# Large Scale Production of Syntactic Annotations to Move Forward

**Patrick Paroubek, Anne Vilnat, Sylvain Loiseau**
LIMSI-CNRS
BP 133 91403 Orsay Cedex
France
prenom.nom@limsi.fr

**Gil Francopoulo**
Tagmatica
126 rue de Picpus 75012 Paris
France
gil.francopoulo@tagmatica.com

**Olivier Hamon**
ELDA and LIPN-P13
55-57 rue Brillat-Savarin 75013 Paris,
France
hamon@elda.org

**Eric Villemonte de la Clergerie**
Alpage-INRIA
Dom. de Voluceau Rocquencourt,
B.P. 105, 78153 Le Chesnay, France
Eric.De_La_Clergerie@inria.fr

## Abstract

This article presents the methodology of the PASSAGE project, aiming at syntactically annotating large corpora by composing annotations. It introduces the annotation format and the syntactic annotation specifications. It describes an important component of the methodolgy, namely an WEB-based evaluation service, deployed in the context of the first PASSAGE parser evaluation campaign.

## 1 Introduction

The last decade has seen, at the international level, the emergence of a very strong trend of researches on statistical methods in Natural Language Processing. In our opinion, one of its origins, in particular for English, is the availability of large annotated corpora, such as the Penn Treebank (1M words extracted from the Wall Street journal, with syntactic annotations; 2nd release in 1995[1], the British National Corpus (100M words covering various styles annotated with parts of speech[2]), or the Brown Corpus (1M words with morpho-syntactic annotations). Such annotated corpora were very valuable to extract stochastic grammars or to parametrize disambiguation algorithms. For instance (Miyao et al., 2004) report an experiment where an HPSG grammar is semi-automatically aquired from the Penn Treebank, by first annotating the treebank with partially specified derivation trees using heuristic rules , then by extracting lexical entries with the application of inverse grammar rules. (Cahill et al., 2004) managed to extract LFG subcategorisation frames and paths linking long distance dependencies reentrancies from f-structures generated automatically for the Penn-II treebank trees and used them in an long distance dependency resolution algorithm to parse new text. They achieved around 80% f-score for fstructures parsing on the WSJ part of the Penn-II treebank, a score comparable to the ones of the state-of-the-art hand-crafted grammars. With similar results, (Hockenmaier and Steedman, 2007) translated the Penn Treebank into a corpus of Combinatory Categorial Grammar (CCG) derivations augmented with local and long-range word to word dependencies and used it to train wide-coverage statistical parsers. The development of the Penn Treebank have led to many similar proposals of corpus annotations[3]. However, the development of such treebanks is very costly from an human point of view and represents a long standing effort, in particular for getting of rid of the annotation errors or inconsistencies, unavoidable for any kind of human annotation. Despite the growing number of annotated corpora, the volume of data that can be manually annotated remains limited thus restricting the experiments that can be tried on automatic grammar acquisition. Furthermore, designing an annotated corpus involves choices that may block future experiments from acquiring new kinds of linguistic knowledge because they necessitate annotation incompatible or difficult to produce from the existing ones.

With PASSAGE (de la Clergerie et al., 2008b), we believe that a new option becomes possible. Funded by the French ANR program on Data

[1] http://www.cis.upenn.edu/~treebank/
[2] http://www.natcorp.ox.ac.uk/
[3] http://www.ims.uni-stuttgart.de/projekte/TIGER/related/links.shtml

Warehouses and Knowledge, PASSAGE is a 3-year project (2007–2009), coordinated by INRIA project-team Alpage. It builds up on the results of the EASy French parsing evaluation campaign, funded by the French Technolangue program, which has shown that French parsing systems are now available, ranging from shallow to deep parsing. Some of these systems were neither based on statistics, nor extracted from a treebank. While needing to be improved in robustness, coverage, and accuracy, these systems has nevertheless proved the feasibility to parse medium amount of data (1M words). Preliminary experiments made by some of the participants with deep parsers (Sagot and Boullier, 2006) indicate that processing more than 10 M words is not a problem, especially by relying on clusters of machines. These figures can even be increased for shallow parsers. In other words, there now exists several French parsing systems that could parse (and re-parse if needed) large corpora between 10 to 100 M words.

Passage aims at pursuing and extending the line of research initiated by the EASy campaign by using jointly 10 of the parsing systems that have participated to EASy. They will be used to parse and re-parse a French corpus of more than 100 M words along the following feedback loop between parsing and resource creation as follows (de la Clergerie et al., 2008a):

1. Parsing creates syntactic annotations;
2. Syntactic annotations create or enrich linguistic resources such as lexicons, grammars or annotated corpora;
3. Linguistic resources created or enriched on the basis of the syntactic annotations are then integrated into the existing parsers;
4. The enriched parsers are used to create richer (e.g., syntactico-semantic) annotations;
5. etc. going back to step 1

In order to improve the set of parameters of the parse combination algorithm (inspired from the Recognizer Output Voting Error Reduction, i.e. ROVER, experiments), two parsing evaluation campaigns are planned during PASSAGE, the first of these already took place at the end of 2007 (de la Clergerie et al., 2008b). In the following, we present the annotation format specification and the syntactic annotation specifications of PASSAGE, then give an account of how the syntactic annotations were compared in the first evaluation campaign, by first describing the evaluation metrics and the web server infrastructure that was deployed to process them. We conclude by showing how the results so far achieved in PASSAGE will contribute to the second part of the project, extracting and refining enriched linguistic annotations.

