Rhapsodie: un Treebank annoté pour l'étude de l'interface syntaxe-prosodie en français parlé

Anne Lacheret(1), Sylvain Kahane(1), Julie Beliao(1), Anne Dister(2), Kim Gerdes(3), Jean-Philippe Goldman(4), Nicolas Obin(5), Paola Pietrandrea(6), Atanas Tchobanov (1)

(1) Modyco, UMR7114, Université Paris Ouest Nanterre
(2) Université Saint-Louis – Bruxelles
(3) LPP, UMR7018,Université Paris Sorbonne Nouvelle & CNRS
(4) Université de Genève
(5) IRCAM - UMR STMS IRCAM-CNRS-UPMC, Paris
(6) LLL, Université François Rabelais & CNRS
anne@lacheret.com; sylvain@kahane.fr; julie@beliao.fr; anne.dister@usaintlouis.be; kim@gerdes.fr; Jean-Philippe.Goldman@unige.ch; Nicolas.Obin@ircam.fr; paolapietrandrea@gmail.com; atanas@u-paris10.fr

1 Introduction

L'objet de notre communication est de présenter la ressource Rhapsodie, un Treebank annoté en syntaxe et en prosodie pour l'analyse du discours en français parlé. Il s'agit d'un corpus multilocuteurs (89 sujets, hommes et femmes) et multigenres (± spontané, ± planifié, entretiens en face à face, interviews, émissions radiophoniques et télévisuelles), composé de 57 échantillons courts (5 minutes en moyenne), soit au total trois heures de parole, transcrites orthographiquement (33000 mots) et phonologiquement, et alignées au son (phonèmes, syllabes, mots, tours de parole, chevauchements, Goldman 2011).

L'objectif majeur du projet, conduit dans le cadre de *l'ANR corpus, données et outils de la recherche en sciences humaines et sociales* (appel 2007), a été de définir des schémas d'annotation explicites et reproductibles en prosodie et en syntaxe, permettant l'étude approfondie de l'interface discours/prosodie/syntaxe, plus spécifiquement le rôle respectif de la syntaxe et de la prosodie (complémentarité et collaboration des modules) dans la segmentation du discours en unités élémentaires dans différents genres discursifs (Lacheret et al. à paraître).

Par rapport aux treebanks syntaxiques existants[i], Rhapsodie présente quatre caractéristiques majeures. D'une part, il vient enrichir le réservoir encore très petit des treebanks syntaxiques dévolus à l'oral (moins d'une dizaine sont distribués à l'heure actuelle, voir notamment the Switchboard Corpus of Penn Treebank : Meter et al. 1995, the British component of the International Corpus of English : Nelson et al. 2002 et, pour le français, the Ester treebank of French : Cerisara et al. 2010). Par ailleurs, dans le sillage du corpus C-Oral-Rom (Cresti et Moneglia 2005), une annotation macrosyntaxique est couplée de façon innovante à l'annotation syntaxique standard que l'on trouve dans les Treebanks actuels. Ensuite, Rhapsodie constitue, à notre connaissance, le premier exemplaire d'un corpus prosodique arboré, toute langue confondue (pour différents projets récents relatifs à l'annotation de la prosodie : objectifs, méthodologies et granularité de l'annotation, voir en particulier: the Spoken Dutch Corpus : Schuurman et al. 2004, the Hong Kong Corpus of Spoken English, HKCSE : Cheng et al. 2008). Enfin, grâce à une implémentation dans une structure orientée objet (Beliao ici même) où les informations temporelles restent accessibles dans les arbres syntaxiques, il permet de façon très flexible, l'exploration simultanée des différents types d'arbres et, ainsi, l'exploration automatique des corrélations intonosyntaxiques à tous les niveaux des arboresences.

La ressource est téléchargeable librement (www.projet-rhapsodie.fr): les enregistrements (wav/mp3), les analyses acoustiques (F0 brutes nettoyées manuellement et F0 stylisées automatiquement, format pitch), la transcription orthographique (txt), l'annotation macrosyntaxique (txt et format tabulaire), l'annotation microsyntaxique (format tabulaire), l'annotation et la segmentation prosodique (textgrid, xml, format tabulaire), et les metadonnées (xml, html) sont téléchargeables selon les termes de la licence Creative

Commons Attribution – Noncommercial - Share Alike 3.0 France. Les métadonnées sont encodées dans le format IMDI-CMFI et peuvent être parsées en ligne.

L'objet de cette communication est de présenter en les justifiant les choix effectués pour l'établissement des métadonnées et les schémas d'annotation retenus en syntaxe et en prosodie.

2 Corpus design et métadonnées

Au cœur du projet Rhapsodie : (i) l'objectif de modéliser l'interface intonosyntaxique sur un jeu de constructions annotées en prosodie et en syntaxe, suffisamment vaste pour permettre les généralisations descriptives sur différents genres de discours, (ii) l'hypothèse selon laquelle il existe une relation étroite entre les caractéristiques typologiques d'un texte, i.e. les patrons textuels définis sur les bases de critères strictement formels et le genre dont il est issu, i.e. les traits situationnels qui le caractérisent(tableau 1)[ii].

Traits situationnels → Marqueurs formels → Genre discursif

Tableau 1. Variables linguistiques, situation et genre de discours : une causalité à modéliser dans la langue parlée

2.1 Corpus design

Les lignes directrices pour l'échantillonnage du corpus Rhapsodie ont été les suivantes : (i) collecter un ensemble d'échantillons suffisamment diversifié en termes de typologie textuelle ; (ii) disposer d'un panel de locuteurs assez large pour éviter les idiosyncrasies individuelles ; (iii) étant donné ces deux premières contraintes et étant donné le coût temporel colossal que suppose une annotation robuste sur le versant de la syntaxe comme de la prosodie, les échantillons sont nécessairement courts (5 minutes en moyenne).

Dans la mesure où il n'existe pas à l'heure actuelle de corpus de référence pour le français parlé dans lequel nous aurions pu aller puiser pour atteindre cet objectif de diversité et d'équilibre typologique, nous avons, dans un premier temps, extrait nos données de sources existantes (sources institutionnelles dont notamment PFC : Durand et al. 2009; ESLO : Eshkol-Taravella et al. 2011, CFPP2000 : Branca et al. 2012, et corpus d'étudiants)[iii]. Ce premier jeu d'échantillons a ensuite été complété par des données plurielles (multimédia, descriptions de films, map task, etc.) récoltées dans le cadre du projet Rhapsodie afin d'assurer l'équilibre de l'échantillonnage fixé au préalable. Ce type de pratique, inexistante jusqu'alors, nous a contraints à mettre en place une procédure ad hoc préalablement à la mise en ligne du corpus Rhapsodie pour le respect de la propriété intellectuelle des équipes à l'origine de la constitution des sources (accès à l'enregistrement original et aux publications de référence qui décrivent la source et sa constitution notamment).

