Rhapsodie: a Prosodic-Syntactic Treebank for Spoken French

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, Université Paris Ouest Nanterre & CNRS

(2) Université Saint-Louis - Bruxelles

(3) LPP, Université Paris Sorbonne Nouvelle & CNRS

(4) Université de Genève

(5) IRCAM, UMR STMS IRCAM-CNRS-UPMC, Paris

(6) LLL, Université François Rabelais & CNRS, Tours-Orléans

anne@lacheret.com, sylvain@kahane.fr, julie@beliao.fr, dister@fusl.ac.be, kim@gerdes.fr, jean-philippe.goldman@unige.ch, nicolas.obin@ircam.fr, paolapietrandrea@gmail.com

Abstract

The main objective of the *Rhapsodie* project (ANR Rhapsodie 07 Corp-030-01) was to define rich, explicit, and reproducible schemes for the annotation of prosody and syntax in different genres (± spontaneous, ± planned, face-to-face interviews vs. broadcast, etc.), in order to study the prosody/syntax/discourse interface in spoken French, and their roles in the segmentation of speech into discourse units (Lacheret, Kahane, & Pietrandrea forthcoming).

We here describe the deliverable, a syntactic and prosodic treebank of spoken French, composed of 57 short samples of spoken French (5 minutes long on average, amounting to 3 hours of speech and 33000 words), orthographically and phonetically transcribed. The transcriptions and the annotations are all aligned on the speech signal: phonemes, syllables, words, speakers, overlaps.

This resource is freely available at www.projet-rhapsodie.fr. The sound samples (wav/mp3), the acoustic analysis (original F0 curve manually corrected and automatic stylized F0, pitch format), the orthographic transcriptions (txt), the microsyntactic annotations (tabular format), the macrosyntactic annotations (txt, tabular format), the prosodic annotations (xml, textgrid, tabular format), and the metadata (xml and html) can be freely downloaded under the terms of the Creative Commons licence Attribution - Noncommercial - Share Alike 3.0 France. The metadata are encoded in the IMDI-CMFI format and can be parsed on line.

Keywords: spoken French Treebank, prosodic annotation, syntactic annotation

1. Introduction

One of the fundamental questions underlying the linguistic analysis of spoken languages is their decomposition into discourse units that can be considered as basic in terms of informational processing and communication. It is well known that, in many languages, prosody and syntax play a crucial role in the identification of these units. However, although widely studied for decades, the relation between these two levels has not been thoroughly explored and a number of general theoretical questions are still unanswered: To what extent do prosodic and syntactic structures interact? To what extent are they autonomous from one another in creating discourse units? Is discourse cohesion always guaranteed by syntax or can we say that prosody supplies cohesion when syntax is absent? Clearly, answering these questions would amount to a precise description of the role that prosody and syntax play in segmenting discourse into pragmatic and textual units.

In order to approach these questions, we first annotated and then analyzed at both the prosodic and the syntactic level a corpus of spoken French. French is a language that presents a particularly interesting interplay between prosody and syntax in discourse structuring. This is due in the first place to the massive presence of so-called paratactic phenomena (Blanche-Benveniste et al. 1990, Béguelin et al. 2010) and in the second place to the fact that supralexical rather than lexical phenomena are relevant for French prosodic organization (Rossi 1979, Lacheret & Beaugendre 1999). For the annotation task we adopted an approach that can be characterized as empirical, inductive and modular: "empirical" because we annotated the entirety of the data in the corpus, without neglecting any segment whatsoever; "inductive" because the set of relevant units for our corpus was identified through a data-driven incremental strategy of annotation; "modular" because we independently annotated prosodic and syntactic units.

2. Corpus design

Given the modelling objectives of our project, we privileged for our corpus the representation of a great variety of textual typologies and of a great number of speakers rather than a balanced sociolinguistic representation. We therefore collected recordings of 89 Central French adult native speakers from early eighties to nowadays. In this section, we present the composition of the *Rhapsodie* Treebank: (i) the innovative strategy chosen to build the Rhapsodie database, (ii) specific legal issues associated with our approach, (iii) the tool used to encode the metadata, and (iii) the discourse features selected to characterize Rhapsodie samples.

2.1. Issues: samples and metadata

The corpus design focused on the selection of samples with a sufficient variety in terms of textual typology. To do so, the *Rhapsodie* repository could not rely on any representative corpus of spoken French, since none exists. Our contribution to this issue has consisted in the elaboration of a rather innovative sampling strategy.

