Bidirectionnal converter between syntactic annotations : from French Treebank Dependencies to PASSAGE annotations, and back

Munshi Asadullah[1] Patrick Paroubek[1] Anne Vilnat[1,2]
[1]LIMSI-CNRS
BP 133 91403 Orsay cedex, FRANCE
{firstName.lastName}@limsi.fr
[2] University Paris-Sud

Abstract

We present the first version of a bidirectional converter between the PASSAGE annotations and the French Tree-bank Dependency (FTB-DEP) annotations. FTB-DEP is the syntactic representation of several freely available parsers and the PASSAGE annotation was used to hand-annotate a relatively large sized corpus, that served as gold-standard in the PASSAGE evaluation campaigns. Our converter will give the means to evaluate these parsers on the PASSAGE corpus. We shall illustrate the mapping of important syntactic phenomena using the corpus made of the examples of the FTB-DEP annotation guidelines, which we have hand-annotated with PASSAGE annotations and employed to compute quantitative performance measures on the FTB-DEP guidelines. The examples we have selected here for illustrating the back converter from ftb-dep to PASSAGE concern passive voice constructions.

1. Introduction

The motivation for a cross formalism conversion system presented in this paper is based on the following two observations: the significant, freely available and ready to use statistical French dependency parsers, are essentially the parsers described in (Candito et al., 2010b). These parsers[1] produce their outputs in an adapted CoNLL data format[2] (Buchholz and Marsi, 2006), following the annotation standards of the dependency tree-bank FTB-UC-DEP (Candito et al., 2010a), extracted and converted from the French Tree-bank (FTB) (Abeillé and Barrier, 2004). On the other hand, for evaluating the performance of these parsers objectively, very little annotated corpus exist in the same annotated format, except for of course the SEQUOIA corpus (Candito and Seddah, 2012), which contains 3,204 annotated sentences. In contrast, the annotated corpus produced during the PASSAGE campaigns (Vilnat et al., 2010; De la Clergerie et al., 2008), contains at least 14,000 manually annotated sentences, of which 8,200 sentences are available without any copyright restrictions[3], part of which includes sentences from various genre, obtained during the EASY campaign (Paroubek et al., 2006).

However, this corpus uses the PASSAGE annotation format, an adaptation of the Grammar Relations[4] GR (Lin, 1998; Carroll et al., 1998), that also introduces constituent like groups and a level of syntactic and grammatical functions specific to the project, which are for the most part a generalization of the standard FTB annotation. It is obvious that the PASSAGE format can provide an evaluation platform with substantial amount of gold standard test data thus a cross-converter (from and to PASSAGE) opens the possibility of cross formalism evaluation.

In this paper we will briefly introduce the two annotation formats. Then, we detail the two converters, and the rules which have been written. The last part will detail the results we obtained on the phenomenon we mostly study, the passive forms. Our goal is to evaluate the converters by a double conversion, from PASSAGE to FTB-DEP and back.

2. Different annotations to deal with identical syntactic phenomena

2.1. FTB-DEP in CoNLL

FTB Surface Dependency Annotation Guide (Candito et al., 2011)[5] lists the basic annotation guideline for FTB-DEP formalism. FTB-DEP is based on the Dependency Grammar (DG) (Tesnière, 1959) formalism and like all DG based formalisms adapted the relation types according to the target language and domain. Among the base relations, there are 12 relations to annotate the relations of a token with the verbal governors and 8 relations to annotate the relations with non verbal governors. There are 8 more more specific relations reserved for manual annotation. The first contrast of this formalism with PASSAGE is the lack of any notion of syntactic group, whereas the groups in PASSAGE are merely an abstraction provided to generalise syntactic relations. Another key difference is the presence of the virtual ROOT element in the FTB-DEP, which is the hierarchical nucleus of a sentence and a natural extension for many formalisms of the DG family (Nivre, 2006).

Among the relations reserved for verbal governors. the SUJ (subject) relation is parallel to its PASSAGE counterpart and it is true for the ATS (attribute to the subject) and ATO (Attribute to the object), although both these relations can be represented in PASSAGE with the ATB_SO relation. ATB_SO, though, takes 3 arguments and the third argument is to distinguish the case of an attribute to

[1]Adapted for French within ANR SEQUOIA project and freely available at: http://alpage.inria.fr/statgram/frdep/fr_stat_dep_parsing.html

[2]https://ilk.uvt.nl/conll/\#dataformat

[3]This corpus is available on: http://www.elda.org, where one can also get access to the copyrighted contents.

