



HAL
open science

Trois schémas d'annotation syntaxique en dépendance pour un même corpus de français oral : le cas de la macrosyntaxe

Kim Gerdes, Sylvain Kahane

► To cite this version:

Kim Gerdes, Sylvain Kahane. Trois schémas d'annotation syntaxique en dépendance pour un même corpus de français oral : le cas de la macrosyntaxe. Atelier sur les corpus annotés du français (ACor4French), Jun 2017, Orléans, France. halshs-01740454

HAL Id: halshs-01740454

<https://shs.hal.science/halshs-01740454v1>

Submitted on 19 Sep 2019

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.

Trois schémas d'annotation syntaxique en dépendance pour un même corpus de français oral : le cas de la macrosyntaxe

Kim Gerdes¹, Sylvain Kahane²

(1) LPP, Université Paris 3 Sorbonne Nouvelle & CNRS

(2) Modyco, Université Paris Nanterre & CNRS

kim@gerdes.fr, sylvain@kahane.fr

RESUME

Nous présentons trois schémas d'annotation appliqués à un même corpus de français oral : Rhapsodie, Orféo et UD (Universal Dependencies). Les deux premiers sont diffusés et le troisième est en cours. Nous mettons ici l'accent sur la macrosyntaxe, c'est-à-dire sur les relations au sein d'un énoncé qui ne relèvent pas de la rection. Nous contrastons les trois schémas et proposons un quatrième schéma qui subsume les trois autres.

ABSTRACT

Three syntactic dependency annotation schemes for the same French spoken corpus: The case of macrosyntax.

We present three dependency annotation schemes applied to the same corpus of spoken French: Rhapsodie, Orféo and UD (Universal Dependencies). The first two are distributed, the latter is work in progress. We emphasize the annotation of “macrosyntactic” phenomena, i.e. relations in the utterance that are not part of the sub-categorization. We contrast the three schemes and propose a fourth scheme that subsumes the three others.

MOTS-CLES : treebank en dépendance, macrosyntaxe, prénoyau, parenthétique, incise

KEYWORDS: dependency treebank, Universal Dependencies, dislocation, parataxis, insert, parenthesis

1 Introduction

L'annotation en dépendance est devenue hégémonique pendant la décennie passée et peu de ressources syntaxiques utilisent encore des structures de constituants. Les raisons de ce chamboulement sont multiples, allant de la légèreté de l'annotation au succès des parseurs en dépendance en passant par la facilité d'encoder les relations entre unités non-adjacentes, notamment pour les langues dites « à ordre libre ». Historiquement, la proximité avec la sémantique mène aussi à moins de variantes d'analyse, car la dépendance permet d'encoder la valence de chaque prédicat. En même temps, des phénomènes très fréquents à l'oral (et loin d'être absents à l'écrit), phénomènes souvent regroupés sous l'étiquette de « macrosyntaxe », n'ont pas encore une annotation générale et consensuelle comme on va le montrer dans la suite.

Cet article présente et compare trois schémas d'annotation syntaxique en dépendance appliqués à un même corpus de l'oral : le schéma Rhapsodie développé dans le cadre de l'ANR Rhapsodie (2008-2012), le schéma Orféo développé dans le cadre de l'ANR Orféo (2013-2017) et le schéma UD issu

du projet Universal Dependencies regroupant des treebanks dans une cinquantaine de langues et appliqué par trois équipes différentes à des corpus de français écrit (McDonald et al. 2013, Candito & Seddah 2012, Sanguinetti & Bosco sous presse), mais encore très peu utilisé pour des corpus oraux. Le corpus Rhapsodie, 33 000 mots, est également annoté en prosodie et est distribué librement sur www.projet-rhapsodie.fr (Lacheret et al. 2014). La partie annotée à la main du corpus Orféo, qui inclut Rhapsodie, comporte environ 150 000 mots et sera distribuée librement prochainement sur www.projet-orfeo.fr. La conversion de Rhapsodie au schéma UD est en cours.

