

# Syntactic annotation of spontaneous speech: application to call-center conversation data

Thierry Bazillon, Melanie Deplano, Frederic Bechet, Alexis Nasr, Benoit Favre

Aix Marseille Univ, LIF-CNRS, Marseille, France

firstname.lastname@lif.univ-mrs.fr

## Abstract

This paper describes the syntactic annotation process of the DECODA corpus. This corpus contains manual transcriptions of spoken conversations recorded in the French call-center of the Paris Public Transport Authority (RATP). Three levels of syntactic annotation have been performed with a semi-supervised approach: POS tags, Syntactic Chunks and Dependency parses. The main idea is to use off-the-shelf NLP tools and models, originally developed and trained on written text, to perform a first automatic annotation on the manually transcribed corpus. At the same time a fully manual annotation process is performed on a subset of the original corpus, called the GOLD corpus. An iterative process is then applied, consisting in manually correcting errors found in the automatic annotations, retraining the linguistic models of the NLP tools on this corrected corpus, then checking the quality of the adapted models on the fully manual annotations of the GOLD corpus. This process iterates until a certain error rate is reached. This paper describes this process, the main issues raising when adapting NLP tools to process speech transcriptions, and presents the first evaluations performed with these new adapted tools.

**Keywords:** Speech Analytics, Dependency Parsing, Semi-supervised annotation process

## 1. Introduction

This study describes the syntactic annotation process developed on the DECODA corpus. This corpus contains transcriptions of Human-Human conversations collected in a French public transport call-center (RATP). The goal of the French ANR DECODA project is to propose new speech analytics methods targetting two applicative frameworks:

- punctual analysis of large dialog corpora for data mining purposes, like detecting a problem in the call-center behaviour, or extracting knowledge about the call-center performance;
- periodic analysis, or monitoring of the call-center by a day-by-day analysis of the call-center dialog logs.

Both frameworks are based on the automatic semantic analysis of Human-Human spoken conversations. The semantic interpretation of a spoken utterance can be split into a two-level process: a tagging process projecting lexical items into basic conceptual constituents and a composition process that takes as input these basic constituents and combines them in a possibly complex semantic interpretation of the utterance, represented, for example, as a set of semantic Frames. Various methods, reviewed in (De Mori et al., 2008), have been proposed for both levels of this process, from statistical tagging approaches to parsing methods.

Syntactic information is useful to perform such an understanding process: at the concept level, syntax can help reducing the ambiguity through semantic role labelling; at the semantic Frame level, syntactic dependencies can be projected into semantic dependencies to obtain structured semantic objects.

Despite its usefulness, syntactic parsing is not always considered when building a Spoken Language Understanding (SLU) system dedicated to process spontaneous speech because of two main issues: firstly transcriptions obtained

through an Automatic Speech Recognition (ASR) process contain errors, the amount of errors increasing with the level of spontaneity in speech; secondly, spontaneous speech transcriptions are often difficult to parse using a grammar developed for written text due to the specificities of spontaneous speech syntax (agrammaticality, disfluences such as repairs, false starts or repetitions). The first issue is currently tackled in the DECODA project with the use of methods dealing with ambiguous inputs, such as word lattices produced by an Automatic Speech Recognition (ASR) system. The second issue is the target of this paper.

## 2. Syntactic parsing of spontaneous speech

Syntactic parsing has been mainly studied for written language. It aims to uncover the relationship between words (e.g. constituency, dependency) within a sentence and guide the construction of the semantic representation in the language processing pipeline. Parsing is traditionally tightly connected to rewriting grammars, usually context free grammars, used together with a disambiguation model. Many current state-of-the-art text parsers are built on this model, such as (Petrov and Klein, 2007). Shallow syntactic processes, including part-of-speech and syntactic chunk tagging, are usually performed in the first stage.

The traditional view of parsing based on context-free grammars is not suitable for processing speech: due to ungrammatical structures in spontaneous speech, writing a generative grammar and annotating transcripts with that grammar remains difficult. New approaches to parsing based on dependency structures and discriminative machine learning techniques (Nivre, 2003; McDonald et al., 2005) are much easier to adapt to speech for two main reasons: (a) they need less training data and (b) the annotation of speech transcripts with syntactic dependencies is simpler than with syntactic constituents. Another advantage is that partial annotation can be performed when the speech is ungrammatical or the ASR transcripts are erroneous (Béchet and Nasr,

2009). The dependency parsing framework also generates parses much closer to predicate argument structuring which eases semantic interpretation.