[4]We are using the term Grammatical Relations to refer to syntactic dependencies between heads and dependent following (Carroll et al., 1998)

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

the subject from an attribute to the object. FTB-DEP also distinguishes between the temporal auxiliary (AUX_TPS) and the auxiliary in a passive construction. Furthermore FTB-DEP has fine grained distinction between direct object of the verb (OBJ) and indirect object types. In FTB-DEP, there even is a distinction between the indirect object with the preposition “de” (DE_OBJ), from those with the preposition“à” (A_OBJ) and from those with all other prepositions (P_OBJ). Verbal modifiers of all sort are expressed with the MOD relation (example 1), whereas PASSAGE even draws distinction for the modifier of a preposition. The final dependency in this group is AFF, for linking a clitic pronoun to its verb in case of a frozen construction like particle verbs in English (example 3).

“*Paul travaille le samedi*” ... (1)
(Paul works on saturdays)

“*la voiture bleue*” ... (2)
(the blue car)

“*Il se souvient*” ... (3)
(He *himself* remembers)

Among the relations that deal with non verbal governors, FTB-DEP have the MOD being a common relation since non verbal governors can also be modified ((example 2)), although there is no distinction. MOD_REL is used to link the antecedent to the verb of the subordinate in relative clauses, while n-ary coordinations are represented using a combination of two dependencies COORD, for the first conjunct and DEP_COORD for the others. Similarly to coordination, ARG dependency will link preposition in case of (partially) frozen constructions (example 4).

“*tomber de Charybde en Scylla*” ... (4)
(on the horns of a dilemma)

The DEP is the most generic relation and HIERARCHICAL the least specific among all the relations and often used to fill the gaps where the system failed to determine any appropriate relation for a pair of tokens. PONCT (punctuations) is another generic relation: it connects all punctuation symbols to the virtual ROOT element, except for the punctuation symbols acting as coordinating conjunctions (as the comma between “*Mobiles*” and “*actifs*” in example 5).

“*Mobiles, actifs et médiatiques, on les entend partout.*” ... (5)
(mobile, active and mediatic, we hear them everywhere)

The relations reserved for manual annotation are primarily further specified form of the automatically generated relations, for example, MOD_LOC has been specified for the semantically locative (literal or figurative) modifiers. There is another specific modifier relation MOD_CLEFT and two specific P_OBJ relations: P_OBJ_AGT for passive and causal agents and P_OBJ_LOC for dependent locative argument.

CONLL is an extensible data format originally developed for evaluation campaigns in dependency parsing (Buchholz and Marsi, 2006) and is used by a large community. It allows to represent the words of a statement, the morphosyntactic information (parts of speech, lemmas, etc..) and syntactic dependencies. It uses a matrix representation where the first column is the token counter, starting at 1 for each new sentence, the second column contains the forms of the statement and the following columns contain their morphosyntactic tags, and finally syntactic dependencies. It is an extensible format where one can add new layers of analysis simply by adding columns to the matrix representation. Dependencies are represented by two columns, one for the type of addiction, one for the token ID of the target, which references a row of the matrix, the source of dependence is the current token or row. .

1	Je	cln	CL:CLS:s=suj:2:suj:2:suj
2	remercie	remercier	V:V:m=ind—n=s—p=3:0:root:0:root
3	le	le	D:DET:g=m—n=s—s=def:4:det:4:det
4	président	président	N:NC:g=m—n=s—s=c:2:obj:2:obj
5	en	en	P:P:p=3:4:dep:4:dep
6	excercice	excercice	N:NC:g=m—n=s—s=c:5:obj:5:obj
7	pour	pour	P:P:-:2:mod:2:mod
8	sa	son	D: DET:g=f—n=s—s=poss:9:det:9:det
9	réponse	réponse	N:NC:g=f—n=s—s=c:7:obj:7:obj
10	.	.	PONCT:PONCT:s=s:2:ponct:2:ponct

Table 1: CONLL Annotation Extract from The Sequoia Corpus v4.0

The SEQUOIA corpus uses the standard CONLL data format used for the CONLL–X[6], in which the last two columns were used for the dependency of the current token with its projective head[7]. An example of the annotation output in CONLL format is presented in Table 1 along with the graphical representation of the same sentence in Figure 1. In contrast, the output generated by the SEQUOIA parsers use the columns slightly differently, for example none of these parsers generates the dependency with the projective head for the tokens. Another example can be found in the output of the MaltParser[8] implementation presented in (Candito et al., 2010b) which has an extra column with word cluster identification numbers.