Firstly, the corpus samples have been mainly derived from existing corpora of spoken French (among others, PFC: Durand et al. 2009, C-Prom: Avanzi et al. 2010, CFPP2000: Branca et al. 2012) and partially created within the framework of the *Rhapsodie* project. Secondly, we had to define a procedure to acknowledge the intellectual property of the creators of the source corpora, as well as strategies to refer to source corpora and to ensure the possibility of retrieving the original samples. Lastly, we had to choose a metadata standard which provides an exhaustive textual description of each sample, in order to provide complete information about source corpora and to precisely describe the annotations of each sample, which are at the core of the *Rhapsodie* project. For this last point, we chose to encode our metadata in the IMDI-CMDI format developed at the Max Planck Institute for Psycholinguistics in Nijmegen (CMDI, http://www.clarin.eu/cmdi, Broeder et al. 2012).

2.2. Maximizing the diversity of Discourse Genres

The description of discourse genres involves a large number of socio-communicative variables that are independent of one another (Koch & Oesterreicher 2001, Biber & Conrad 2009). Since representing the complete variability of discourse genres is totally unrealistic, the objective of the *Rhapsodie* project was to maximize the diversity of the discourse genres by including a number of speech samples for which rich syntactic/prosodic annotations could be manually processed. The selection of speech samples was therefore derived from general principles that are commonly used for the description of discourse genres. The first principle was to balance the distribution within the corpus between public and private speech, then each type of speech is made of monologues and dialogues. Second, the following variables were used to represent discourse features of each sample (Table 1): (i) the degree of speech planning, (ii) the degree of interactivity, (iii) the channel of communication, and (iv) the type of discourse sequence mostly characterizing the speech (from argumentation to neutral description).

Type of speech	**Private, public**	**monologues**
		dialogues
	Planning type (planned, semi-spontaneous, spontaneous)	
	Interactivity (non interactive, semi-interactive, interactive)	
	Channel (broadcasting, face-to-face)	
	Discourse sequence (oratory, argumentation, description, procedural)	

Table 1. Discourse features taken into account in Rhapsodie Corpus

3. Annotation schemes

The first step in processing the *Rhapsodie* corpus was to produce manual orthographical transcriptions (Dister & Simon 2008) and speech/text alignment (phonemes, syllables, words, pauses), performed automatically with EasyAlign (Goldman 2011), then manually corrected, and on which annotations were conducted. The remainder of the paper describes the schemes used for syntactic and prosodic annotation.

3.1. Syntactic annotation

Combining the syntactic model proposed by the Aix School (Blanche-Benveniste et al. 1990) and the pragmatic model developed within the Lablita project (Cresti 2000), we annotated two levels of syntactic cohesion: microsyntax, i.e., syntactic cohesion guaranteed by government and macrosyntax, i.e. syntactic cohesion guaranteed by illocutionary dependency.

The macrosyntactic level describes the whole set of relations holding between all the segments that make up one and only one illocutionary act. The annotation was conducted manually by the syntactician team of the *Rhapsodie* project on distributional syntactic properties (Deulofeu et al. 2010). Basically, each sample is segmented into a string of illocutionary units (henceforth IU); each IU is composed of 3 kinds of components: a nucleus (obligatory), pre-nuclei (optional) and post nuclei (optional); see below: (1) and (2), where '<' follows a pre-nucleus and precedes a nucleus or another pre-nucleus; '>' precedes a post-nucleus and follows a nucleus or a previous post-nucleus; and '//' indicates the right boundary of a IU (nuclei are in bold).

(1) alors < là < la psychiatrie < **c'est autre chose** // [Rhap-D0006, CFPP2000]
well < now < psychiatry < that's something else //

(2) ***ça a duré dix ans*** *> le silence autour de moi //* [Rhap-D2001, Mertens corpus]
it lasted two years > the silence around me //

We also propose a complete annotation and a functional tagging of pile structures (Kahane & Pietrandrea 2012). By piles we mean the multiple realization of one and the same structural position, which occurs in continuous speech in various types of segments, especially syntactic disfluencies (see 3 in bold).