2.2 Métadonnées

Concernant le choix des descripteurs situationnels et l'établissement des métadonnées, si les linguistes s'accordent pour considérer qu'un genre de discours peut être décrit comme un objet multifactoriel qui intègre un ensemble de variables socio-communicatives a priori orthogonales (Biber et al. 1999, Koch et Oesterreicher 2001)[iv] et si, en conséquence, les angles d'attaque pour caractériser un genre sont pluriels, nous avons privilégié six traits situationnels majeurs. En premier lieu, en distinguant les monologues des dialogues, nous opposons les discours produits par un unique locuteur à l'intention d'une large audience ou d'un petit auditoire et les discours produits par au moins deux locuteurs, plus ou moins interactifs. Ensuite, l'opposition parole privée vs. parole publique différencie d'une part les échantillons tirés d'entretiens en face à face qui peuvent être extraits d'interactions au quotidien ou avoir été réalisés en présence d'un informateur, d'autre part, les conférences et les émissions télévisuelles ou radiophoniques. Dans les deux types de parole, les thèmes de discours sont larges et diversifiés ; dans la parole publique,

les émissions de nature variée (entretiens politiques, débats d'idées, émissions de vulgarisation scientifique, etc.). Au final, chaque type de parole (monologale vs dialogale), qu'il soit privé ou publique est renseigné selon les variables suivantes : taux de planification, degré d'interactivité, type de tâche (interview, sermon, émission sportive, etc.), canal de communication et type de séquence discursive majoritairement représenté dans l'échantillon, de l'argumentation à la description neutre (tableau 2).

Type de parole	**Privé, public**	**Monologue**
		Dialogue
	<Planification>	Spontané, semi-spontané, planifié
	<Interactivité>	Interactif, semi-interactif, non-interactif
	<Canal de communication>	Face à face vs. Conférence, émission radiophonique ou télévisuelle
	<Type de séquence>	Argumentative, descriptive, procédurale, oratoire

Tableau 2. Variables situationnelles dans Rhapsodie

Quant au choix du standard pour consigner les métadonnées, nous avons cherché un outil flexible et inclusif qui permette une description textuelle exhaustive, afin, d'une part, de fournir des informations complètes sur l'origine des sources (auteurs, objectifs scientifiques, accès à l'échantillon dans le réservoir d'origine, etc.), d'autre part, de décrire précisément les annotations pour chaque échantillon (type d'annotation, annotateurs) qui constituent le cœur du projet Rhapsodie. Le format IMDI-CMDI, développé au Max Planck de Nijmegen (CMDI, http://www.clarin.eu/cmdi, Broeder et al. 2012), nous a semblé le plus adapté à ces besoins spécifiques, et a donc été sélectionné pour encoder les métadonnées Rhapsodie.

3 Annotation syntaxique

Combinant le modèle syntaxique développé par l'Ecole d'Aix-en-Provence (Blanche-Benveniste et al. 1990) et le modèle pragmatique élaboré dans le cadre du projet Lablita (Cresti 2000), deux niveaux de cohésion syntaxique ont été fixés : le niveau macrosyntaxique qui pilote la cohésion illocutoire à l'intérieur de l'énoncé, i.e. l'ensemble des relations formelles qui se déploient entre segments pour former un acte illocutoire, et le niveau microsyntaxique qui pilote la cohésion syntaxique et pour lequel sont annotés les catégories, les fonctions et les dépendances entre unités grammaticales. Le premier ayant déjà été largement discuté dans différents supports (voir en particulier Benzitoun et al. 2010, Deulofeu et al. 2010, Lacheret et al 2011), nous en faisons une présentation synthétique ici pour nous consacrer davantage à l'annotation microsyntaxique et à l'interface micro-macro tel qu'il a été conçu dans Rhapsodie.

3.1 Annotation macrosyntaxique

Le modèle macrosyntaxique défini dans le cadre du projet Rhapsodie se différencie d'autres modèles stratificationnels, comme par exemple celui proposé par Berrendonner (1990) qui considère les unités maximales de la microsyntaxe, comme des points d'ancrage pour la macrosyntaxe. L'annotation macrosyntaxique se fonde ici plutôt sur une approche modulaire, comme celle élaborée par l'école d'Aix-en-Provence (Blanche-Benveniste 1990), ou par Cresti (2000), où les deux types d'organisation sont envisagés comme orthogonaux et donc comme pouvant opérer de concert sur les mêmes séquences.

De la même façon, notre modèle considère l'organisation macrosyntaxique comme un principe de cohésion opérant de manière indépendante de la prosodie. Au final, deux critères principaux ont été retenus pour segmenter un énoncé en unités illocutoires (désormais UI) : (i) caractériser l'organisation syntaxique indépendamment de toute organisation prosodique ou du moins de tout cadre théorique prosodique; (ii) proposer des critères syntaxiques de segmentation explicites permettant aux annotateurs d'appliquer de manière aussi homogène que possible les critères présidant à l'annotation du corpus. Le principal critère retenu est la non-autonomie : les segments qui ne peuvent former un énoncé autonome sont considérés comme dépendants macrosyntaxiquement.

L'annotation a été effectuée manuellement par l'équipe « syntaxe » du projet Rhapsodie à partir de critères formels, i.e. les propriétés distributionnelles de segments à annoter. (Deulofeu et al. 2010). Chaque échantillon est segmenté en une succession d'UI; chaque UI est potentiellement composée de deux types d'unités : un noyau obligatoire (en gras dans les exemples ci-dessous), des ad-noyaux optionnels (pré-, post et in-noyaux), qui peuvent se placer avant, après, voire à l'intérieur du noyau (cf. en infra, les exemples (1) et (2), où '<' suit un pré-noyau et précède un noyau ou un autre pré-noyau, '>' précède un post-noyau et suit un noyau ou un autre post-noyau et '//' marque la frontière droite d'une UI).

(1) alors < là < la psychiatrie < **c'est autre chose** // [Rhap-D0006, CFPP2000]

(2) **ça a duré dix ans** > le silence autour de moi // [Rhap-D2001, Corpus Mertens]

L'UI est l'unité maximale de notre annotation macrosyntaxique. Il s'agit, comme son nom l'indique, d'une unité porteuse d'une force illocutoire (assertion, question, exclamation). Le noyau est le segment qui porte la force illocutoire proprement dite et possède à ce titre une autonomie illocutoire (il constitue à lui seul un énoncé bien formé). Les ad-noyaux sont des composantes périphériques qui n'ont pas d'autonomie illocutoire et sont donc dans la dépendance illocutoire du noyau. Un certain nombre de critères syntaxiques permettent de caractériser les noyaux, comme la possibilité de les nier ou de les interroger ou de les mettre dans la portée d'un adverbe d'énonciation.