2.2. PASSAGE

PASSAGE (Vilnat et al., 2010; De la Clergerie et al., 2008) annotates both groups and dependency relationships[9], where groups are non-recursive minimum components. Six groups are defined: the nominal group (GN) the prepositional Group (GP), the verbal nucleus (NV), the adjectival group (GA), the adverbial phrase (GR) and the

[6]CoNLL–X Shared Task: Multi-lingual Dependency Parsing (http://ilk.uvt.nl/conll/)

[7]The dependency structure resulting from the this column is guaranteed to be projective.

[8]http://www.maltparser.org/

[9]The annotation guide for French: http://perso.limsi.fr/anne/PEAS_reference_annotations_v2.2.html

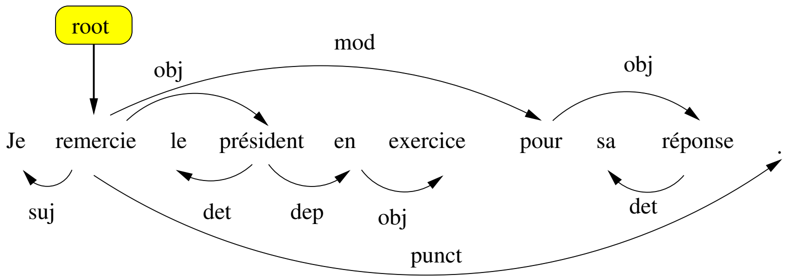

Figure 1: CoNLL Annotation Extract from The Sequoia Corpus v4.0 (Graphical Representation)

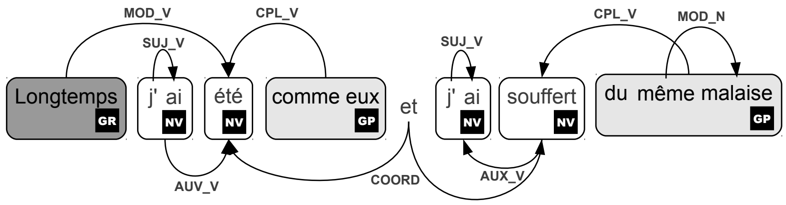

Figure 2: Example of the passage annotation Format (Graphical)

prepositional verbal nucleus (PV). 14 relations linked the groups or the word forms within these groups. There are, (SUJ-V) to link the subject to the verb, (AUX-V) to link the auxiliary to the verb, the direct object to the verb (COD-V), or the other objects to the verb (CPL-V) whether indirect or adjunct (it has not been distinguished), or any other optional modifiers to the verb (MOD-V). It can also annotate all other types of modifiers such as the modifiers to the nouns (MOD-N), to the adjectives (MOD-A), to the adverbs (MOD-R) and to the prepositions (MOD-P).

As explained before, the attribute of the subject or object (ATB_-SO) is annotated, and the relation between the introducer of a complement clause and its verb kernel (COMP). The last three relations are coordination (COORD), juxtaposition (JUXT) and apposition (APP). Figure 2 illustrates this annotation scheme. A comparison between PASSAGE annotations translated to English and those of SD adapted for PARK was presented by Paroubek et al. (2009). It is particularly illustrated that the PASSAGE annotation is much closer to the linguistic intuition than that of the SD based formalisms because it has the explicit relationship for prepositional modifier (MOD-P) while in SD formalisms, the modifiers of prepositions are attached to the head of the clause that contains it and not to the prepositions themselves (De Marneffe and Manning, 2008). However, although having very close resemblance with the GR formalism, PASSAGE does not explicitly represent passive constructions or deep analysis in its current form, but there are no restrictions to upgrade the formalism.