(3) *alors < { { **j'a~ | j'avais } beaucoup | j'avais beaucoup** } trop peur de m'installer (comme ça) seule { d~ | dans } la brousse //* [Rhap-D2004, Rhapsodie]
'well < { { I wa~ | I was } much | I was much } too scared of moving (like that) alone { i~ | into } the outback //'

Albeit extremely frequent in spoken language, this cohesion device, which can be regarded as a particular type of microsyntactic relation, is often disregarded in corpus annotation. By extensively annotating and tagging pile phenomena we were able to guarantee an exhaustive

microsyntactic annotation of all our data, including disfluencies, repetitions, and reformulations generally considered as performance errors and not analyzed in spoken language treebanks.

The microsyntactic structure is encoded as a dependency tree. Note that we do not consider IU or turn-taking as boundaries of microsyntactic dependencies. In the following exchange over two turns (figure 1), the question of the second speaker is analyzed as an adjunct to the nucleus of the assertion of the first speaker.

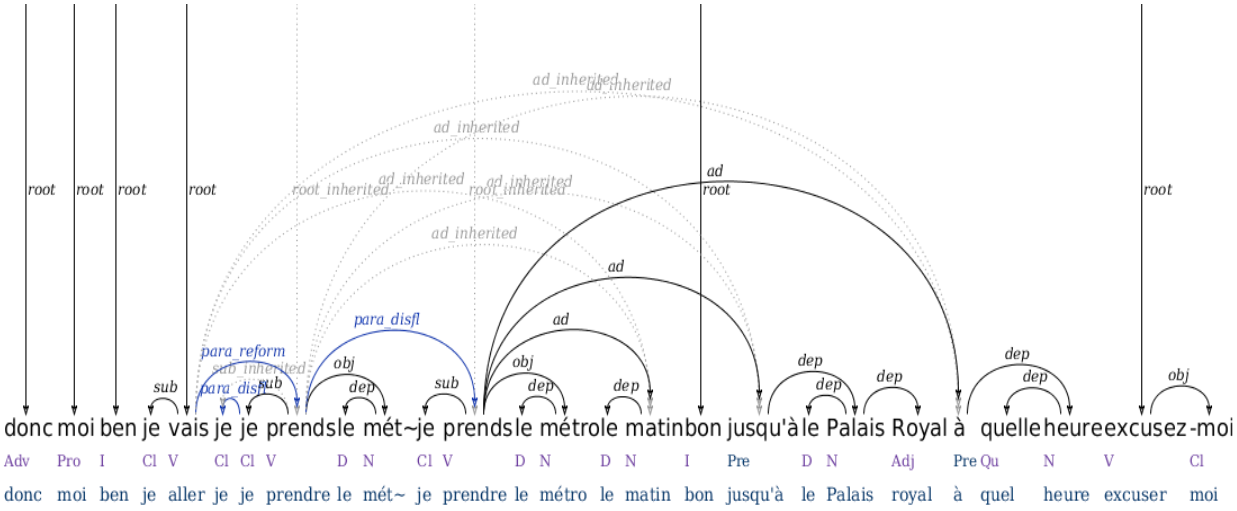

Figure 1. Dependency tree of the two turns: *$L1 donc < moi < "ben" { je vais | { je | je } prends le mét~ | je prends } le métro le matin "bon" jusqu' au Palais Royal //+ $L2 à quelle heure "excusez-moi" //* [Rhap-D0001, CFPP2000 corpus]
'$L1 so < me < "well" { I go | { I | I } take the met~ | I take the metro } in the morning "well" until Palais-Royal //+ $L2 at what time "excuse me" //'

3.2. Prosodic annotation

The model used for the prosodic annotation is based on the theoretical hypothesis formulated by the Dutch-IPO school ('t Hart et al. 1990) stating that, out of the total information characterizing the acoustic domain, only some perceptual cues selected by the listener are relevant for linguistic communication (see also Wightman 2002). From this starting point, the prosodic annotation was processed into three parts: 1) the manual annotation of relevant perceptual prosodic events, 2) the automatic derivation of the prosodic structure based on this manual annotation, and 3) the automatic stylization of melodic contours and the tonal annotation associated with the constituents contained in the prosodic structure.

Two types of event were retained for the manual annotation: prosodic prominences - that are widely considered as the core prosodic event for the annotation of speech prosody (Buhmann et al. 2002; Tamburini & Caini 2005) - and disfluencies. Prosodic boundaries were not considered, due to the poor inter-annotator agreement that was obtained during preliminary experiments (Lacheret et al. 2010).

