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

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

Aix Marseille Univ, LIF-CNRS, Marseille, France

firstname.lastname@lif.univ-mrs.fr

1. Context of this study

This study describes the syntactic annotation process developped on the DECODA corpus. This corpus contains transcriptions of Human-Human conversations collected in a French public transport call-centre (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 combine them in a possibly complex semantic interpretation of the utterance, represented, for example, as a set of semantic Frames. Various methods, reviewed in [3], 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 word relationships (e.g. word order, constituency, dependency) within a sentence and control 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 [7]. 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 [6, 4] are much easier to adapt to speech for two main reasons: (a) they need less training data and (b) the annotation with syntactic dependencies of speech transcripts 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 [1]. 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 [4] based on the [2] implementation has been trained and evaluated on this corpus. This paper descibes the annotation pro-

cess as well as the first results obtained.

3. Annotation process

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 [5][1] then manually checked for errors. Syntactic dependencies between chunks were manually added thanks to a WEB interface. We chose to put the links at the chunk level rather than the word level in order to speed up the annotation process. An automatic process based on the POS patterns of the chunks is in charge of projecting the links from the chunk to the word level.

The syntactic model used in this study is derived from the French TreeBank annotation guide[2]. 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 mentionned, the DECODA corpus is a spontaneous speech corpus. Some words may not be connected to others in the syntactic tree due to speech pecularities.

4. Corpus and first 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 partientez, je regarde, je vous reprends après*, the three 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 syntacticaly 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.

A first evaluation on the usefulness of this annotated corpus has been made by training on it [2] statistical dependency parser. We used 80% of the speakers turns to train the parser, 10% for tuning the parameters and 10% for evaluating its performances. The results are given in table 1. Although the raw results are lower than those obtained on written text, they are encouraging considering the spontaneous speech phenomenon occurring in the corpus.

[1] `http://macaon.lif.univ-mrs.fr/`

[2] http://alpage.inria.fr/statgram/frdep/Publications/FTB-GuideDepSurface.pdf

Evaluation	(LAS)	Unlabelled accuracy score
Dev	82.4	89.2
Test	80.8	87.4

Table 1: 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 should be very useful to study spontaneous speech grammatical structures, as most of the other works related to syntactic dependencies were focused either on part-of-speech level, either on prepared speech. This annotated corpus will give the opportunity to study how verbal constructions work in oral conversations, with which agents, in which way, etc. As far as automatic treatment of language is concerned, it should be an interesting starting point to develop tools that are able to extract informations about verbs valence and oral grammatical structures.

6. References

[1] Frédéric Béchet and Alexis Nasr. Robust dependency parsing for spoken language understanding of spontaneous speech. In *Interspeech*, Brighton, United Kingdom, 2009.

[2] Bernd Bohnet. Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of COLING*, 2010.

[3] R. De Mori, F. Bechet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur. Spoken language understanding. *Signal Processing Magazine, IEEE*, 25(3):50–58, May 2008.

[4] Ryan McDonald, Koby Crammer, and Fernando Pereira. Online Large-Margin Training of Dependency Parsers. In *Association for Computational Linguistics (ACL)*, 2005.

[5] Alexis Nasr, Frederic Bechet, Jean-Francois Rey, and Joseph Le Roux. Macaon:a linguistic tool suite for processing word lattices. In *The 49th Annual Meeting of the Association for Computational Linguistics: demonstration session*, 2011.

[6] J. Nivre. An efficient algorithm for projective dependency parsing. In *Proceedings of the 8th International Workshop on Parsing Technologies (IWPT*, 2003.

[7] Slav Petrov and Dan Klein. Improved Inference for Unlexicalized Parsing. In *HLT-NAACL*, pages 404–411, 2007.