In order to train such a dependency parser for the DECODA applicative frameworks we have selected a set of dialogs from the DECODA corpus. These dialogs have been manually annotated at the POS and dependency parse levels. A Graph-based dependency parser (McDonald et al., 2005) based on the (Bohnet, 2010) implementation has been trained and evaluated on this corpus. This paper describes the annotation process as well as the first results obtained.

### 3. Annotation process

#### 3.1. The DECODA corpus

In the DECODA project<sup>1</sup> we are dealing with the call-center of the Paris transport authority (RATP). This applicative framework is very interesting because it allows us to easily collect large amount of data, from a large range of speakers, with very few personal data. Indeed people hardly introduce themselves while phoning to obtain bus or subway directions, ask for a lost luggage or for information about the traffic. Therefore this kind of data can be anonymised without erasing a lot of signal. The DECODA corpus currently collected within the project has been fully anonymized, manually segmented and transcribed. The current state of the corpus is made of 1514 dialogs, corresponding to about 74 hours of signal. The average duration of a dialog is about 3 minutes.

#### 3.2. POS and chunk annotation

We performed three levels of annotation on the manual transcriptions of the DECODA corpus: Part-Of-Speech tags, chunk tags and dependency links. The first two levels were performed by the NLP suite MACAON (Nasr et al., 2011)<sup>2</sup>. The MACAON POS tagger is based on an HMM approach. The baseline models have been trained on the French TreeBank (Abeillé et al., 2003) corpus containing articles from the French newspaper *Le Monde*. Therefore a lot of tagging errors occurred when these models were applied to the DECODA corpus, since written and spoken French have a lot of dissimilarities. Among these, the main sources of errors were the use of personal pronouns, lexical ambiguities due to discourse markers or syntactic forms specific to oral French and spoken disfluencies. Table 1 presents some examples of these ambiguities.

Issue	Example
personal pronouns	<i>toi tu veux prendre cet appel ?</i>
discourse markers	<i>bon c'est vrai qu'il a pas tort quoi</i>
disfluencies	<i>c'est un peu le le le principe</i>

Table 1: Main Part-Of-Speech tagging ambiguities on the DECODA corpus

In order to adapt the HMM models to the specificities of oral French, we developed an iterative process consisting

in manually correcting errors found in the automatic annotations thanks to a WEB-based interface. This interface allows to write regular expressions on the POS tags and the lexical forms in order to correct the annotations on the whole DECODA corpus. Then the HMM models of the MACAON tagger are retrained with this corrected corpus. A small subset of the DECODA corpus has been set aside and fully manually annotated at the POS level. This sub-corpus, called *DECODA-GOLD*, is used to control the quality of the tagger retrained after correcting the DECODA training corpus. When the POS error rate is considered acceptable, this correction process stops.

This process is described in figure 1.

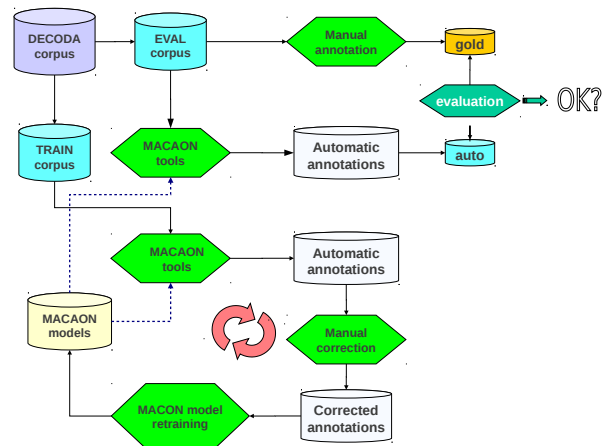


Figure 1: Semi-supervised annotation process

The same process was applied at the syntactic chunk level. This level is implemented in MACAON by a set of regular grammars representing all the possible chunks described by sequences of POS tags. Most of the chunking errors were due to the analysis of non-grammatical sentences which lead the chunk grammars to erroneous matches: the grammars have been defined on written French, and oral particularities like detachments make some rules non applicable, as shown in table 2: the sequences "determiner+noun+pronoun" and "determiner+proper noun+noun" refer to a single nominal chunk in written text, but may be much more ambiguous in an oral conversation.

Moreover, some chunks may be "broken" by speech disfluencies (*les horaires du bus numero je pense trois cent trente*), which make the chunk grammar inapplicable in those cases.