As for the annotation of prominence (Table 2), we chose a three-level scale: a syllable can be strongly prominent (label 'S'), weakly prominent, (label 'W') or not prominent ('0').

	C'était assez assez terrible							
S	se	tɛ	a	se	a	se	te	ribl
P			W			W		S

Table 2. Annotation of prominences for the speech sequence *c'était assez assez terrible* (it was quite quite horrible), [Rhap-D0003, PFC corpus].

Regarding disfluencies, it can be seen as a generic label to designate numerous phenomena, which are traditionally named *filled pauses*, *fillers* (*euh*, which corresponds to English 'um's or 'er's or syllabic extra-lengthening), *repetitions*, *self-repairs*, *false starts*, and *truncations* (of morphemes, words or syntagms). These phenomena often appear together in the speech flow. In Rhapsodie, only disfluencies which are perceptually linked to specific prosodic profiles such as extra-lengthening, infra-low register and creaky voice are labeled at the prosodic level (Table 3).

Orthographic string	eh bien euh		
Syllabic string	e	bjɛ̃	∅
Disfluency	B	I	I

Table 3. Example of extra-lengthening followed by an 'um' in the sequence *eh bien euh* 'well um', [Rhap-D0003, PFC corpus]; where B and I indicate syllables at the beginning or inside a disfluent segment.

The prosodic structure automatically derived from this manual annotation is presented in Table 4. This structure is organized around rhythmic and melodic components. From the largest to the smallest constituent, these are: (i) global macroprosodic units called *intonational periods* (Lacheret & Victorri 2002), (ii) intonational units internal to periods called *intonational packages*, (iii) *rhythmic groups* internal to intonational packages and (iv) *metrical feet* inside rhythmic groups.

For the segmentation of intonational periods (henceforth IPE), only melodic variations in time and silent pauses are used, regardless of any segmental and syntactic constraints. Each pause of at least 300 ms is assigned a temporary marking and becomes a potential candidate for an IPE boundary. In other words, a silent is a necessary but not a sufficient marker to locate a potential IPE boundary, the localization of a boundary can be envisaged only with respect to the combination of several parameters. In practice, two other criteria are also used: (i) the detection of a F0 pitch movement reaching a certain amplitude; defined according to the melodic interval, measured in semitones, between the last extreme F0 value (before the silent pause) and the average F0 over the whole segment preceding the pause; (ii) the detection of a melodic jump which corresponds to the melodic interval which separates the points of F0 before and after the pause (melodic resetting); and (iii) absence of 'um' in the immediate vicinity of the pause (Figure 2).

Figure 2. Segmentation in three intonational periods for the speech *sequence je suis heureux de me retrouver ce soir # parmi vous # après ma visite à Landivisiau et à l'île Longue ce matin* 'I feel very pleased to be with you this evening after my visit to Landivisiau and L'île Longue this morning' [Rhap-M2001, C-Prom corpus].

Then, from the bottom to the top, the internal units of a IPE are generated as follow (Table 4):

- Metrical foot (MF): Each non-disfluent prominent syllable inside a syllabic string marks the end and the right head of a metrical foot (henceforth RHF).
- Each RHF that is the terminal syllable of a phonetic word marks the right boundary of a rhythmic group (RG).
- When there are several contiguous rhythmic groups, the first one that ends with a strong prominence forms an intonation package (IPA) with the preceding ones.

IPE	que vous soyez devenue une vedette vous étiez normalement entrainée																	–
IPA	que vous soyez devenue une vedette vous étiez normalement entrainée																	
RG	que vous soyez devenue					une vedette			vous étiez			normalement			entrainée			
MF	kvuswajedəvny					ynvədɛt			vuzetje			nɔr	malmã		ãtrene			
syllable	kvu	swa	je	dəv	ny	yn	və	dɛt	vu	ze	tje	nɔr	mal	mã	ã	tre	ne	
Prom	0	0	0	0	W	0	0	W	0	0	W	S	0	0	0	0	S	

Table 4. Prosodic tree derived from manual tagging. Segmentation of the period *que vous soyez devenue une vedette vous étiez normalement entraînée* 'the fact that you became a star you were normally trained', [Rhap-D2001, Mertens corpus]. From top to bottom: the *period*, the *intonation packages*, the *rhythmic groups*, the *metrical feet*, the *syllables* and the syllabic prominences