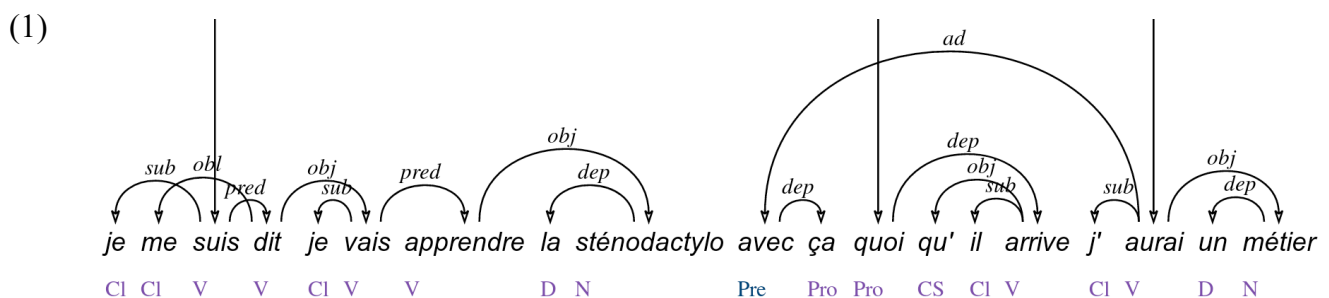
Nous présentons d’abord les schémas Rhapsodie et Orféo qui ont déjà été présentés ailleurs (Benzitoun et al. 2011, Kahane et al. 2017) en nous focalisant sur la macrosyntaxe (section 2). Dans la section 3, nous nous penchons sur les différences de ces schémas avec celui d’UD et nous proposons des aménagements d’UD pour prendre en compte certains traits importants des schémas précédents. L’accent est mis sur ce dernier schéma d’annotation.

2 Macrosyntaxe dans Rhapsodie et Orféo

Le schéma Rhapsodie s’est appuyé à la fois sur une tradition de la syntaxe de dépendance (Tesnière 1959, Mel’čuk 1988, Kahane 2001) et sur une tradition de la syntaxe de l’oral (Blanche-Benveniste 1990, 2010, Deulofeu 2003). Ce schéma présente deux originalités principales : une tentative de définir les limites de la dépendance syntaxique et une prise en compte des phénomènes de listes et plus généralement d’entassements très présents dans les corpus oraux (Blanche-Benveniste 1990, Gerdes & Kahane 2009, Kahane & Pietrandrea 2012).

Les projets d’annotation syntaxique de corpus écrits ne se posent généralement pas la question des limites de la syntaxe, considérant que celles-ci sont données par la structuration du texte en phrases, lesquelles phrases étant délimitées par les ponctuations majeures. Nous ne nous appesantirons pas sur le fait que la définition de la phrase ne va pas de soi, même à l’écrit (Pietrandrea et al. 2014). Il est en tout cas clair qu’on ne peut se baser sur la ponctuation à l’oral et qu’ajouter celle-ci à des transcriptions pose divers problèmes. Dans le cadre de Rhapsodie, nous nous sommes basés sur les travaux de Blanche-Benveniste (1990, 2010) qui distinguent deux niveaux de cohésion syntaxique.

Le premier niveau, appelé *microsyntaxe*, repose sur une définition restrictive de la *rection* qui ne prend en compte que les syntagmes susceptibles d’être pronominalisés ou clivés. (Le terme *rection* est à prendre dans un sens plus large que celui de valence ou de sous-catégorisation et inclut des ajouts et compléments circonstanciels.) Les relations microsyntaxiques sont encodées par des relations de dépendance entre mots. Considérons l’exemple (1), annoté selon le schéma de Rhapsodie :



L’énoncé (1) est divisé en trois unités microsyntaxiques, qui sont les projections des trois arbres de dépendance, chacun avec sa racine, marquée par une dépendance verticale. Nous considérons qu’il y a une relation de rection entre *je me suis dit* et *je vais apprendre la sténodactylo*, car *je me suis dit* n’est pas saturé et *je vais apprendre la sténodactylo* est pronominalisable (*je me suis dit ça*). Dans la deuxième partie de l’énoncé, on peut contraster *avec ça* qui est clivable (*c’est avec ça que j’aurai un*

métier) et *quoi qu'il arrive* qui ne l'est pas (**c'est quoi qu'il arrive que j'aurai un métier*) et forme donc une unité microsyntactique indépendante.