The PASSAGE corpus produces its outputs in XML format which allows the formalism to be independent of the possible constrains posed by the representation formats such the text format of SD. Furthermore, De Marneffe and Manning (2008), while explaining the limitation of SD mentioned the formalisms inability to represent ternary dependencies and having less linguistic information than PARK. In contrast, PASSAGE is virtually free from these limitations thus, the open and extensible XML representation allows ternary dependencies and additional linguistic features such as Named Entity (NE) etc. PASSAGE XML allows to incorporate low level information such as the token list and the formation of groups as well as high level information of the relations between groups or word elements.

3. How to convert?

We have developed two converters: from FTB-DEP to PASSAGE, and from PASSAGE to FTB-DEP

3.1. From FTB-DEP to PASSAGE

FTB-DEP	PASSAGE
A_OBJ(verb1, ?var1[10])	CPL_V(?var1, verb1)
ATS(?var1, ?var2)	ATB_SO(?var2, ?var1, “subject”)
ATO(?var1, ?var2)	ATB_SO(?var2, ?var1, “object”)
AUX_PASS(?var1, ?var2)	AUX_V(?var2, ?var1)
AUX_TPS(?var1, ?var2)	AUX_V(?var2, ?var1)
DE_OBJ(?var1, ?var2)	CPL_V(?var2, ?var1)
MOD(noun1, ?var1)	MOD_N(?var1, noun1)
MOD(verb1, ?var1)	MOD_V(?var1, verb1)
MOD(adj1, ?var1)	MOD_A(?var1, adj1)
MOD(adv1, ?var1)	MOD_R(?var1, adv1)
MOD_REL(noun1, ?var1)	MOD_N (?var1, noun1)
MOD_REL(verb1, ?var1)	MOD_V (?var1, verb1)
OBJ(prep1, noun1)	[GP(prep1, noun1)]
OBJ(verb1, ?var1)	COD_V(?var1, verb1)

Table 2: Simple Projection Rules from FTB-DEP to PASSAGE

In comparison to FTB-DEP, PASSAGE is a more generalized annotation format thus, the conversion from FTB-DEP to PASSAGE was dealt with a rule-based system with hierarchical rules triggered by the pattern sequence of the relations and the specific details of the arguments provided. Table 2 provides a detailed list of the simple projection rules form FTB-DEP to PASSAGE. Simple rules are composed of

FTB-DEP	PASSAGE
A_OBJ(verb1, “à”) + OBJ(“à”, verb2)	COD_V([PV(“à”, verb2)], verb1)
AUX_CAUS(?var1, ?var2) + SUJ(?var1, ?var3)	COD_V(?var2, ?var1) + SUJ_V(?var3, ?var2)
DE_OBJ(verb1, prep1) + OBJ(prep1, verb2)	CPL_V([PV(prep1, verb2)], verb1)
DE_OBJ(?var1, “que”) + OBJ(“que”, ?var2)	CPL_V(?var2, ?var1) + COMP(“que”, ?var2)
DE_OBJ(verb1, “que”) + OBJ(“que”, verb2)	COD_V(verb2, verb1) + COMP(“que”, verb2)
COORD(?var1, ?var2) + DEP_COORD(?var2, ?var3)	COORD(?var2, ?var1, ?var2)
MOD(verb1, prep1) + OBJ(prep1, noun1)	MOD_V([GP(prep1, noun1)], verb1)
MOD(verb1, prep1) + OBJ(prep1, verb2)	MOD_V([PV(prep1, verb2)], verb1)
MOD(verb1, soc1[11]) + OBJ(soc1, verb2)	CPL_V(verb1, verv2) + COMP(soc1, verb2)
MOD(noun1, cwl1[12]) + SUJ(verb1, noun1)	SUJ_V(noun1, verb1) + SUJ_V([GN(cwl1)], verb1)
MOD(noun1, adj1) + SUJ(?var1, noun1)	MOD_N(adj1, noun1) + SUJ_V(noun1, ?var1) + ATB_SO(adj1, ?var1, “subject”)
OBJ(verb1, prep1) + OBJ(prep1, verb2)	COD_V([PV(prep1, verb2)], verb1)
OBJ(verb1, “que”) + OBJ(“que”, verb2)	COD_V(verb2, verb1) + COMP(“que”, verb2)
OBJ(?var1, “de”) + OBJ(“de”, vinf1[13])	COD_V([PV(“de”, vinf1)], ?var1)
P_OBJ(?var1, prep1) + OBJ(prep1, ?var2)	CPL_V([GP(prep1, ?var2)], ?var1)
P_OBJ(verb1, “comme”) + OBJ(“comme”, ?var1)	MOD_V([GP(“comme”, ?var1)], verb1)
P_OBJ(verb1, “comme”) + OBJ(“comme”, ?var1)+ OBJ(?var1, verb2)	COMP(“comme”, verb2) + MOD_V(verb2, verb1)