Finally, stylized melodic contours and tonal annotation were automatically computed for each constituent of the Rhapsodie Treebank (Obin et al. 2014). In the proposed method, the F0 contour is represented by a set of five acoustic values for each given unit: (i) the initial value of the F0 on the unit, (ii) the final value of the F0 on the unit, (iii) the main saliency, i.e. the value corresponding to the most salient F0 peak – if one exists, (iv) the main saliency position, i.e. the time position of the main saliency, relative to the boundaries of the unit, and (v) the local register which corresponds to the mean F0 over the unit. All frequency values are expressed in semi-tones, with respect to the overall mean F0 of the speaker. Frequency values are represented with respect to 5 pitch levels covering the whole F0 range of the speaker: H (extreme high), h (high), m (medium), l (low), and L (extreme low). Each pitch level covers a range of 4 semi-tones centered on the average F0 value of the speaker. Figure 3 illustrates the output for some prosodic constituents.

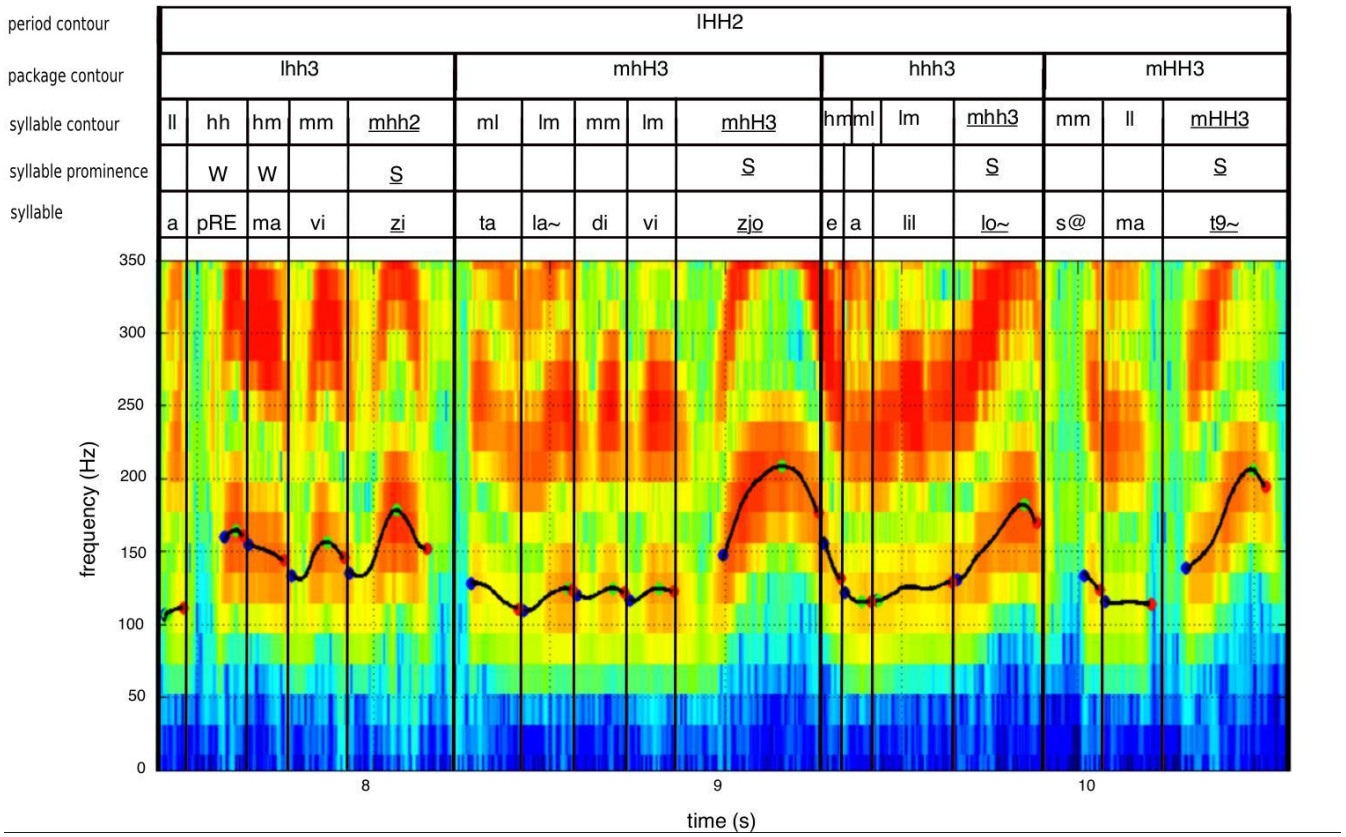

Figure 3. Tonal representation and annotation for the speech sequence *après ma visite à Landivisiau et à l'île Longue ce matin* 'after my visit to Landivisiau and L'île Longue this morning' [Rhap-M2001, C-Prom corpus]. On top: tonal annotation of melodic contours over syllables, intonation packages, and periods. Below: stylization of melodic contours. Blue and red dots denote initial and final melodic values, respectively, and green dots intermediate melodic saliencies.

4. Conclusion

We presented the Rhapsodie resource freely available at www.projet-rhapsodie.fr. The different steps of treatment, are summarized in Figure 4.

The development of prosodic and syntactic annotation schemes for French speech was guided by the objective of modeling the interface between prosody and syntax in discourse segmentation and structuring. The main contribution is to propose novel annotation schemes based on bottom-up principles, simultaneously as neutral as possible at the theoretical level and guided by the principles developed by the Dutch-IPO school for prosody, dependency grammars and macrosyntactic theory for syntax. The main advantage of the proposed annotation scheme is that it can be widely shared and used by the syntactic/prosodic community and can then be adapted to different linguistic approaches and representations.