Le deuxième niveau de cohésion syntaxique, la macrosyntaxe, repose sur l'appartenance à une même *unité illocutoire* et l'impossibilité de former une unité autonome (Benzitoun et al. 2010). L'unité illocutoire, qui est donc l'unité maximale de la macrosyntaxe, est organisée autour d'un *noyau*, qui porte la force illocutoire et réunit autour de ce noyau des composantes périphériques non autonomisables. Dans l'exemple précédent, *j'aurai un métier* forme un noyau tandis que *avec ça* et *quoi qu'il arrive* sont des composantes périphériques, que nous appelons *prénoyaux*. Le tout forme une unité illocutoire. D'autre part, *je vais apprendre la sténodactylo* forme un noyau et une unité illocutoire, qui est elle-même enchâssée dans l'unité illocutoire que forme *je me suis dit* avec son objet. Nous notons cela en utilisant les balises suivantes : // pour marquer la fin d'une unité illocutoire, < pour noter la fin d'un prénoyau, + pour indiquer que cette frontière macrosyntaxique n'est pas une frontière micro et [] pour marquer l'enchâssement :

(2) *je me suis dit [je vais apprendre la sténodactylo //] // avec ça <+ quoi qu'il arrive < j'aurai un métier //*

Dans Rhapsodie, micro et macrosyntaxe sont annotées séparément. L'un des avantages est de mettre en évidence une relative indépendance des deux niveaux d'analyse. Ainsi comme on le voit dans notre premier exemple, nous distinguons parmi les prénoyaux, ceux qui remplissent un rôle microsyntactique (*avec ça*) et ceux qui n'en remplissent pas (*quoi qu'il arrive*). L'indépendance entre micro et macrosyntaxe est également illustrée par les *épexégèses* ou complément différés (angl. *afterthought*), qui sont des syntagmes régis mais qui forment de nouvelles unités illocutoires. C'est le cas dans l'exemple (3) où le complément *et en particulier lorsqu'on enseigne une langue* forme un noyau séparé du précédent (ce que montre la présence de la conjonction comme *et*, en plus de la prosodie) tout en étant régi (*c'est lorsqu'on enseigne une langue que ça peut être dans la région*).

(3) *donc < ça peut être dans la région comme "euh" à l'étranger //+ et en particulier lorsqu'on enseigne une langue //*

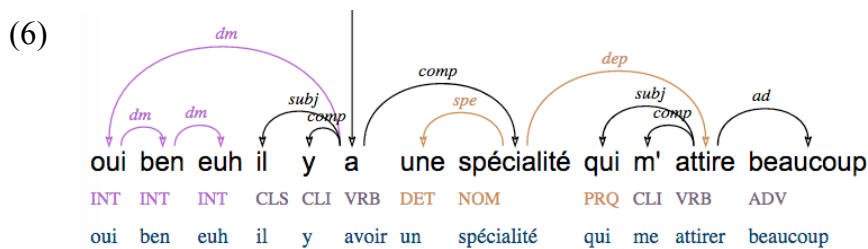
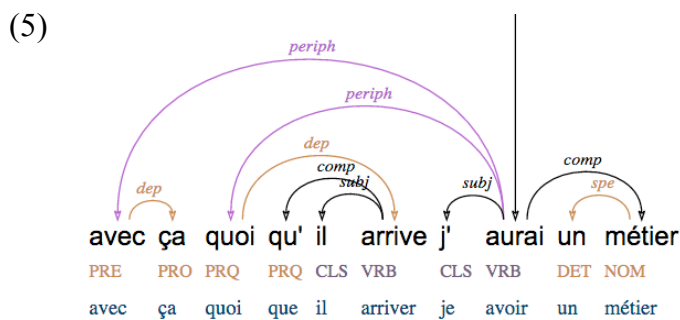
Notons encore que l'on distingue parmi les composantes périphériques, des unités que nous nommons les *noyaux associés*, car elles possèdent une forme de force illocutoire qui les empêche d'être régis (Kahane & Pietrandrea 2009, Kahane & Mazziotta 2015). Ces unités, plus souvent appelées marqueurs de discours, sont marqués par "...". Il peut s'agir d'interjections ou de constructions verbales très contraintes.