Table 3: Compound Projection Rules from CoNLL-FTB to PASSAGE

```
16. MOD(adv1, ?var1) >
     MOD_R(?var1, adv1) ("mod",pos="R",null)|
    ("MOD_R",[0:1],[0:0])|
     null


17. MOD(verb1, prep1) + OBJ(prep1, noun1) >
     MOD_V([GP(prep1, noun1)], verb1)
       ("mod",pos="V",pos="P")+("obj",pos="P",pos="n")|
    ("MOD_V",[1:1],[0:0])+("GP",[0:1],[1:1])|([0:1]=[1:0])
```

Figure 3: Examples of the CoNLL-FTB to PASSAGE Rules

a single pattern to be detected in the FTB-DEP representation and converted to a single PASSAGE relation or group. Table 2 illustrates these rules. They have least priority of consuming a FTB-DEP relation pattern since the compound forms (presented in Table 3) suit the original purpose of the PASSAGE format, being less specific. Compound rules have more than one FTB-DEP pattern to be detected and express a constraint assigned to the arguments. Currently only the equality constraint has been implemented, i.e. at least one argument must be common between at least two relations. The converted pattern ranges from one or more group formation to multiple PASSAGE relations. Since the relations are treated differently in these formalisms, the arguments are often distributed differently in the resultant PASSAGE relations.

The pattern recogniser triggers rules from most specific to least specific primarily on the basis of the number of relations. A secondary hierarchy has been defined for the patterns having same number and types of relation on the basis of the specificity of their arguments. Each relation can be specified with surface form, lemma and part of speech, with surface form being most specific. Only the FTB-DEP relations specified for automatic annotation has been addressed by the conversion rules , yet some relations do not have any equivalent relation in the PASSAGE formalism, e.g. PONCT. The converter itself was written in Python 2. Each rule has three main declarations, the FTB-DEP relation pattern to be searched, the equivalent PASSAGE output and an optional last part, which contains the constraints that must be satisfied for the arguments of the FTB-DEP relations. There are also separate declaration sections for declaring the detail output patterns, list of words etc.

3.2. From PASSAGE to FTB-DEP

The working principle of the converter for the direction PASSAGE to FTB-DEP is illustrated in Figure 4. The map-