Précisons qu'une UI peut être intégrée dans une autre UI selon deux modalités : (i) les enchâssements d'UI (i) le discours rapporté en (3) et les greffes[v] en (4), notés [ ]), et (ii) les insertions d'UI (les parenthèses en (5) notées ( ))

(3) Marcel Achard écrivait [elle est très jolie // elle est même belle // elle est élégante //] // [Rhap-D2001, Corpus Mertens]

(4) vous suivez la ligne du tram qui passe vers la [je crois que c'est une ancienne caserne //] // [Rhap-M003, Corpus Avanzi]

(5) alors que Heinze (c'est quand même assez extraordinaire hein //) c'est le patron de la défense // [Rhap-D2003, Rhapsodie]

Enfin, les noyaux associés (appelés *unités illocutoires associées* dans Kahane & Pietrandrea 2012) sont également annotés. Il s'agit de segments munis d'un opérateur illocutoire, mais qui n'ont pas toutes les propriétés d'autonomie illocutoire d'un noyau ordinaire (cf. en infra l'exemple 6 où les noyaux associés sont entre "…"). Ces unités sont en partie équivalentes à celles définies par Morel & Danon-Boileau (1998) sous le terme de *marqueurs discursifs*, mais comprennent aussi des segments généralement traités comme parenthétiques (exemple 7).

(6) "ouais" il y a un accident "quoi" // [Rhap-M0023, Rhapsodie]

(7) mes souvenirs les plus anciens sont "je crois" des souvenirs de Mulhouse sûrement pas de Cannes // [Rhap-D2004, Corpus Lacheret].

3.2 Annotation microsyntaxique

L'analyse microsyntaxique est l'analyse syntaxique de surface usuelle, celle que considère la plupart des treebanks en syntaxe, de l'écrit comme de l'oral (Hoekstra et al. 2003, Abeillé & Crabbé 2013). Dans la mesure où notre annotation microsyntaxique est complétée par l'annotation macrosyntaxique, nous pouvons adopter une définition assez restrictive de la rection syntaxique. Cependant, nous ne limitons pas la microsyntaxe aux unités illocutoires (c'est-à-dire aux frontières macrosyntaxiques, cf. infra, figure 3), ni même aux tours de parole. De plus, avec la notion d'entassement (Kahane et Pietrandrea 2012, figure 4 et tableau 5 en infra), nous gérons un certain nombre de phénomènes massifs à l'oral (disfluences, reformulations, etc.) qui ne sont généralement pas pris en compte dans les systèmes actuels.

Tout comme l'annotation macrosyntaxique, l'annotation de la microsyntaxe, réalisée avec le logiciel d'annotation collaboratif Arborator (Gerdes, 2013), prend comme matériel d'entrée la transcription en mots orthographiques qui sont étiquetés morphosyntaxiquement (tableau 3).

Etiquettes	Parties du discours
V	verbe (un trait mode est ajouté distinguant indicatif, subjonctif, infinitif, participe passé et participe présent)
N	nom
Adj	adjectif
Adv	adverbe
Pre	préposition
CS	conjonction de coordination
J	joncteur (conjonction de coordination, ainsi que des marqueurs de reformulation comme *c'est-à-dire* ou des clotureurs de listes comme *etcetera*)
D	déterminant
I	interjection, y compris des marqueurs de discours comme *bon, ben, euh, hein*
Qu-	mot *qu-* (relatifs et interrogatifs)
Cl	clitique, y compris *ne* et clitiques sujets
Pro	pronom
X	élément dont on ne peut déterminer la catégorie

Tableau 3. Les parties du discours dans Rhapsodie

La microsyntaxe se limite aux relations de type rection. On parle de rection lorsqu'un élément impose à un autre élément sa nature, ses marqueurs et/ou sa place. Par exemple, le complément d'objet d'un verbe est régi par ce verbe. Dans *Pierre admire le paysage*, *le paysage* est régi par la forme verbale *admire*. En effet : (i) la forme est imposée : le paradigme des éléments qui peuvent commuter avec *le paysage* se limite à des groupes nominaux ; (ii) les marqueurs sont imposés : dans le cas du complément d'objet direct en français, il n'y a pas de marqueur explicite, mais si le complément est pronominalisé (*Pierre l'admire*), une forme particulière du pronom doit être utilisée ; (iii) la place est imposée : le complément d'objet direct doit suivre le verbe (sauf formes pronominales particulières ou rares cas d'antéposition : *deux euros ça coûte*).

Nous retenons comme tests majeur pour caractériser les éléments régis par un verbe la possibilité d'être clivés (*c'est le paysage que Pierre admire*). Dans :

(8) simplement < vous êtes un peu plus jeune que moi // [Rhap-D0001, CPP2000]

simplement est considéré comme non régis, puisque non clivables (**c'est simplement que vous êtes un peu plus jeune que moi*). Il forme donc une unité microsyntaxique indépendante (figure 1). La cohésion de (8) est saisie au niveau macrosyntaxique, où *simplement*, non autonome illocutoirement, est un pré-noyau de *vous êtes plus jeune que moi*.

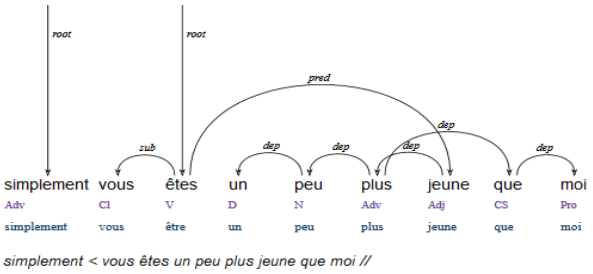

Figure 1. Annotation microsyntaxique de (8)

Comme illustré ci-dessus, la structure microsyntaxique est encodée par un graphe de dépendance. Formellement, une dépendance est une relation orientée entre deux mots (tableau 4), que nous représentons par une flèche du gouverneur vers le dépendant. Les éléments non régis, comme *simplement* ou le verbe principal *êtes* reçoivent une dépendance verticale.

Etiquettes	Fonctions syntaxiques
root	unité gouvernée par aucune autre
sub	sujet grammatical
obj	complément d'objet direct.
obl	complément actanciel oblique, incluant les compléments d'objet indirect
ad	adjoint au verbe (complément circonstanciel)
pred	tous les éléments qui forment un prédicat complexe avec un verbe (adjectif attribut, forme verbale complexe…)
dep	tous les dépendants de formes non verbales
junc	jonction entre les conjoints et le joncteur, en suivant l'analyse asymétrique de la coordination de Mel'čuk (1988)

Tableau 4. Les types de dépendance dans Rhapsodie

Une unité rectionnelle (UR) est une unité maximale pour la rection, c'est-à-dire la projection maximale d'un lexème non régi. Nous distinguons l'UR de l'unité illocutoire (UI). Une UI peut être composée de plusieurs UR. Dans l'exemple fourni dans la figure 2, nous avons quatre UR, le groupe nominal *mon premier choix*, la proposition *c'était psychologie* et les deux occurrences du marqueur de discours *euh*.

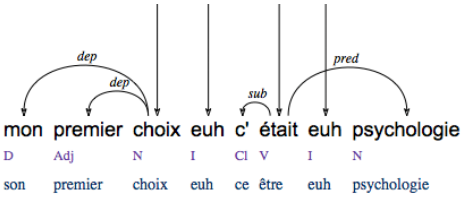

Figure 2. Exemple d'UI, *mon premier choix « euh » < c'était « euh » psychologie//,* ,[Rhap-M1001, Rhapsodie], découpée en plusieurs unités de rection

Inversement, une UR ne s'arrête pas obligatoirement à la frontière de l'UI ou même du tour de parole. Dans l'exemple suivant :

(9) $L1 donc < moi < "ben" je vais je je prends le mét~ je prends le métro le matin "bon" jusqu'au Palais Royal //+
$L2 à quelle heure "excusez-moi" //
$L1 "oui oui" je prends le métro le matin à huit heures et demie // [Rhap-D0001, CFPP2000]

la question de $L2 (*à quelle heure*) continue la construction microsyntaxique qui précède (*je prends le métro le matin jusqu'à Palais Royal*) et la réponse de $L1 (*je prends le métro le matin à huit heures et demie*) a exactement la même structure que la concaténation des deux tours de parole qui précèdent (*je prends le métro le matin à quelle heure*), comme le montre la figure 3.