(4) *il y a des spécialités délaissées parce que ça ne rapporte pas assez d'argent "je sais pas" la psychiatrie des choses comme ça // [...] "oui" "ben" "euh" il y a une spécialité qui m'attire beaucoup //*

L'encodage de la microsyntaxe dans Rhapsodie distingue parmi les dépendants du verbe *sub* (sujet), *pred* (prédicatif, pour les formes verbales composées, les constructions avec verbe modal et aussi les constructions prédictives de l'adjectif), *obj* (objet direct), *obl* (oblique) et *ad* (ajout). Les dépendants des autres catégories sont tous *dep* (dépendant). Voir (1). Rhapsodie propose également une analyse fine des phénomènes d'entassements (coordination, reformulation, etc.) que nous n'avons pas la place de présenter ici (Kahane & Pietrandrea 2012, Bawden et al. 2014).

Dans le projet Orféo, nous souhaitons passer à l'échelle, pouvoir analyser davantage de données manuellement et entraîner un analyseur avec des résultats acceptables pour le reste du corpus (Debaisieux et al. 2016, Kahane et al. 2017). La micro et la macrosyntaxe ont été fusionnées dans un même schéma en dépendance. Les prénoyaux ont été systématiquement rattachés à la racine du noyau par la relation *periph* (périphérique), ce qui écrase la distinction entre régis et non régis, comme en

(5). Les noyaux associés ont été rattachés au mot qui précède par la fonction *dm* (discourse marker) (leur ancrage dans le reste de l'énoncé est généralement non pertinent), cf. (6).



Le schéma Orféo introduit également la relation *parenth* pour les unités illocutoires parenthétiques (voir exemple en section 3) et la fonction *insert* pour les incises, qui ont des propriétés différentes (Bonami & Godard 2008) et étaient mal analysées dans Rhapsodie. Le nombre de fonctions microsyntaxiques a été réduit, une unique fonction *comp* (compléments) couvrant *obj*, *obl* et *pred* (à l'exception des auxiliaires marqués *aux*). Les épexégèses ne sont plus annotées, l'analyse micro s'arrêtant à la frontière des unités illocutoires et des tours de parole.

3 Un schéma UD pour l'oral

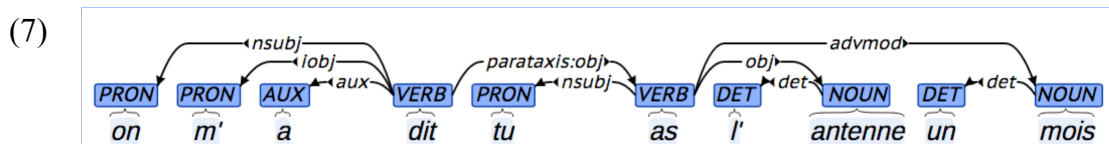
UD est un projet qui propose un schéma d'annotation universel qui a été appliqué à une cinquantaine de langues (universaldependencies.org). Le schéma UD a été développé à partir de corpus écrit et ne prend guère en compte la macrosyntaxe. Nous allons présenter le schéma UD, les différences avec les schémas Rhapsodie et Orféo et nous ferons en même temps des propositions pour un schéma d'annotation qui subsume les trois.

Les noms de relation en UD distinguent en partie la fonction grammaticale (sujet et objet direct sont distingués des autres dépendants du verbe), mais dépendent surtout de la catégorie des gouverneurs et dépendants de la relation : ainsi un groupe prépositionnel dépendant d'un verbe sera systématiquement déclaré *obl* qu'il s'agisse d'un actant du verbe (*parler à qqn*, *compter sur qqn*), d'un ajout (*travailler dans la chambre*, *partir à trois heures*) ou d'un syntagme non régi (*à mon avis*, *d'après lui*), alors qu'il sera *nmod* (nominal modifier) s'il dépend d'un nom (*la visite de Paul*, *la visite dans ma chambre*). À l'inverse, des syntagmes qui appartiennent au même paradigme instancieront des relations différentes (*iobj* pour *lui parler*, *advmod* pour *compter dessus*, *travailler ici*, *maintenant*). De même, l'objet de *vouloir* sera *obj* pour un nom (*vouloir une banane*) ou un pronom (*la vouloir*), *xcomp* pour un infinitif (*vouloir manger*) et *ccomp* pour une complétive (*vouloir qu'il mange*). Par ailleurs, les mots fonctionnels tels que prépositions et conjonctions de subordination sont systématiquement dépendants de la tête lexicale qu'ils introduisent. Ici notre principale modification est de maintenir la distinction entre véritables compléments obliques (*obl:comp* pour *parler à qqn*), ajouts (*obl:mod* pour *partir à trois heures*) et non régis (*obl:periph* pour *à mon avis*).