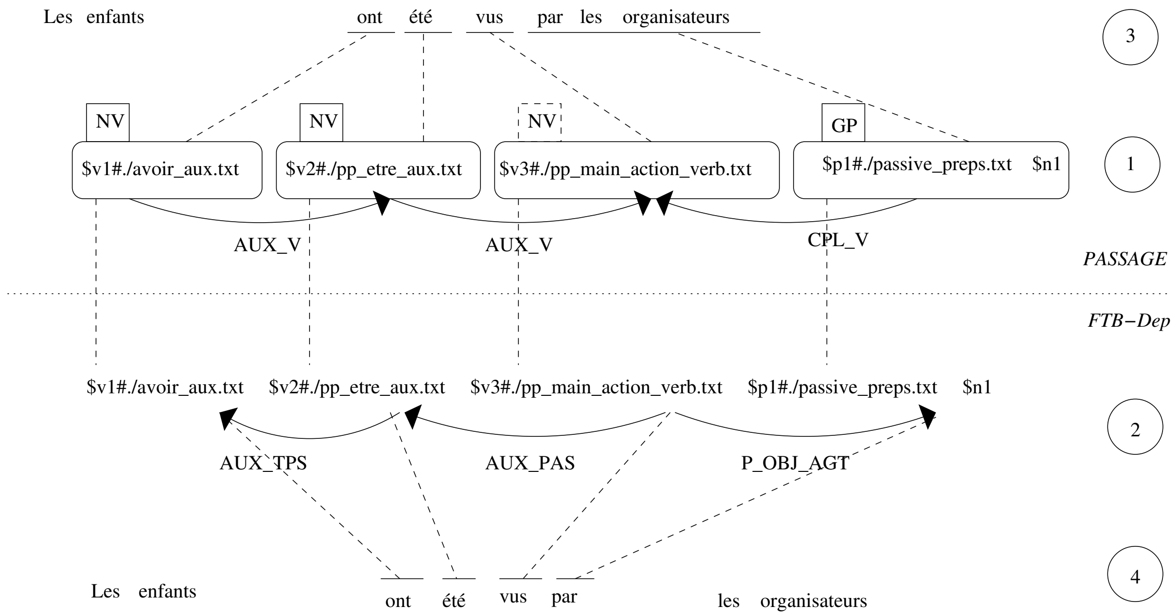

Figure 4: Example of a passive sentence with the both annotations

ping rules comprises two sets of aligned underspecified sentences: a set in PASSAGE that will match the input (step 1 in Fig. 4) and a set in FTB-DEP (step 2 in Fig. 4) to guide the generation of the dependencies. With respect to the representation of a regular syntactically annotated sentence, the converter replaces word forms by variables for pattern matching. These variables can be associated to arbitrary constraints bearing on the form, the part-of-speech or the lemma. Once the converting patterns have been written, pattern matching is applied on the input sentence and the bindings between the PASSAGE variables and the input word forms (step 3 in Fig. 4) are used to produce the FTB-DEP dependencies (step 4 in Fig. 4). The mapping between the PASSAGE and the FTB-DEP patterns is realized through the similarity of the sentences word forms and by the fact that there is a one to one correspondance between the sentences in the two annotation systems.
This process is illustrated on the passive sentence: *Les enfants ont été vus par les organisateurs.*[14]. For the passive form, PASSAGE annotates two AUX_V relations (from *ont* to *été* and from *été* to *vus*, without any distinction between them. Then CPL-V encodes the link between the verb and the agent which is annotated as all the other prepositional phrases related to a verb.
In FTB-DEP, the two auxiliary relations are distinguished: the first one is a temporal auxiliary (AUX-TPS), the second is a passive one (AUX-PAS). During the matching step, the main verb is recognized as an action verb, the second auxiliary as the auxiliary *be*, so this auxiliary relation is correctly recognized for the FTB-DEP annotation. For identifying action verbs we define a constraint on the corresponding verb variable using the content fo the LVF (Dubois and Dubois-Charlier, 1997)[15] This matching process also allows to annotate the relation P-OBJ-AGT, where PASSAGE annotates the CPL-V relation. The ROOT element is determined automatically from the PASSAGE annotations. Four patterns are enough to cover all cases of passive voice constructions (simple or compound tense for the verb, presence or absence of a deep subject).
We have taken different linguistic phenomena under consideration to design the different matching patterns. Our goal is to obtain a complete (at least, as complete as possible) coverage of the linguistic phenomena which have been observed, both in the French Treebank (to design the FTB-DEP annotation , encoded in CoNLL) and in the PASSAGE corpus.

4. Evaluation

A first evaluation performed with the FTB-DEP to PASSAGE converter is presented in (Asadullah et al., 2013). It gives the measures of precision, recall and fmeasure for PASSAGE relations obtained with the Berkeley Parser adapted to French (Candito et al., 2010c)[16] on an excerpt of 1584 sentences taken from the PASSAGE corpus (European Parliament EP & JRC). The highest performance is obtained for the AUX-V relation with a precision of 0.88, a recall of 0.75 and an f-measure of 0.81. At the time of writing we do not have yet evaluation results with the back converter (PASSAGE to FTB-DEP), but we already know that it identifies correctly the 6 sentences which hold passive constructions, out of the 184 sentences contained in the FTB-DEP guidelines, for which it produces correct FTB-DEP outputs.