## 2 PASSAGE Annotation Format

The aim is to allow an explicit representation of syntactic annotations for French, whether such annotations come from human annotators or parsers. The representation format is intended to be used both in the evaluation of different parsers, so the parses' representations should be easily comparable, and in the construction of a large scale annotation treebank which requires that all French constructions can be represented with enough details.

The format is based on three distinct specifications and requirements:

1. MAF (ISO 24611)[4] and SynAF (ISO 24615)[5] which are the ISO TC37 specifications for morpho-syntactic and syntactic annotation (Ide and Romary, 2002) (Declerck, 2006) (Francopoulo, 2008). Let us note that these specifications cannot be called "standards" because they are work in progress and these documents do not yet have the status Published Standard. Currently, their official status is only Committee Draft.
2. The format used during the previous TECHNOLANGUE/EASY evaluation campaign in order to minimize porting effort for the existing tools and corpora.
3. The degree of legibility of the XML tagging.

From a technical point of view, the format is a compromise between "standoff" and "embedded" notation. The fine grain level of tokens and words is standoff (wrt the primary document) but higher levels use embedded annotations. A standoff notation is usually considered more powerful but less

[4] http://lirics.loria.fr/doc_pub/maf.pdf
[5] http://lirics.loria.fr/doc_pub/N421_SynAF_CD_ISO_24615.pdf

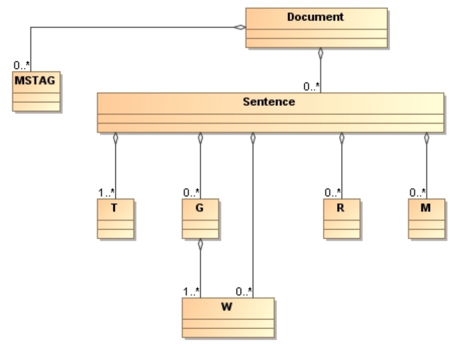


Figure 1: UML diagram of the structure of an annotated document

readable and not needed when the annotations follow a (unambiguous) tree-like structure. Let us add that, at all levels, great care has been taken to ensure that the format is mappable onto MAF and SynAF, which are basically standoff notations.

The structure of a PASSAGE annotated document may be summarized with the UML diagram in Figure1. The document begins by the declaration of all the morpho-syntactic tagsets (MSTAG) that will be used within the document. These declarations respect the ISO Standard Feature Structure Representation (ISO 24610-1). Then, tokens are declared. They are the smallest unit addressable by other annotations. A token is unsplittable and holds an identifier, a character range, and a content made of the original character string. A word form is an element referencing one or several tokens. It has has two mandatory attributes: an identifier and a list of tokens. Some optional attributes are allowed like a part of speech, a lemma, an inflected form (possibly after spelling correction or case normalization) and morpho-syntactic tags. The following XML fragment shows how the original fragment *"Les chaises"* can be represented with all the optional attributes offered by the PASSAGE annotation format :

```
<T id="t0" start="0" end="3">
   Les
</T>
<W id="w0" tokens="t0"
   pos="definiteArticle"
   lemma="le"
   form="les"
   mstag="nP"/>
<T id="t1" start="4" end="11">
   chaises
</T>
<W id="w1" tokens="t1"
   pos="commonNoun"
   lemma="chaise"
   form="chaises"
   mstag="nP gF"/>
```

Note that all parts of speech are taken from the ISO registry [6] (Francopoulo et al., 2008). As in MAF, a word may refer to several tokens in order to represent multi-word units like *"pomme de terre"*. Conversely, a unique token may be refered by two different words in order to represent results of split based spelling correction like when *"unetable"* is smartly separated into the words *"une"* and *"table"*. The same configuration is required to represent correctly agglutination in fused prepositions like the token *"au"* that may be rewritten into the sequence of two words *"à" "le"*. On the contrary of MAF, cross-reference in token-word links for discontiguous spans is not allowed for the sake of simplicity. Let us add that one of our requirement is to have PASSAGE annotations mappable onto the MAF model and not to map all MAF annotations onto PASSAGE model. A **G** element denotes a syntactic group or a constituent (see details in section 3). It may be recursive or non-recursive and has an identifier, a type, and a content made of word forms or groups, if recursive. All group type values are taken from the ISO registry. Here is an example :

```
<T id="t0" start="0" end="3">
   Les
</T>
<T id="t1" start="4" end="11">
   chaises
</T>
<G id="g0" type="GN">
   <W id="w0" tokens="t0"/>
   <W id="w1" tokens="t1"/>
</G>
```