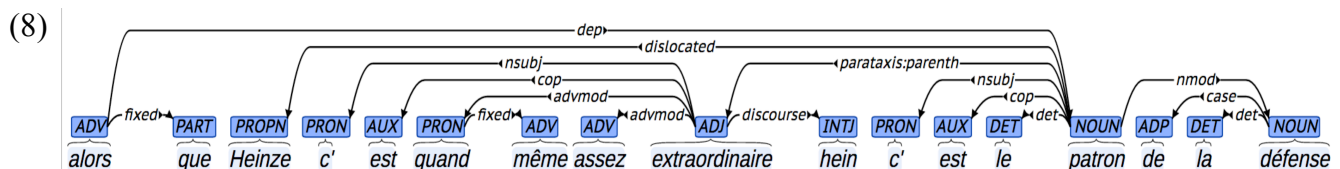
Pour comprendre ces choix atypiques par rapport à des analyses dépendantielles antérieures, il faut rappeler l'histoire des Universal Dependencies : le schéma UD est une évolution du schéma d'annotation dépendantielle de Stanford qui est lui-même le résultat d'une conversion des sorties d'un analyseur en constituants (De Marneffe et al. 2006). Dans un arbre de constituants, les relations syntaxiques sont des relations entre catégories syntaxiques de constituants, ce que traduisent encore les noms de relations de UD comme *nsubj* ou *advmod*. Mélanger catégorie et fonction a des désavantages : un exemple comme *il est architecte et fier de l'être* est problématique si on raisonne en terme de catégorie (on a coordonné un nom et un adjectif), mais ne l'est plus dès lors qu'on prend en compte qu'il s'agit d'une coordination de deux éléments remplissant la même fonction vis-à-vis de la copule.

Dans le guide UD actuel, pour les éléments en prénoyau, deux relations particulières sont considérées : *dislocated* pour les dislocations (*lui, je ne le connais pas*) et *vocative* pour les adresses à un interlocuteur (*vous, mettez-vous là*). Les autres éléments en prénoyau reçoivent une relation « microsyntaxique » en fonction de leur catégorie, qu'ils soient régis ou non. Pour notre part, nous souhaitons faire la différence entre les éléments régis et les autres afin d'encoder la valence. Les éléments non régis, qui ne peuvent être clivés et ne peuvent être déplacés dans le noyau qu'en incise, seront donc marqués comme intrinsèquement périphériques (*advmod:periph* pour *franchement, il ne parle pas beaucoup* ; *obl:periph* pour *à mon avis, ça va bien se passer* ; *advcl:periph* pour *si je comprends bien, tout est à refaire*).

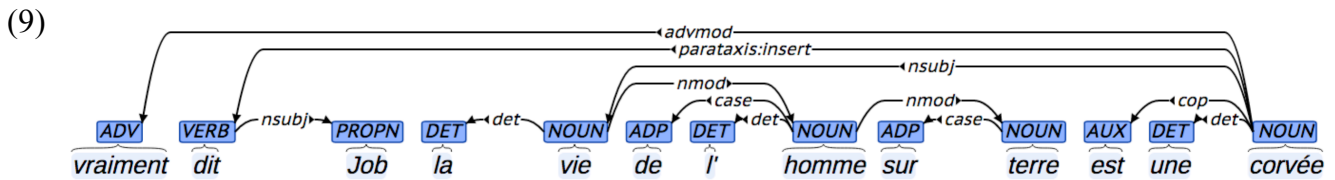
Les marqueurs de discours ont la relation *discourse* et dépendent de la tête du syntagme où ils apparaissent. Ceci s'applique aux interjections (*il va venir, hein ?*), mais pas aux noyaux associés verbaux (*il va venir, tu penses ?*) qui ont la relation *parataxis*. Il s'agit encore une fois d'une distinction basée sur la catégorie qui nous semble injustifiée. La relation *parataxis* de UD couvre tous les cas où une construction verbale finie n'est pas introduite par un complémenteur. Il y a là au moins six situations différentes que nous souhaitons distinguer. La première est celle des discours directs introduits par un verbe recteur. Dans l'exemple (7), *on m'a dit* ne peut constituer un noyau sans le syntagme *tu as l'antenne un mois*. Il s'agit donc bien de l'objet du verbe *dire* auquel nous donnons la fonction *parataxis:obj*.