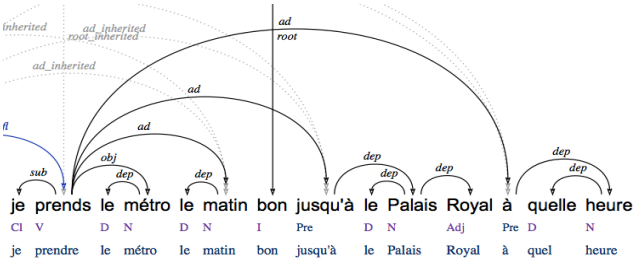

Figure 3. Annotation microsyntaxique de (9) : exemple d'une unité de rection sur deux unités illocutoires

Une annotation complète des structures de pile, dénommées aussi *entassements paradigmatiques* a également été effectuée. Il s'agit par là de repérer les constructions caractérisées par le fait que plusieurs éléments viennent occuper la même position régie (Kahane & Pietrandrea, 2012, voir 10 en gras ci-dessous). Bien qu'extrêmement fréquentes en parole spontanée, ces constructions, qui peuvent être vues comme un type particulier de relation microsyntaxique, sont ignorées des protocoles d'annotation actuels et considérées comme des erreurs de performance dans les autres treebanks[vi], ce qui de fait limite les traitements. Leur repérage systématique, quelle que soit leur nature (disfluences, répétitions, reformulations, …), étape incontournable pour une annotation syntaxique robuste, constitue une des spécificités majeures du projet Rhapsodie (tableau 5).

Etiquette	Relation paradigmatique
para_coord	coordination (*des pneus et des voitures*)
para_intens	intensification par répétition (*des dizaines et des dizaines d'années*)
para_disfl	disfluence caractérisée par une répétition (*c'était* ***un un un un*** *enfin* ***une*** *super expérience*)
para_dform	double formulation, c'est-à-dire une nouvelle dénotation pour un même référent (par une apposition par exemple : Philippe Lemaire l'avocat des parties civiles). Le deuxième élément peut être préfixé par *c'est-à-dire*. Cas particulier : une réponse occupant la même position qu'un pronom interrogation (***quand*** *vient-elle ?* ***demain***)
para_reform	reformulation visant à corriger ou à préciser une précédente dénotation (*une sorte de halle quoi de de de structure métallique*). Le deuxième élément peut être préfixé par *je veux dire*.
para_hyper	hyperonyme construit par une liste de conjoints formant chacun un sous-ensemble de la dénotation (*un ordinateur une souris euh tout ça*)
para_negot	négociation : demande de confirmation, réfutation ou correction (*des des Français enfin des Français pas simplement des Français des de l'humanité et de la lecture*)

Tableau 5. Les types de relations paradigmatiques annotés dans Rhapsodie

(10) puisque **les les les les c~ les capitales les grandes villes** ne me disaient rien du tout [Rhap-D2004, Corpus Lacheret]

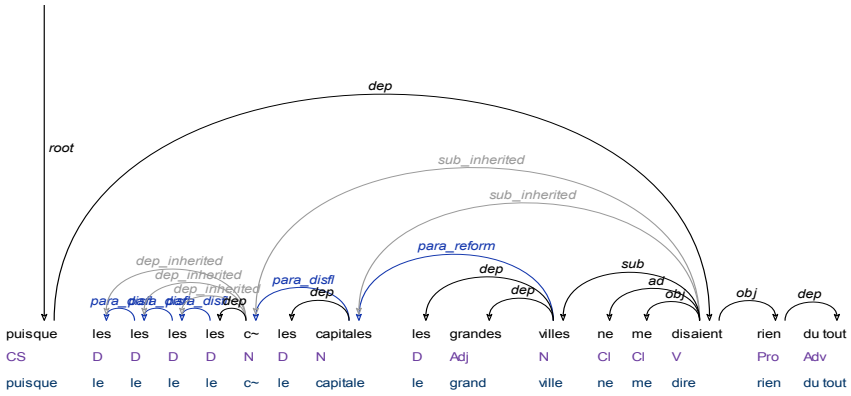

Figure 4. Annotation microsyntaxique de (10) : exemple de reformulations et de disfluences

La présence de relations paradigmatiques implique la présence de relations de dépendance héritées (en gris clair dans la figure 4). En effet, un seul des conjoints est considéré comme le dépendant du gouverneur d'un entassement, mais les autres conjoints héritent de cette relation.

4 Annotation prosodique

L'annotation prosodique de la parole continue demeure un processus complexe, en particulier pour la prosodie du français, dont la singularité typologique est caractérisée d'abord par un syncrétisme entre accentuation et intonation (Rossi 1979)[vii] et, par conséquent, un empan très variable des marqueurs dans la structure de surface. La prosodie en français a essentiellement une fonction de démarcation et de hiérarchisation, ce qui n'exclue pas, bien sûr, ses fonctions pragmatiques plurielles (focalisation d'éléments et marquage de la structure informationnelle, marqueurs de prise en charge énonciative, expression des attitudes et des émotions). Ce sont les marqueurs de segmentation et de hiérarchisation qui ont été pris en compte dans l'annotation Rhapsodie. Cette dernière repose sur deux principes élémentaires : (i) comme pour l'annotation syntaxique, elle est totalement autonome, i.e. indépendante des contraintes fonctionnelles (syntaxiques et pragmatiques) qui peuvent sous-tendre l'organisation prosodique[viii] ; (ii) l'approche qui fonde l'annotation est basée sur le modèle de l'Ecole IPO : Institut de Recherche en Perception, Eindhoven ('t Hart et al. 1990), selon lequel dans l'ensemble de l'information acoustique disponible, seuls certains traits sont sélectionnés par l'auditeur pour interpréter la structure prosodique, et c'est uniquement ceux-là dont l'annotation est censée rendre compte (voir aussi Wightman 2002).

Compte-tenu de ces deux principes, le traitement s'articule autour de trois étapes : l'annotation manuelle d'indices perceptifs jugés pertinents pour la suite du traitement ; la dérivation automatique de la structure prosodique sur les bases de cette annotation ; enfin, l'annotation tonale automatique des unités contenues dans la structure.

4.1 Annotation manuelle : proéminences et disfluences

L'annotation a été réalisée sur des fichiers préalablement alignés semi-automatiquement avec le script Easyalign (Goldman 2011) : alignement en mots, en syllabes et en phonèmes ; détection automatique des pauses (figure 5).