In order to deal with these issues, all the repetitions, false starts and discourse markers have been annotated in the corpus, then removed before applying the chunking process. In addition to syntactic chunks, all the named entities such as bus number, adress, metro lines, etc. have been also annotated in the DECODA training corpus.

Finally all the MACAON models have been retrained on the corrected corpus.

<sup>1</sup><http://decoda.univ-avignon.fr>

<sup>2</sup><http://macaon.lif.univ-mrs.fr>

Patterns	Written	Oral
det+nc+pro	<i>le gouvernement lui se réserve le droit d'intervenir</i>	<i>le trajet moi ça me semble très long</i>
det+np+nc	<i>le Molière comédien est moins célèbre que le Molière auteur</i>	<i>le Navigo monsieur c'est 15 euros</i>

Table 2: Chunking patterns applied to oral and written language

### 3.3. Syntactic dependencies

In order to train a statistical dependency parser directly on the corpus, as presented in section 2., we added word dependency annotation to the whole DECODA corpus. Manually annotating a large corpus such as the DECODA corpus with syntactic dependency links for each word is a difficult and costly task. Therefore we applied a process similar to the one presented in figure 1. A first dependency parser, originally trained on written French is adapted to the DECODA corpus thanks to the manual annotation of a subset of the corpus. This parser is applied to the whole DECODA corpus, the corpus is corrected and the statistical models of the parser are retrained on this corrected corpus.

In order to speed up the manual annotation process of dependency parses, we decided to perform this annotation at the chunk level: instead of connecting each word of a sentence, only the chunks are connected with each other. In a second stage, an automatic process is in charge of projecting this annotation at the word level.

All the syntactic dependencies between chunks were manually added thanks to a WEB interface on the *DECODA GOLD* corpus.

The syntactic model used in this study is derived from the French TreeBank annotation guide<sup>3</sup>. However, our annotation process focuses on the chunk level, whereas the French TreeBank focuses on the part-of-speech level. As a consequence a few annotation conventions have been simplified, since word-to-word links weren't needed here.

15 types of syntactic dependencies have been used in our annotation process:

- subject (suj): Jean←est mon ami
- impersonal subject (suj\_imp): il←pleut beaucoup ce matin
- direct object (obj): je lis→le journal
- indirect object with *de* preposition (de\_obj): il se souvient→de ses vacances
- indirect object with *à* preposition (a\_obj): il pense→à toi
- indirect object introduced with another preposition (p\_obj): il compte→sur toi
- locative object (p\_obj\_loc): j'habite→à Marseille
- coordination (coord): du pain→et des jeux
- dependant of the coordination (dep\_coord): du pain et→des jeux
- subject attribute (ats): je suis→content

- object attribute (ato): il me trouve→intelligent
- reflexive pronoun (aff): je me←lève
- relative subordinate clause (mod\_rel): l'homme→qui rit
- comparative (arg\_comp): il est plus grand→que toi
- adverbial phrase (mod): il travaille→depuis deux jours

The syntactic dependency annotation process does not always lead to complete parses since, as already mentioned, the DECODA corpus is a spontaneous speech corpus. Some words may not be connected to others in the syntactic tree due to speech peculiarities, which will be illustrated in the next section.

After having manually established the syntactic dependencies at the chunk level, an automatic process based on the POS patterns of the chunks was in charge of projecting the links from the chunk to the word level. About 195 of these patterns have been used, covering most of DECODA chunks structures. A few examples of them are given in table 3:

In a lot of cases, lexicalized patterns were needed because a generic one would have been too ambiguous. For example, in the pattern "prep+clo+vinf" (preposition+object clitic+infinitive verb), the clitic may be analyzed in three different ways: direct object in "pour le prendre"; indirect object in "pour me dire" and locative object in "pour y aller". As a consequence, more than 200 word-to-word patterns have been built to solve these ambiguities.

## 4. Experiments

The annotation corpus was made of 156 dialogs, containing 34K words. The dialog durations are between thirty seconds and twelve minutes. All the dialogs have been annotated by two human annotators. A subset of 20 dialogs has been annotated by both annotators in order to check inter-annotator agreement. Every dialog is segmented into chunks, and every chunk is displayed on an horizontal line which indicates: the chunk position in the dialog; the chunk content; the POS tagging of each word inside the chunk; the chunk type.