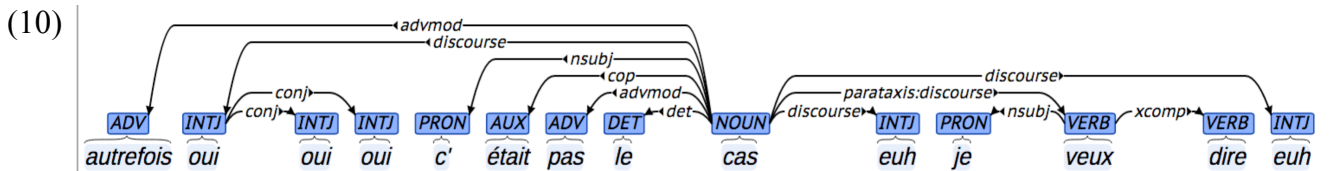
La deuxième est celle des parenthétiques, qui sont de véritables unités illocutoires qui pourraient être autonomes, mais qui viennent s'insérer au milieu d'une autre unité illocutoire, comme *c'est quand même assez extraordinaire* en (8). Nous les notons *parataxis:parenth*.



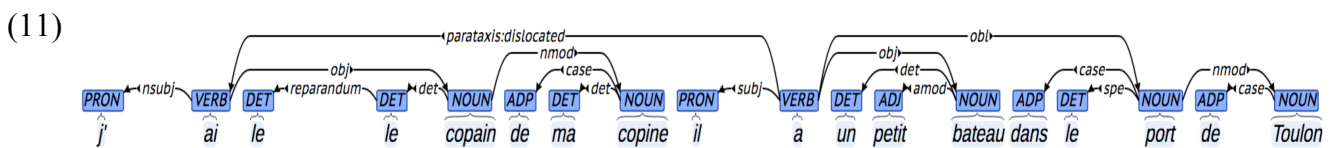
La troisième situation est celle des incises, comme *dit Job* en (9), qui se distinguent des parenthétiques par le fait qu'elles ne sont pas saturées et ne pourrait pas former un énoncé indépendant (Bonami & Godard 2008). Elles relèvent d'un registre plus formel (news, contes, etc.). En français, elles se caractérisent en plus par l'inversion du sujet. Nous les notons *parataxis:insert*. Une incise peut se trouver en fin d'unité illocutoire.



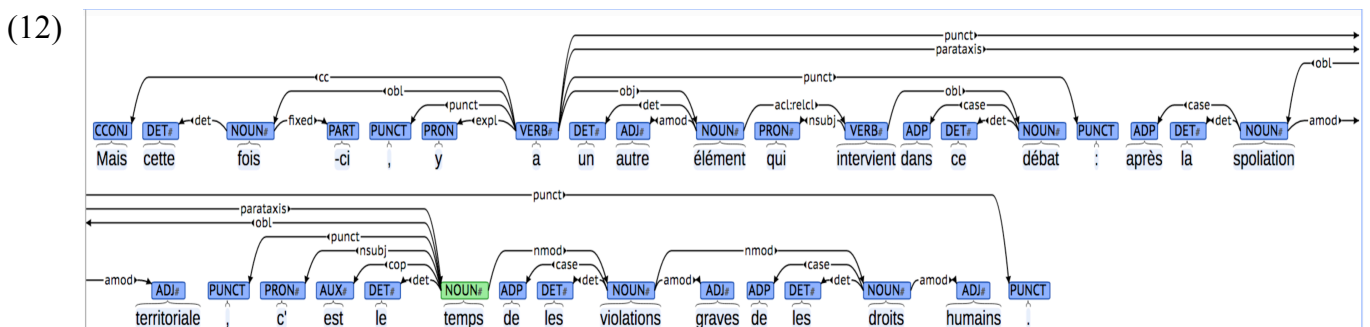
La quatrième est celle des noyaux associés, comme *je veux dire* en (10), qui se distinguent des parenthétiques par le fait de ne pas accepter d'expansions (**je veux dire maintenant c'est très varié*) et d'avoir des sujets pronominaux à la 1^{ère} ou 2^e personne (*je pense, tu pense pas ?, on dirait*). Nous les notons *parataxis:discourse*.