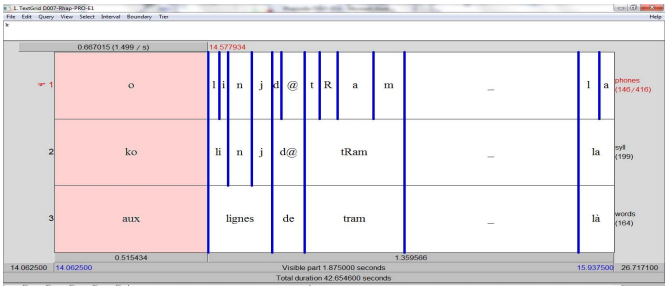

Figure 5. Matériel fourni à l'annotateur en sortie du traitement Easyalign, où '_' représente une pause

Deux indices ont été retenus pour cette phase d'annotation manuelle : les proéminences syllabiques, communément considérées comme le point d'ancrage pour la segmentation prosodique d'un énoncé (Buhmann et al. 2002, Tamburini & Caini 2005), et les disfluences. L'annotation a été effectuée sous PRAAT (Boersma & Weenink 2010) par cinq codeurs naïfs. Etant donné, le score médiocre du taux d'accord inter-annotateurs pour les proéminences[ix] (compris entre 0,23 et 0,52)[x], cette première phase d'annotation a été complètement revue par trois experts pour générer l'annotation de référence (kappa score sur quatre échantillons tests avant l'annotation : 0,78 à 0,81 pour les proéminences et 0,84 à 0,87 pour les disfluences)[xi]. En pratique une syllabe proéminente est une syllabe qui se détache perceptivement de son environnement phonétique (Terken 1991) ; une unité (mot, syntagme, etc.) pourvue d'une ou de

plusieurs syllabes proéminentes est une unité qui se dégage comme une figure sur un fond discursif (Arnold et al. 2012). La proéminence est donc un indice crucial que l'auditeur utilise pour traiter la prosodie on-line. Une échelle à trois degrés a été retenue pour le codage des proéminences (tableau 6) : syllabe non proéminente ('0'), syllabe modérément proéminente ('W'), syllabe fortement proéminente ('S')[xii].

Syl	se	tE	a	se	a	se	te	Ri	ble	le	le	mE	zo~	o~	bRy	le
Prom	0	0	W	S	0	W	0	S	0	0	0	0	0	0	0	S

Tableau 6. Etiquetage des proéminences dans l'énoncé *c'était assez assez terrible et les les maisons ont brûlé*, [Rhap-D0003, Corpus PFC] ; avec de haut en bas : la tire des syllabes ('Syl') et la tire des proéminences ('Prom')

Le concept de *disfluence* est habituellement utilisé pour désigner des points dans la chaîne parlée correspondant à une perturbation du programme syntagmatique (Blanche-Benveniste & Jeanjean 1986). Il s'agit d'une classe générique (tableau 7) qui regroupe les pauses remplies (ou pauses d'hésitation) précédées et/ou suivies de 'euh', les allongements syllabiques supérieurs au seuil d'allongement qui marque une syllabe accentuée, les répétitions, les faux départs et les inachèvements de morphèmes, de mots ou de syntagmes. Ces phénomènes apparaissent souvent de façon combinée dans le flux discursif et sont marqués, du moins nous en posons l'hypothèse, par des patrons temporels et/ou mélodiques spécifiques.

Syl	I	la	vE	de	_	de	pRO	ZEd	par	ti	R2	_	9
Disfl				H							H		H

Tableau 7. Etiquetage des disfluences (marqueur 'H') dans l'énoncé *Il avait des _ des projets de partir euh _ euh* [Rhap-D0003, Corpus PFC] ; avec de haut en bas : la tire des syllabes ('Syl') et la tire des disfluences ('Disfl).

4.2 Génération automatique de la structure prosodique

La structure prosodique, dérivée automatiquement de l'annotation manuelle, est une structure hiérarchique non récursive composée de quatre niveaux de constituance, avec du plus bas au plus haut de la hiérarchie : le pied métrique, le groupe rythmique, le paquet intonatif et la période intonative (figure 9). En pratique, la génération s'articule autour des étapes suivantes : un échantillon est segmenté en une succession de périodes intonatives ; chaque période ainsi constituée est segmentée en pieds métriques puis on remonte progressivement de bas en haut de la structure pour effectuer les regroupements en groupes rythmiques et en paquets intonatifs.

- **Génération de la période intonative, ou unité prosodique maximale (PEI) :** L'énoncé est segmenté automatiquement et vérifié manuellement avec le logiciel Analor (Lacheret et Victorri 2002, http://www.lattice.cnrs.fr/analor) en une succession de périodes intonatives (PI) sur la base de deux types de critères : (i) des critères acoustiques pour les fichiers monologaux (tableau 8) et des contraintes liées à la distribution des tours de parole pour les fichiers dialogaux.

Détection d'une pause d'au moins 300 ms
Détection d'un mouvement de F0 ample : trait [±ample] fixé en fonction de l'intervalle mélodique, mesuré en demi-tons, entre le dernier extremum de F0 (avant la pause) et la moyenne de F0 sur toute la portion qui précède la pause
Détection d'un saut mélodique (ou 'resetting'), i.e. intervalle mélodique qui sépare les points de F0 avant et après la pause

Tableau 8. Critères de coupures en périodes intonatives pour les échantillons monologaux

Il convient de souligner que la décision de reconnaissance d'une rupture périodique repose sur un principe de compensation de seuils. En d'autres termes, la détection ne dépend pas des valeurs exactes des paramètres, mais de leurs seuils respectifs d'activation et des poids associés (activation très forte : poids '2', forte : '1', moyenne : '0', en dessous du seuil : '-1) : quand un paramètre est légèrement au-dessous du seuil, une frontière de période est détectée si les autres paramètres ont des valeurs au-dessus du seuil (figure 6).

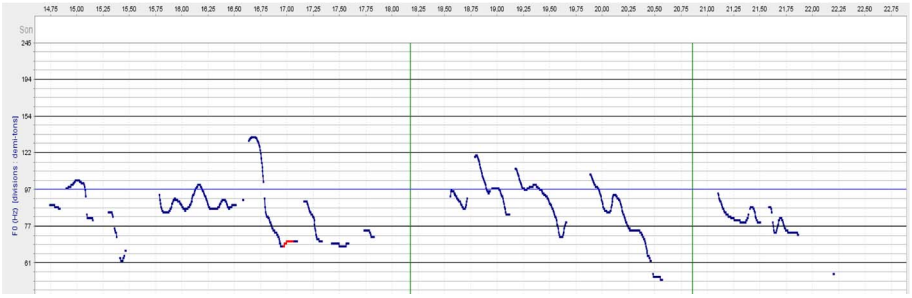

Figure 6: Segmentation en trois périodes de l'énoncé *je pense aux nombreuses victimes de la tempête*$_{\mathbf{PI}}$ *et à toute leur famille endeuillée*$_{\mathbf{PI}}$ *dont nous partageons la peine*$_{\mathbf{PI}}$, [D2004, Corpus Rhapsodie][xiii]. Les deux lignes verticales représentent les frontières des trois périodes. Avec pour la première période (les seuils 2, -1, 1 ; pour la seconde : 1, 1, 2 ; pour la dernière : 2, 1, 2 ; soit respectivement des seuils globaux de 2, 4 et 5.