Unlike written texts, sentences in our corpus can contain chunks or groups of chunks not connected to the rest of the sentence. It may happen in the case of spoken disfluencies such as false starts or juxtaposed structures. Indeed juxtaposed structures are very frequent in oral conversations, where speakers don't always use relative pronouns, subordinative conjunctions or coordinative conjunctions to articulate their speech. As a consequence, in a sentence like *vous patientez, je regarde, je vous reprends après*, the three

<sup>3</sup><http://alpage.inria.fr/statgram/frdep/Publications/FTB-GuideDepSurface.pdf>

Chunks	Structure	Links	Exemples
Nominal Chunk	det nc det adj nc	det(1,2) det(1,3), mod(2,3)	le bus la mauvaise ligne
Verbal Chunk	clneg v advneg v vinf	mod(1,2), mod(3,2) obj(2,1)	ne peux pas pouvez venir
Prepositional Chunk	prep pro prep det nc np	obj(2,1) obj(3,1) det(2,3) mod(4,3)	devant vous avec un forfait Navigo

Table 3: Some POS patterns used for automatic word-level tagging

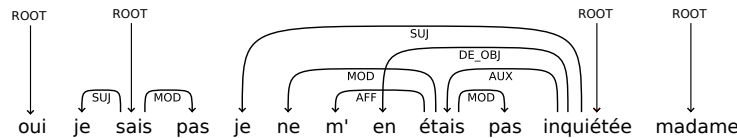


Figure 2: Example of multi-root dependency parse of an utterance

verbs can't be linked together because there isn't any syntactical articulation between them. Such sentences are very frequent in DECODA.

Annotation may sometimes be ambiguous, especially when a speaker uses grammatical structures that are clearly syntactically incorrect. Sentences like *la personne que vous avez dit que vous me passerez* or *je voudrais savoir qu'est-ce que je dois faire* are frequent in oral conversations, but agrammatical. In these cases, annotators had to try establishing coherent links, as if the erroneous structures were the correct ones. Besides, several other ambiguities due to spoken phenomena had to be dealt with during annotation:

- lots of multiple relations: *lui il passe a Villemomble ce bus*
- dependencies difficult to assess: *j'arrive pas a acceder au / ouais ça marche pas / site web*
- sequences of chunks with no dependencies (different dialog acts): *bonne journee // merci // au revoir*
- cleft sentences: *ce que vous voulez, ce sont les horaires*

The part-of-speech model adaptation has been evaluated on this annotated corpus. Two kinds of models have been considered: baseline models (trained on the French Treebank) and DECODA models (trained on the corrected "train" corpus). In addition to the DECODA corpus, another corpus has been taken into account: EPAC, which is made of broadcast conversations (radio interviews, radio talk-shows). The results shown in table 4 indicate more than 50% error reduction on both corpora by using the DECODA adapted models.

Corpus / Models	Baseline	DECODA models
DECODA	21.0%	8.5%
EPAC	13.3%	4.5%

Table 4: Evaluation of the POS model adaptation

As a consistency check, we performed a first evaluation of a syntactic parser trained on the DECODA corpus. The parser is a graph-based dependency parser (McDonald et al., 2005) with second order features (involving any combination of three words) and the maximum spanning tree decoder by (Carreras, 2007). Features are based on words, part-of-speech tags and dependency direction and labels. The MATE parser (Bohnet, 2010) is trained with the passive-aggressive perceptron update rule (Crammer et al., 2006) and optimized for speed. It performed at the state of the art during multiple evaluation campaigns.

We used 80% of the speakers turns to train the parser, 10% for tuning the parameters and 10% for evaluating its performance. The parser was trained and tested on the reference texts with reference tags in order to suppress the confounding effect of ASR errors. The results are given in table 6. The specificities of speech add three issues to the reference parse trees:

- Non-projectivity: arcs in the dependency tree may intersect, which might prevent from using certain classes of parsing algorithms.
- Multiple-roots: speaker turns are composed of juxtaposed speech acts, as stated previously.
- Overlap: disjoint trees from speech acts might overlap (the most frequent case is when a small speech act is embedded under the tree of a larger speech act).

In order to reduce the multiple root issue, we applied two simple strategies that consist in completing the turn-level tree with artificial dependencies. The first approach links the governor of a disjoint subtree as dependent of the governor of the following tree (*Governor after*), and the second links it to the previous tree (*Governor before*). Figure 3 shows illustration of this tree completion strategy. The two strategies, when applied to the DECODA corpus, remove the multiroot issue and decrease non-projectivity at the cost of a small increase in the number of overlapping subtrees (Table 5).