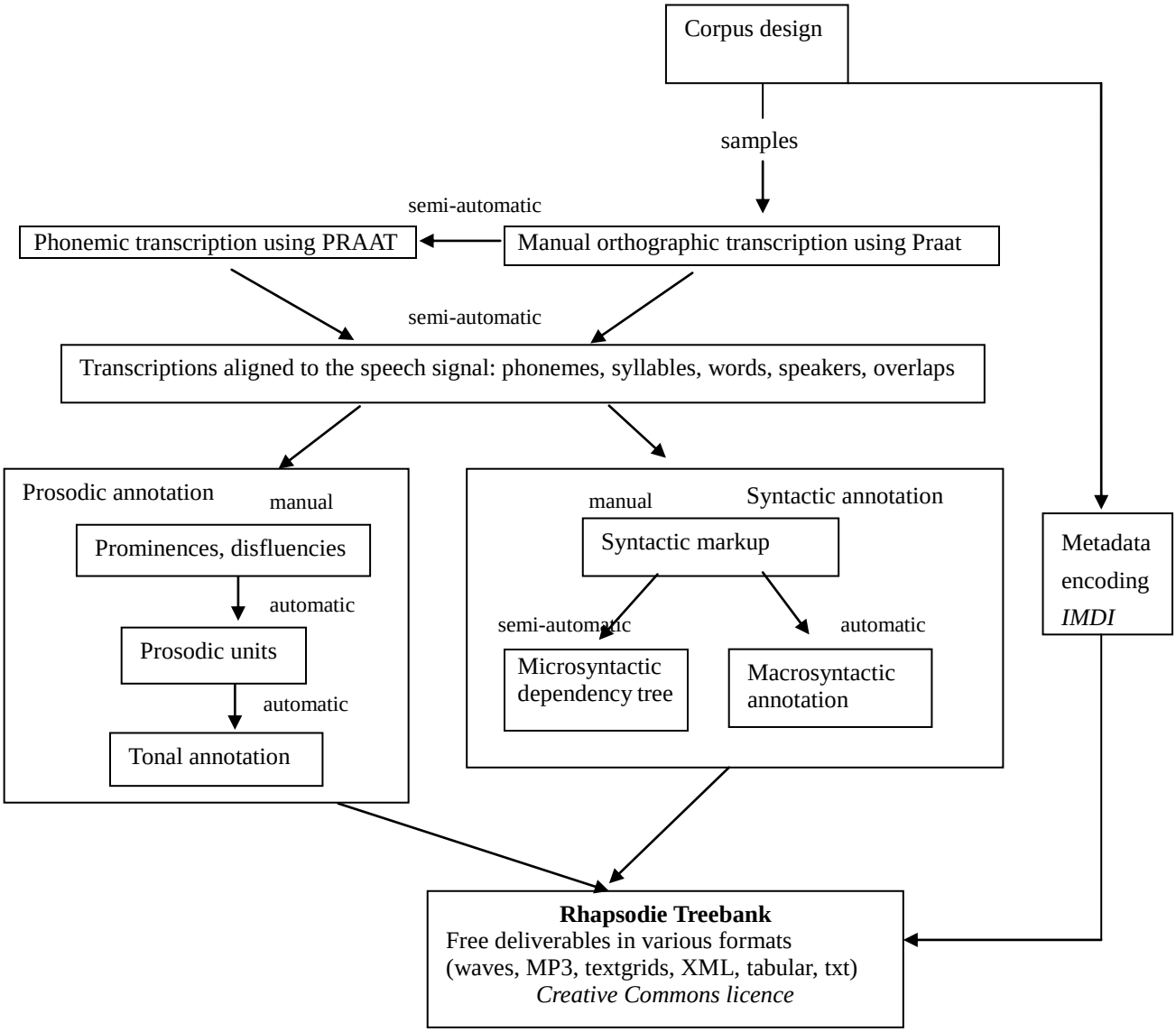

Figure 4. Processing chain for the Rhapsodie Treebank : an overview

5. Acknowledgements

The Rhapsodie project was supported by the French National Agency for Research (ANR).

6. References

't Hart, J., Collier, R. and Cohen, A. (1990). *Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge: Cambridge University Press.

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

Béguelin, M-J, Avanzi, M. and Corminboeuf G. (Eds.) (2010). *La parataxe*, Bern, Peter-Lang.

Biber, D. and Conrad, S. (2009). *Register, Genre and Style*. Cambridge, CUP.

Blanche-Benveniste, Cl., Bilger, M., Rouget, Ch. and Van den Eyende, K. (1990). *Le français parlé. Etudes grammaticales*. Paris: Editions du Centre National de la Recherche Scientifique.

Branca, S., Fleury, S., Lefeuvre, Fl., and 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., and Windhouwer, M. (2012). Standardizing a component metadata infrastructure. In *Proceedings of LREC*, Istanbul, Turkey, pp.1387-1390.

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

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

Deulofeu, J., Gerdes, K., Kahane S., and Pietrandrea, P. (2010). Depends on what the French say: Spoken corpus annotation with and beyond syntactic function. In *Proceedings of the 4th Linguistic Annotation Workshop* (*LAW IV*), ACL, Uppsala, Sweden, pp. 274-281.

Dister, A., Simon, A.C. (2008). La transcription synchronisée des corpus oraux. Un aller-retour entre théorie, méthodologie et traitement informatisé. *Arena Romanistica* 1/1, pp. 54-79

Durand, J., Laks, B. and 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, pp. 19-61.