Pour les fichiers dialogaux : (i) dans l'état actuel des choses, le premier indice pour la détection d'une période est la pause, or, deux tours peuvent s'enchaîner sans pause, (ii) une période intonative ne peut pas être attribuée simultanément à plusieurs locuteurs, il faut donc pouvoir faire des choix dans les contextes de chevauchement.

Dans les simples contextes d'enchaînement, chaque transition de tour correspond à une frontière de période, quelle que soient les configurations acoustiques en jeu. Les contextes de chevauchement sont beaucoup plus complexes à traiter et demandent des prises de décision manuelles. Trois cas de figure peuvent se présenter (figure 7), qu'il est important de distinguer dans la perspective de l'articulation entre l'annotation syntaxique, où les segments chevauchés ne sont pas gommés, et l'annotation prosodique, où on ne peut pas manipuler plusieurs signaux superposés. (i) Une période est tronquée à sa finale : la fin du tour de parole (syllabe, morphème, mot ou suite de mots) est masquée par le début du tour suivant et, en conséquence, ne peut pas être alignée au signal. (ii) Une période est tronquée à l'initiale : le début d'un tour est maqué par la fin du tour précédant et ne peut donc pas être aligné au signal. (iii) Un segment correspond à la superposition complète de deux tours de parole, en conséquence, n'est consigné dans la période que celui qui peut faire l'objet d'un alignement phonétique.

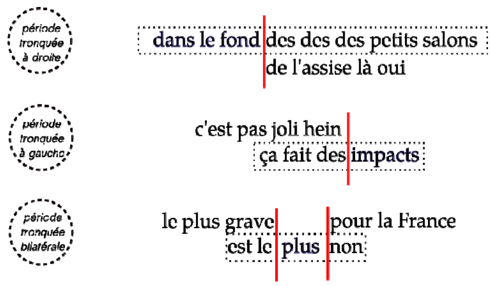

Figure 7. Illustration des périodes tronquées, avec de gauche à droite : une période tronquée à sa finale [Rhap-D0009, Corpus PFC], une période tronquée à l'initiale [Rhap-D0009, Corpus PFC], une superposition complète [Rhap-D0001, Corpus CFPP].

- **Génération.du pied métrique** : Toute syllabe proéminente non disfluente d'une période marque la frontière droite et la tête d'un pied métrique (PM).
- **Génération.du groupe rythmique** : Chaque tête de pied métrique qui s'aligne avec la frontière droite d'un mot orthographique marque la fin d'un groupe rythmique (GR).
- **Génération.du paquet intonatif** : Pour plusieurs GR contigus, le premier ponctué par une proéminence terminale forte ('S') indique la frontière droite d'un paquet intonatif (PAI).

Ainsi, pour l'annotation manuelle ci-dessous (figure 8) de la séquence *que vous soyez devenue une vedette vous étiez normalement entraînée*,

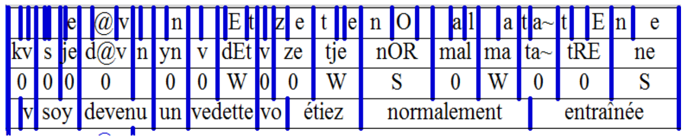

Figure 8. Annotation manuelle des proéminences ([Rhap-D2001, Corpus Mertens]

La structure prosodique générée est fournie dans la figure 9 :

PEI	que vous soyez devenue une vedette vous étiez normalement entrainée																
PAI	que vous soyez devenue une vedette vous étiez normalement entrainée																
GR	que vous soyez devenue					une vedette			vous étiez			normalement			entrainée		
PM	kvuswajed↔vny					ynv↔dEt			vuzetje			n□r	malmã		ãtrene		
Syl	kvu	swa	je	d↔	ny	yn	v	d	vu	ze	tje	n□r	mal	m	ã	tre	n

Figure 9. Exemple de génération automatique de la structure prosodique à partir de l'annotation manuelle

4.3 Annotation tonale

Une unité, quelle que soit sa nature (prosodique ou syntaxique) et son empan, est porteuse d'un contour mélodique global que l'on peut représenter symboliquement par une suite de tons mélodiques. La première représentation de l'intonation du français en termes de contours globaux avait comme objectif de caractériser l'intonation modale en français (Delattre 1966) : à une modalité donnée (affirmation, question, exclamation, etc.) serait associé un contour particulier. Plus récemment, ce paradigme a été modélisé par Aubergé (1991), qui, dans le cadre de la génération automatique de la prosodie pour la synthèse de la parole, s'est donné comme tâche de représenter l'intonation comme une superposition de contours mélodiques élémentaires portés par des unités syntaxiques variables. Dans la continuité de ces travaux, nous posons l'hypothèse que pour toutes les unités dérivées des annotations Rhapsodie, les unités

syntaxiques, tout comme les unités prosodiques, nous pouvons construire un lexique de contours prototypiques. Le traitement ici conduit s'articule autour de cinq points : (i) la transcription intonative (l'alphabet utilisé pour décrire les contours) résulte d'un traitement bottom-up, fondé sur les caractéristiques de la courbes mélodiques, (ii) une représentation temps-fréquence est utilisée pour décrire les contours complexes avec un jeu limité de symboles, (iii) la représentation n'est pas restreinte à un type spécifique de segment mais peut être généralisée pour n'importe quel type d'unité, (iv) cette représentation tient compte du registre mélodique global d'un locuteur, (v) elle est suffisamment robuste pour être utilisée en parole interactive, y compris dans des contextes de chevauchement de parole.

Dans une première étape, la représentation acoustique est fondée sur une stylisation préalable de la courbe mélodique, dont la fonction est de gommer les variations micro-temporelles non pertinentes pour sa description, parce que non perçues par l'auditeur. Cette représentation repose sur un jeu de cinq valeurs acoustiques qui renseignent sur les points de F0 initial et final du contour, la saillance principale, la position de cette saillance dans le contour (initiale, médiane ou finale) pour les contours complexes[xiv] et le registre local associé au contour ; cinq nivaux de hauteur sont considérés, de l'infra-bas au suraigu (tableau 9). La représentation est ensuite utilisée pour l'annotation d'un contour mélodique en une succession de tons élémentaires (figure 10). Formellement, l'annotation est composée de quatre champs :

Initial-Final-[Saillance mélodique]-[position temporelle de la saillance]

Où les deux derniers champs ne sont indiqués que pour les contours complexes.