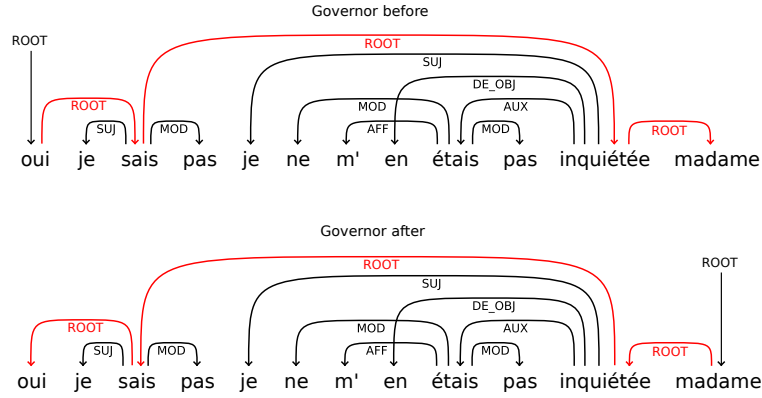


Figure 3: Two linking strategies for multi-root utterance: *Governor after* and *Governor before*

	Non projective	Multiroot	Overlapping
Baseline	8.21%	40.82%	1.29%
Governor after	7.83%	0%	4.81%
Governor before	6.49%	0%	2.68%

Table 5: Characteristics of the dependency parses obtained on the DECODA corpus according to the type of treatment applied to the multi-root sentences

Performance wise, the results are in the same range as newspaper text dependency parsing, in term of unlabeled accuracy score (UAS, rate of correct dependency governor/dependent couples), and labeled accuracy score (LAS, also accounting for dependency label correctness). This is remarkable given the added difficulty of missing punctuation and case information, and spontaneous-speech-specific phenomena. However, utterances are relatively small compared to text, and therefore contain less difficult syntactic structures such as relative clauses.

	DEV		TEST	
	LAS	UAS	LAS	UAS
Baseline	87.72%	91.41%	87.71%	91.19%
Governor after	87.33%	91.15%	87.87%	91.36%
Governor before	87.57%	91.41%	87.37%	90.80%

Table 6: Evaluation of the parsing accuracy with the MATE parser on the DECODA corpus. *LAS* is for Labelled Accuracy Score and *UAS* stands for Unlabelled Accuracy Score

## 5. Conclusion

In the medium and long term, the result of this annotation task will be very useful to study spontaneous speech syntactic structures. More precisely, we plan to study verbal subcategorisation in spontaneous speech and contrast it with subcategorisation in written language. An accurate description of the syntactic behaviour of verbs is of main importance for the design of Natural Language Processing tools that extract valuable information from conversational data.

## 6. Acknowledgements

This work is supported by the French agency ANR, Project DECODA, contract no 2009-CORD-005-01, and the French business clusters Cap Digital and SCS. For more information about the DECODA project, please visit the project home-page, <http://decoda.univ-avignon.fr>

## 7. References

- A. Abeillé, L. Clément, and F. Toussnel. 2003. Building a treebank for french. In Anne Abeillé, editor, *Treebanks*. Kluwer, Dordrecht.
- Bernd Bohnet. 2010. Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of COLING*.
- Frédéric Béchet and Alexis Nasr. 2009. Robust dependency parsing for spoken language understanding of spontaneous speech. In *Interspeech*, Brighton.
- X. Carreras. 2007. Experiments with a higher-order projective dependency parser. In *Proceedings of the CoNLL Shared Task Session of EMNLP-CoNLL*, volume 7, pages 957–961.
- K. Crammer, O. Dekel, J. Keshet, S. Shalev-Shwartz, and Y. Singer. 2006. Online passive-aggressive algorithms. *The Journal of Machine Learning Research*, 7:551–585.
- R. De Mori, F. Bechet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur. 2008. Spoken language understanding. *Signal Processing Magazine, IEEE*, 25(3):50–58, May.
- Ryan McDonald, Koby Crammer, and Fernando Pereira. 2005. Online Large-Margin Training of Dependency Parsers. In *Association for Computational Linguistics*.
- Alexis Nasr, Frederic Bechet, Jean-Francois Rey, and Joseph Le Roux. 2011. Macaon:a linguistic tool suite for processing word lattices. In *The 49th Annual Meeting of the Association for Computational Linguistics: demonstration session*.
- J. Nivre. 2003. An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies*.
- Slav Petrov and Dan Klein. 2007. Improved Inference for Unlexicalized Parsing. In *HLT-NAACL*, pages 404–411.