiquement sujet < objet < obliques). A noter cependant que les sujets canoniques de nos REPRSYNTPROF peuvent correspondre aussi bien à des proto-agents qu'à des proto-patients (au sens de (Dowty, 1991)), cette distinction étant considérée comme sémantique. Nos REPRSYNTPROF ne capturent donc pas que dans une alternance causative/inchoative, par exemple *X baisse Y* versus *Y baisse*, l'argument *Y* joue le même rôle sémantique. Cette généralisation est capturée notamment dans la ressource PropBank (Palmer *et al.*, 2005) en utilisant un même rôle arg1 pour l'argument *Y* dans les deux variantes. D'autres travaux prévoient plutôt un lien au niveau du lexique, entre le lexème transitif et le lexème intransitif : c'est ce qui est prévu dans le dictionnaire explicatif et combinatoire de la TST (Mel'čuk *et al.*, 1999), et dans le DeepBank (Flickinger *et al.*, 2012), constitué de représentations syntactico-sémantiques dérivées d'analyses HPSG de l'anglais (Oepen, p.c). Quoi qu'il en soit, nous espérons que la libre mise à disposition du corpus et de son guide d'annotation permettront tant le développement de ressources syntaxico-sémantiques plus riches que des perspectives nouvelles d'application.

## References

- ABEILLÉ A. (2004). Guide des annotateurs, annotation fonctionnelle. *LLF Annotation Guide*.
- ABEILLÉ A. & BARRIER N. (2004). Enriching a French treebank. In *Proc. of LREC*, Lisbonne, Portugal.
- ABEILLÉ A., TOUSSENEL F. & MARTINE C. (2004). Corpus le monde, annotations en constituants, guide pour les correcteurs. *LLF Annotation Guide*.
- BONFANTE G., GUILLAUME B., MOREY M. & PERRIER G. (2011). Enrichissement de structures en dépendances par réécriture de graphes. In *Proc. of TALN*, Montpellier, France.
- CANDITO M. (1999). *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées application au français et à l'italien*. PhD thesis, Université Paris Diderot.
- CANDITO M., CRABBÉ B. & DENIS P. (2010). Statistical french dependency parsing: Treebank conversion and first results. In *Proc. of LREC*, La Valette, Malte.
- CANDITO M., PERRIER G., GUILLAUME B., RIBEYRE C., FORT K., SEDDAH D. & DE LA CLERGERIE É. (2014). Deep Syntax Annotation of the Sequoia French Treebank. In *International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Islande.
- CANDITO M. & SEDDAH D. (2012a). Effectively long-distance dependencies in French: annotation and parsing evaluation. In *Proc. of TLT 11*, Lisbonne, Portugal.
- CANDITO M. & SEDDAH D. (2012b). Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In *Proc. of TALN*, Grenoble, France.
- DOWTY D. (1991). Thematic proto-roles and argument selection. *Language*, **67**(3), 547–619.
- FLICKINGER D., ZHANG Y. & KORDONI V. (2012). Deepbank: A dynamically annotated treebank of the wall street journal. In *Proc. of the Eleventh International Workshop on Treebanks and Linguistic Theories*, p. 85–96.
- HAJIC J., PANEVOVÁ J., HAJICOVÁ E., SGALL P., PAJAS P., ŠTEPÁNEK J., HAVELKA J., MIKULOVÁ M., ZABOKRTSKÝ Z. & RAZIMOVÁ M. Š. (2006). Prague dependency treebank 2.0. *CD-ROM, Linguistic Data Consortium, LDC Catalog No.: LDC2006T01, Philadelphie*, **98**.
- MEL'ČUK I. (1988). *Dependency syntax: theory and practice*. State University Press of New York.
- MEL'ČUK I., ARBATCHEWSKY-JUMARIE N., IORDANSKAJA L., MANTHA S. & POLGUÈRE A. (1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques IV* [Explanatory and Combinatorial Dictionary of Contemporary French. Lexico-Semantic Research IV]. Montréal: Les Presses de l'Université de Montréal.
- PALMER M., GILDEA D. & KINGSBURY P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, **31**(1), 71–106.
- PERLMUTTER D. (1983). *Studies in Relational Grammar 1*. University of Chicago Press.
- SGALL P., HAJIČOVÁ E. & PANEVOVÁ J. (1986). *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht:Reidel Publishing Company and Prague:Academia.
- ZABOKRTSKY Z. (2005). Resemblances between Meaning-Text Theory and Functional Generative Description. In *Proc. of MTT 2005*, p. 549–557, Moscou, Russie.