5. Conclusion

We have presented a bidirectionnal converter from FTB-DEP syntactic annotations toward the PASSAGE annotation scheme. Our goal is to be able to go from one to the other of this annotation format, to be able to evaluate the results

[14]The children have been seen by the organizers.
[15]A browsing interface is available at: `http://rali.iro.umontreal.ca/Dubois/`
[16]`http://alpage.inria.fr/statgram/frdep/fr\_stat\_dep\_bky.html`

produced by different parsers, taking advantage of the validated annotations resulting of the PASSAGE project. From our experience, it seems to be easier in terms of data management to organise the mapping between the two annotations in terms of linguistic phenomena (e.g. passive voice constructions) rather than in terms of particulare dependencies or relations.

6. References

Abeillé, A. and Barrier, N. (2004). Enriching a french treebank. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC)*, Lisbon, Portugal, May. European Language Resources Association.

Asadullah, M., Vilnat, A., and Paroubek, P. (2013). Converting from the french treebank dependencies into passage syntactic annotations. In *Proceedings of the 6th Language and Technology Conference*, pages 188–192, Poznan, December.

Buchholz, S. and Marsi, E. (2006). Conll–x shared task on multilingual dependency parsing. In *Proceedings of the Tenth Conference on Computational Natural Language Learning*, CoNLL-X '06, pages 149–164, Stroudsburg, PA, USA. Association for Computational Linguistics.

Candito, M. and Seddah, D. (2012). Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical. In *Actes de la 19e conférence sur le Traitement Automatique des Langues Naturelles (TALN'2012)*, pages 321–334, Grenoble, France.

Candito, M., Crabbé, B., and Denis, P. (2010a). Statistical french dependency parsing: Treebank conversion and first results. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).

Candito, M., Nivre, J., Denis, P., and Henestroza Anguiano, E. (2010b). Benchmarking of statistical dependency parsers for french. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 108–116, Stroudsburg, PA, USA. Association for Computational Linguistics.

Candito, M., Nivre, J., Denis, P., and Henestroza Anguiano, E. (2010c). Benchmarking of statistical dependency parsers for french. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 108–116, Stroudsburg, PA, USA. Association for Computational Linguistics.

Candito, M., Crabbé, B., and Falco, M., (2011). *Dépendances syntaxiques de surface pour le français – Schéma d'annotation pour un corpus en dépendances obtenu par conversion du FrenchTreebank*, may.

Carroll, J., Briscoe, T., and Sanfilippo, A. (1998). Parser Evaluation: A Survey and a New Proposal. In *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pages 447–454, Granada, Spain.

De la Clergerie, E. V., Hamon, O., Mostefa, D., Ayache, C., Paroubek, P., and Vilnat, A. (2008). PASSAGE: from French Parser Evaluation to Large Sized Treebank. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 3570–3577, Marrakech, Morocco, may. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.

De Marneffe, M.-C. and Manning, C. D. (2008). The stanford typed dependencies representation. In *Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation.*, CrossParser '08, pages 1–8, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dubois, J. and Dubois-Charlier, F. (1997). *Les verbes français*. Larousse-Bordas.

Lin, D. (1998). A dependency-based method for evaluating broad-coverage parsers. *Natural Language Engineering*, 4(02):97–114, 5.

Nivre, J. (2006). *Inductive Dependency Parsing (Text, Speech and Language Technology).* Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Paroubek, P., Robba, I., Vilnat, A., and Ayache, C. (2006). Data, Annotations and Measures in EASY - the Evaluation Campaign for Parsers of French. In *proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)*, pages 315–320, Genoa, Italy, May. ELRA.

Paroubek, P., De la Clergerie, É., Loiseau, S., Vilnat, A., and Francopoulo, G. (2009). The PASSAGE Syntactic Representation. In *Proceedings of the 7th International Workshop on Treebanks and Linguistic Theories*, pages 91–102, Gröningenm Pays-Bas, January. Netherlands Graduate Schools of Linguistics (LOT).

Tesnière, L. (1959). *Éléments de Syntaxe Structurale.* Éditions Klinksieck, Paris, France.

Vilnat, A., Paroubek, P., de la Clergerie, E. V., Francopoulo, G., and Guénot, M.-L. (2010). PASSAGE Syntactic Representation: a Minimal Common Ground for Evaluation. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta, may. European Language Resources Association (ELRA).