Etiquettes pour les tons élémentaires	Niveau de hauteur	Valeur en demi-tons (par rapport au registre global du locuteur)
H	Suraigu	>+6
H	Haut	+2/+6
M	Moyen	-2+2
L	Bas	-2/-6
L	Infra-bas	<-6

Tableau 9. Les tons élémentaires et leurs niveaux de hauteur (en demi-tons)

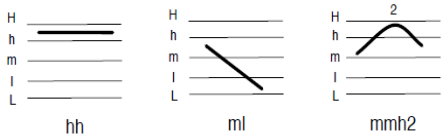

Figure 10. Exemples d'annotation. Avec de gauche à droite : un plateau mélodique dans le niveau aigu, un contour simple descendant du niveau moyen vers le niveau grave et un contour complexe (niveau moyen, aigu et moyen). Pour les deux premiers contours, aucune saillance interne n'est observée, les deux derniers champs sont donc vides ; en revanche, pour le dernier contour, une saillance au niveau aigu (marqueur 'H') est repérée ; cette saillance se situe au 2/3 du contour (marqueur '2').

Une illustration du traitement est fournie dans la figure 11 pour le calcul des contours des syllabes, des paquets intonatifs et des périodes.

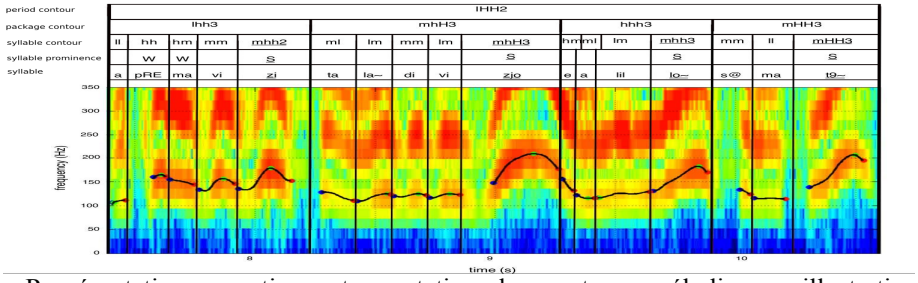

Figure 11. Représentation acoustique et annotation des contours mélodiques ; illustration pour les syllabes, les paquets intonatifs et la période *Après ma visite à Landivisiau et à l'île Longue ce matin*

[Rhap-M001, C-Prom]. Les cercles bleus et rouges indiquent les valeurs initiales et finales, les cercles verts : les saillances intermédiaires.

5 Conclusion

Le développement de schémas d'annotation pour l'établissement du treebank Rhapsodie a été motivé par l'objectif prioritaire de sonder l'interface prosodie/syntaxe et ses variations à travers divers genres de discours. La contribution spécifique du projet Rhapsodie au sein de la linguistique de corpus se résume en six points : (i) une ressource accessible en ligne (données primaires, donnés secondaires, tutoriels d'annotation pour les différents niveaux annotés, outils d'annotation, d'analyse et/ou de visualisation qui ont fait l'objet de développements dans le cadre du projet), (ii) la définition de schémas d'annotation modulaires de la syntaxe et de la prosodie du français parlé, explicites et reproductibles, (iii) une approche corpus-driven, bottom-up, où les unités ne sont pas posées a priori mais émergent empiriquement de l'observation des données et se construisent dynamiquement, (iv) des solutions innovantes pour traiter les chevauchements de parole, les disfluences et les empilements paradigmatiques, et, par conséquent, donner à ces éléments la place qu'ils méritent dans l'analyse linguistique, (vi) l'implémentation d'un modèle pour l'annotation tonale automatique d'unités de nature et d'empan variable, (v) une approche flexible caractérisée par une annotation suffisamment neutre pour permettre à des cadres théoriques pluriels de s'en saisir.

Références bibliographiques

'THart, J., Collier, R. and Cohen, A. (1990). *Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

Abeillé, A. & Crabbé, B. (2013). Vers un treebank du français parlé, *actes TALN 2013*, 174-187.

Arnold, D. & Wagner, P. (2008). The influence of top-down expectations on the perception of syllable prominence. *Proceedings of the second ISCA Workshop on Experimental Linguistics*, 25-28.

Arnold, D., Wagner, P., Möbius, B. (2012). Obtaining prominence judgments from naïve listeners. *Proceedings of Interspeech 2012*, http://interspeech2012.org/.

Aubergé, V. (1991). *La synthèse de la parole: des règles aux lexiques*. Thèse de Doctorat, Grenoble : Université Pierre Mendès-France.

Avanzi, M., Simon, A. C., Goldman, J.-P., Auchlin, A. (2010). C-PROM. An annotated corpus for French prominence studies. *Proceedings of Prosodic Prominence: Perceptual and Automatic Identification, Speech Prosody 2010*, http://speechprosody2010.illinois.edu/.

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 CMLF 2010*, 2075-2090.

Berrendonner, A. (1990). Pour une macro-syntaxe. *Travaux de linguistique* 21, 25-31.

Biber, D. & Conrad, S. (2009). *Register, Genre and Style*. Cambridge : Cambridge University Press.

Blanche-Benveniste, Cl, & Jeanjean, C. (1986). *Le français parlé. Editions et transcription*. Paris: Didier Erudition.

Blanche-Benveniste, Cl., Bilger, M., Rouget, Ch., Van den Eyende, K. (1990). *Le français parlé. Etudes grammaticales*. Paris: Editions du CNRS .

Boersma, P. & Weenink, D., (2010). Praat: doing phonetics by computer (Version 5.3). www.praat.org.

Branca, S., Fleury, S., Lefeuvre, Fl., Pires, M. (2012). Discours sur la ville. Corpus de Français Parlé Parisien des années 2000 (CFPP2000). http://cfpp2000.univ-paris3.fr/.

Broeder, D., Van Uytvanck, D., Gavrilidou, M., Trippel, T., Windhouwer, M. (2012). Standardizing a component metadata infrastructure. *Proceedings of LREC*, 1387-1390.

Buhmann, J., Caspers, J., van Heuven, V., Hoekstra, H., Martens, J.-P., Swerts, M. (2002). Annotation of Prominent Words, Prosodic Boundaries and Segmental Lengthening by Non Expert Transcribers in the Spoken Dutch Corpus. *Proceedings of LREC*, 779-785.

Cerisara, C., Gardent, C., Anderson, C. 2010. Building and Exploiting a Dependency Treebank for French Radio Broadcast. *Proc. 9th international workshop on Treebanks and Linguistic Theories* (*TLT9*), Tartu : Estonie.

Cheng, W., Greaves, C., Warren, M. 2008. *Discourse Intonation systems. A Corpus-driven Study of Discourse Intonation The Hong Kong Corpus of Spoken English* (Prosodic). Benjamins.

Cresti, E. (2000). *Corpus di italiano parlato*. Florence: Accademia della Crusca.

Cresti, E. & Moneglia M. (eds.) 2005. *C-ORAL-ROM. Integrated reference corpora for spoken romance languages*, DVD + vol., Benjamins.

Delattre, P. (1966). Les dix intonations de base du français. *The French Review*, 40(1):1-14.

Deulofeu, J., Gerdes, K., Kahane S., Pietrandrea, P. (2010). Depends on what the French say: Spoken corpus annotation with and beyond syntactic function. *Proceedings of LAW IV, ACL*, 274-281.

Durand, J., Laks, B., Lyche, C. (2009). Le projet PFC (phonologie du français contemporain): une source de données primaires structurées, In J. Durand, B. Laks & C. Lyche (eds.), *Phonologie, variation et accents du français*. Paris : Hermès, 19-61.