Le cinquième cas est celui des constructions verbales en prénoyau qui fonctionnent comme des dislocations. Dans l'exemple (11), *j'ai le le copain de ma copine* n'est clairement pas une unité illocutoire (ça ne pourrait pas fonctionner comme une assertion autonome). Il s'agit d'un prénoyau, qui peut commuter avec un syntagme nominal comme *le copain de ma copine*. Nous proposons de lui attribuer la fonction *parataxis:dislocated* ou simplement *dislocated*.



UD fait un sixième usage de la relation *parataxis*, pour relier des propositions juxtaposées, comme en (12), extrait des corpus UD.



Nous ne retenons pas ce cas, car nous considérons que nous avons alors deux unités illocutoires formant deux arbres de dépendance séparés. Si cette relation était maintenue, il faudrait la traiter comme un cas particulier de coordination (*parataxis:conj*).

4 Conversion

La table 1 présente la correspondance entre les trois schémas d'annotation discutés ici. Les relations en gras sont les relations que nous introduisons dans le schéma UD pour ne pas perdre les distinctions faites dans les schémas Rhapsodie ou Orféo. Même si UD utilise beaucoup plus de noms de relations,

la plupart des conversions sont triviales puisqu'entièrement déductibles du nom de la relation dans Rhapsodie ou Orféo et de la catégorie du dépendant. Par exemple, la relation *dep* deviendra *nmod*, *amod*, *acl*, *nummod*, *case* ou *mark*, selon que le dépendant est un nom, un adjectif, un verbe, un numéral, une préposition ou une conjonction. La seule distinction qui n'est pas automatique est entre *dislocated* et *vocative* et elle devra donc être faite manuellement. Nous n'avons pas discuté les relations paradigmatiques et les expressions polylexicales qui posent d'autres problèmes.

Notons également que les noms de relations ne sont pas les seules modifications à faire. Nous devons également revoir la structure et l'orientation de certaines relations, puisque les relations *case* et *mark* sont inversées par rapport à la relation *dep* (*à dep* > *Paris* donne *à* < *case Paris*). Par ailleurs, les marqueurs de discours qui sont attachés au mot précédent dans Orféo sont attachés à l'élément « à la tête de la proposition voisine la plus pertinente », ce que nous interprétons comme le mot le plus haut dans l'arbre de dépendance accessible sans couper une autre dépendance.

Rhapsodie	Orféo	UD+
sub	sub	nsubj, csubj
obj, obl	comp	obj, iobj, xcomp, ccomp, obl:comp
obj + [...//]		parataxis:obj
ad	ad	advmod, advcl, obl:mod
pred	aux, comp	aux, cop
dep	dep	nmod, amod, acl, nummod, case, mark
	spe	det
root + "..."	dm	discourse, parataxis:discourse
root + <	periph	dislocated, parataxis:dislocated vocative, obl:periph , advmod:periph , advcl:periph ,
root + (...//)	parenth	parataxis:parenth
	insert	parataxis:insert

Table 1. Correspondance entre les 3 schémas d'annotation

5 Conclusion

Concernant la macrosyntaxe, le schéma UD est parfois plus riche que les schémas Rhapsodie et Orféo (cf. la distinction entre *dislocated* et *vocative*), parfois comparable (*discourse*), mais souvent plus pauvre (adverbes, groupes prépositionnels ou propositions non régis, *parataxis*). Ceci nous a amené à proposer un schéma d'annotation qui subsume les trois. Ce schéma pourrait être simplifié, puisque certaines distinctions (*obl:comp* vs *iobj*, *discourse* vs *parataxis:discourse*, ...) peuvent être récupérées à partir des catégories des éléments en jeu et nous espérons que le schéma UD évoluera dans cette direction, en prenant davantage en compte les travaux menés depuis des dizaines d'années en syntaxe de dépendance et dans d'autres domaines de la linguistique.

Remerciements

Nous remercions Bernard Caron, qui nous a entraîné dans ce travail, ainsi que les étudiants qui travaillent avec nous : Sandra Bellato, Aurélie Collomb, Marine Courtin, Aline Etienne et notamment Chunxiao Yan, qui a réalisé la conversion automatique en UD.