Goldman, J.-Ph. (2011). Easyalign: an automatic phonetic alignment tool under praat. In *Proceedings of Interspeech-2011*, Florence, Italy, pp. 3233-3236.

Kahane, S., Pietrandrea, P. (2012). La typologie des entassements en français, In *Proceedings of the 3rd congrès mondial de linguistique française* (*CMLF*), Lyon, France, pp. 1809 – 1828.

Koch, P, Oesterreicher (2001). Langage parlé et langage écrit. In *Lexicon der Romanitischen Linguistik*, T1-2, Tübingen, Max Niemeyer Verlag, pp. 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*, pp. 55-.73.

Lacheret, A., Obin, N. and Avanzi, M. (2010). Design and evaluation of shared prosodic annotation for spontaneous French speech: from expert knowledge to non-expert annotation. In *Proceedings of the 4th Linguistic Annotation Workshop (LAWIV),* Uppsala, Sweden, pp. 265-273.

Lacheret, A., Kahane, S. and Pietrandrea, P. (Eds.) (forthcoming). *Rhapsodie. A Prosodic-Syntactic Treebank for Spoken French.* Amsterdam/ Philadelphia, John Benjamins Publishing Company.

Obin, N., Beliao, J., Veaux Ch. and Lacheret A. (2014). SLAM: Automatic Stylization and labelling of Speech Melody. In *Proceedings of Speech Prosody*, Dublin, Ireland http://www.speechprosody2014.org/.

Rossi, M. (1979). Le français, langue sans accent ? in I. Fonagy & P. Léon (Eds.), *L'accent en français contemporain*, Studia Phonetica, 15, Paris, Didier, pp. 13-51.

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

Wightman, Colin W. (2002). Tobi Or Not Tobi?. In *Proceedings of Speech Prosody*, Aix-en-Provence, France, http://sprosig.isle.illinois.edu/sp2002/.