Eshkol-Taravella, I., Baude, O., Maurel, D., Hriba, L., Dugua, C., Tellier, I., (2011). Un grand corpus oral disponible : le corpus d'Orléans 1968-2012. *TAL, 52 – n° 3,* 17-46.

Fleiss, Joseph L., Cohen, Jacob. 1971. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Measures of Reliability. *Educational and Psychological Measurement*, 33, 613-619.

Gerdes, K. (2013). Collaborative Dependency Annotation. *DepLing 2013*, 88.

Goldman, J.-Ph. (2011). Easyalign: an automatic phonetic alignment tool under praat. *Proceedings INTERSPEECH*, 3233-3236. http://latlntic.unige.ch/phonetique.

Hoekstra, H., Moortgat, M., Renmans, B., Schouppe M., Schuurman I., Van der Wouden, T. (2003). CGN Syntactische Annotatie. http://lands.let.kun.nl/cgn/doc_Dutch/topics/version_1.0/annot/syntax/syn_prot.pdf

Kahane, S.& Pietrandrea, P. (2012). La typologie des entassements en français. *Actes CMLF*, 1809-1828.

Koch, P.& Oesterreicher, W. (2001). Langage parlé et langage écrit. *Lexicon der Romanistischen Linguistik*, T1-2, 584-627.

Lacheret A. & Beaugendre F. (1999). *La prosodie du français*. Paris, Editions du CNRS.

Lacheret, A., Victorri, B. (2002). La période intonative comme unité d'analyse pour l'étude du français parlé : modélisation prosodique et enjeux linguistiques. *Verbum*, 24/1-2, 55- 73.

Lacheret A., Kahane S. , Pietrandrea P., Avanzi M., Victorri B. (2011). Oui mais elle est où la coupure, là? Quand syntaxe et prosodie s'entraident ou se complètent. *Langue Française, 170*, 61-80.

Lacheret A., Simon, A-C, Goldman, J-Ph, Avanzi, M. (2013). Prominence perception and accent detection in French: from phonetic processing to grammatical analysis. *Language Sciences*, 39, 95-106.

Lacheret, A., Kahane, S., Pietrandrea, P. (2015, à paraître). Rhapsodie. A Prosodic-Syntactic Treebank for Spoken French. Amsterdam-Philadelphia: John Benjamins Publishing Company.

Mel'čuk, I. (1988). *Dependency Syntax : Theory and Practice*, SUNY Press.

Meteer, M. (1995). Disfluency annotation stylebook for the switchboard corpus. Rapport technique, Upenn.

Morel, M.-A., Danon-Boileau, L. (1998). *Grammaire de l'intonation*, Gap-Paris : Ophrys.

Nelson, G., Wallis S., Aarts B. (eds.) 2002. Exploring Natural Language: Working with the British Component of the International Corpus of English, John Benjamins Publishing Company, Varieties of English Around the World G29.

Nilsson, J., Riedel S., Deniz Yuret, D. (2007). The CoNLL 2007 shared task on dependency parsing. *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*, 915-932.

Obin, N. Lacheret, A. Beliao, J. (2014). Tonal annotation: stylization of complex melodic contours over arbitrary linguistic units, Speech Prosody, Dublin, http://www.speechprosody2014.org/.

Schuurman, I., Goedertier, W., Hoekstra, H., N. Oostdijk, R. Piepenbrock, Schouppe, M. 2004. Linguistic annotation of the Spoken Dutch Corpus: If we had to do it all over again ..., Proc. LREC, Lisbon, 57-60.

Tamburini, F. & Caini, C. (2005). An Automatic System for Detecting Prosodic Prominence in American English Continuous Speech. *International Journal of Speech Technology*, 8, 33-44.

Terken, J. (1991). Fundamental Frequency and Perceived Prominence. *Journal of the Acoustical Society of America*, 89, 1768-1776.

Tesnière, L. (1959). *Eléments de syntaxe structurale*, Klincksieck.

Villemonte de la Clergerie, E. (2010). Building factorized TAGs with meta-grammars. *Proceedings of the 10th International Conference on Tree Adjoining Grammars and Related Formalisms - TAG+10*, 111-118.

Wightman, Colin W. (2002). Tobi Or Not Tobi? *Proceedings of Speech Prosody*, http://sprosig.isle.illinois.edu/sp2002/pdf/wightman.pdf.

---

[i] Pour un panorama, voir http://en.wikipedia.org/wiki/Treebank#Syntactic_treebanks.

[ii] Voir aussi le concept de « registre » chez Biber et Conrad (2009), caractérisé par une série de traits lexicaux-grammaticaux fréquents et récurrents dans les textes d'une certaine variété et qui servent des fonctions communicatives majeures.

[iii] La description des sources est accessible sur le site Rhapsodie http://www.projet-rhapsodie.fr/propriete-intellectuelle.html .

[iv] Un genre de discours peut être décrit selon la nature de la situation de communication (localisation, but communicationnel, degré de formalité), mais aussi selon l'environnement spatial et le canal de communication, ou encore en fonction du contenu thématique, etc.

[v] Nous appelons greffe une UI qui vient occuper une position régie où est attendu, en principe, un nom ou un groupe nominal : « *cette fille, c'est une pousse toi de là que je m'y mette* ». Les discours rapportés sont également des UI qui viennent occuper une position régie, mais, à la différence des greffes, il s'agit d'une position où est attendue une telle construction.

[vi] Les disfluences sont bien annotées dans la plupart des treebanks syntaxiques de l'oral, mais comme des segments à « éliminer » de l'analyse syntaxique (Abeillé & Crabbé 2013, Hoekstra et al. 2003).

[vii] Ni ton lexical, ni accent de mot mais un accent primaire final de groupe, doté le cas échéant d'accents secondaires internes (Lacheret et Beaugendre 1999).

[viii] Autrement dit, l'interface forme-fonction n'intervient pas au niveau de l'annotation, ce n'est que dans l'étape ultérieure d'analyse des annotations que cette question est traitée. Il s'agit ici de rendre compte uniquement de la dimension formelle.

[ix] En revanche, un bon score a été observé pour l'annotation des disfluences (compris entre 0,78 et 0,81).

[x] Test Fleiss-Kappa test (Fleiss & Cohen 1971).

[xi] La perception des proéminences et des disfluences est un phénomène complexe qui comporte une part de subjectivité certaine. Ainsi pour les proéminences, si l'auditeur s'appuie sur des indices acoustiques pour leur identification, il ne peut y avoir de perception brute indépendamment des contraintes grammaticales qui déterminent en partie les attentes des auditeurs (Arnold et Wagner 2008, Lacheret et al. 2013). Une campagne inter-annotateurs supervisée par un contrôle d'experts s'avère donc indispensable.

[xii] Où les marqueurs 'W' et 'S' correspondent aux valeurs *weak* et *strong.*

[xiii] Discours oratoire : vœux présidentiels pour la bonne année.

[xiv] L'information sur la saillance n'est fournie que si celle-ci ne correspond pas au premier ou au dernier point fréquentiel du contour.