Références

- BENZITOUN C., DISTER A., GERDES K., KAHANE S., PIETRANDREA P., SABIO F. (2010). tu veux couper là faut dire pourquoi : propositions pour une segmentation syntaxique du français parlé, *Actes de CMLF*, La Nouvelle-Orléans, 16 p.
- BLANCHE-BENVENISTE, C. et al. (1990). *Le français parlé: études grammaticales*. Paris, Éditions du CNRS.
- BLANCHE-BENVENISTE, C. (2010). *Approches de la langue parlée*. Paris, Éditions Ophrys.
- BONAMI, O., GODARD, D. (2008). Vers une syntaxe générale des constituants incidents. *Actes du séminaire du LIDILE*, Université de Haute Bretagne.
- CANDITO, M., SEDDAH, D. (2012). Le corpus Sequoia: annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. Actes de *TALN*.
- CRESTI, E. (2000). *Corpus di italiano parlato*. Florence, Accademia della Crusca.
- DEBAISIEUX, J. M., BENZITOUN, C., & DEULOFEU, H. J. (2016). Le projet ORFEO : Un corpus d'études pour le français contemporain. *Revue Corpus*, 15, 91-114.
- DE MARNEFFE, M.-C., MACCARTNEY B., MANNING C. D. (2006) Generating typed dependency parses from phrase structure parses. *Proceedings of LREC*.
- DEULOFEU, J. (2003). L'approche macrosyntaxique en syntaxe: un nouveau modèle de rasoir d'Occam contre les notions inutiles. *Scolia*, 16, 77-95.
- GERDES, K., KAHANE, S. (2009) Speaking in piles: Paradigmatic annotation of French spoken corpus, *Processing of the fifth Corpus Linguistics Conference*, Liverpool, 15 p.
- KAHANE S. (2001). Grammaires de dépendance formelles et théorie Sens-Texte, Tutoriel, Actes de *TALN*, vol. 2, 17-76.
- KAHANE S., PIETRANDREA P. (2009). Les parenthétiques comme « Unités Illocutoires Associées » : un approche macrosyntaxique. In M. Avanzi, J. Glikman (éd.), *Les Verbes Parenthétiques : Hypotaxe, Parataxe ou Parenthèse ?*, in *Linx*, 61, 49-70.
- KAHANE S., PIETRANDREA P. (2012). La typologie des entassements en français, *Actes du 3e congrès mondial de linguistique française (CMLF)*, Lyon, 1809-1828.
- KAHANE S., DEULOFEU J., GERDES K., VALLI A., NASR A. (2017). Annotation micro- et macrosyntaxique manuelle et automatique de français parlé, *Journées Floral*, Orléans, 4 p.
- LACHERET A., KAHANE S., BELIAO J., DISTER A., GERDES K., GOLDMAN J.-P., OBIN N., PIETRANDREA P. TCHOBANOV A. (2014). Rhapsodie: un Treebank annoté pour l'étude de l'interface syntaxe-prosodie en français parlé, *Actes du 4e congrès mondial de linguistique française (CMLF)*, Berlin.
- MCDONALD R., NIVRE J., QUIRMBACH-BRUNDAGE Y., GOLDBERG Y., DAS D., GANCHEV K., HALL K., PETROV S., ZHANG H., TACKSTROM O., BEDINI C., BERTOMEU CASTELLO N., LEE J. (2013) Universal Dependency Annotation for Multilingual Parsing, *Proceedings of ACL*.

- MEL'CUK I. A. (1988) *Dependency syntax : Theory and practice*, State University of New York Press.
- PIETRANDREA P., KAHANE S., LACHERET A., SABIO F. (2014). The notion of sentence and other discourse units in corpus annotation. In T. Raso, H. Mello, M. Pettorino (eds.), *Spoken Corpora and Linguistic Studies*, John Benjamins, Amsterdam.
- SANGUINETTI M., BOSCO C. (sous presse). PartTUT: The Turin University Parallel Treebank". In Basili, Bosco, Delmonte, Moschitti, Simi (eds) *Harmonization and development of resources and tools for Italian Natural Language Processing within the PARLI project*, LNCS, Springer Verlag.
- TESNIERE L. (1959) *Eléments de syntaxe structurale*, Klincksieck.