A group may also hold optional attributes like syntactic tagsets of MSTAG type. The syntactic relations are represented with a standoff annotations which refer to groups and word forms. A relation is defined by an identifier, a type, a source, and a target (see details in section 3. All relation types, like "subject" or "direct object" are mappable onto

[6]Data Category Registry, see http://syntax.inist.fr

the ISO registry. An unrestricted number of comments may be added to any element by means of the mark element (i.e. M). Finally, a “Sentence” element gathers tokens, word forms, groups, relations and marks and all sentences are included inside a “Document” element.

## 3 PASSAGE Syntactic Annotation Specification

### 3.1 Introduction

The annotation formalism used in PASSAGE[7] is based on the EASY one(Vilnat et al., 2004) which whose first version was crafted in an experimental project PEAS (Gendner et al., 2003), with inspiration taken from the propositions of (Carroll et al., 2002). The definition has been completed with the input of all the actors involved in the EASY evaluation campaign (both parsers' developers and corpus providers) and refined with the input of PASSAGE participants. This formalism aims at making possible the comparison of all kinds of syntactic annotation (shallow or deep parsing, complete or partial analysis), without giving any advantage to any particular approach. It has six kinds of syntactic “chunks”, we call constituents and 14 kinds of relations The annotation formalism allows the annotation of minimal, continuous and non recursive constituents, as well as the encoding of relations wich represent syntactic functions. These relations (all of them being binary, except for the ternary coordination) have sources and targets which may be either forms or constituents (grouping several forms). Note that the PASSAGE annotation formalism does not postulate any explicit lexical *head*.

### 3.2 Constituent annotations

For the PASSAGE campaigns, 6 kinds of constituents (syntactic “chunks”) have been considered and are illustrated in Table 3.2:

- the Noun Phrase (GN for *Groupe Nominal*) may be made of a noun preceded by a determiner and/or by an adjective with its own modifiers, a proper noun or a pronoun;
- the prepositional phrase (GP, for *groupe prépositionnel* ) may be made of a preposition and the GN it introduces, a contracted determiner and preposition, followed by the introduced GN, a preposition followed by an adverb or a relative pronoun replacing a GP;
- the verb kernel (**NV** for *noyau verbal* ) includes a verb, the clitic pronouns and possible particles attached to it. Verb kernels may have different forms: conjugated tense, present or past participle, or infinitive. When the conjugation produces compound forms, distinct NVs are identified;
- the adjective phrase (**GA** for *groupe adjectival*) contains an adjective when it is not placed before the noun, or past or present participles when they are used as adjectives;
- the adverb phrase (**GR** for *groupe adverbial* ) contains an adverb;
- the verb phrase introduced by a preposition (**PV**) is a verb kernel with a verb not inflected (infinitive, present participle,...), introduced by a preposition. Some modifiers or adverbs may also be included in PVs.

#### 3.2.1 Syntactic Relation annotations

The dependencies establish all the links between the minimal constituents described above. All participants, corpus providers and campaign organizers agreed on a list of 14 kinds of dependencies listed below:

1. subject-verb (**SUJ_V**): may be inside the same NV as between *elle* and *était* in *elle était* (*she was*), or between a GN and a NV as between *mademoiselle* and *appelait* in *Mademoiselle appelait* (*Miss was calling*);
2. auxiliary-verb (**AUX_V**), between two NVs as between *a* and *construit* in: *on a construit une maison* (*we have built a house*);
3. direct object-verb (**COD_V**): the relation is annotated between a main verb (NV) and a noun phrase (GN), as between *construit* and *la première automobile* in: *on a construit la première automobile* (*we have built the first car*);
4. complement-verb (**CPL_V**): to link to the verb the complements expressed as GP or PV

[7] Annotation guide: `http://www.limsi.fr/Recherche/CORVAL/PASSAGE/eval_1/2007_10_05PEAS_reference_annotations_v11.12.html`

| | |
|---|---|
| GN | - la très grande porte[8]<br>(*the very big door*);<br>- Rouletabille<br>- eux (*they*), qui (*who*) |
| GP | - de la chambre (*from the bedroom*),<br>- du pavillon (*from the lodge*)<br>- de là (*from there*), dont (*whose*) |
| NV | - j'entendais (*I heared*)<br>- [on ne l'entendait][9] plus<br>(*we could no more hear her*)<br>- Jean [viendra] (*Jean will come*)<br>- [désobéissant] à leurs parents<br>(*disobeying their parents*),<br>- [fermée] à clef (*key closed*)<br>- Il [ne veut] pas [venir]<br>(*He doesn't want to come*),<br>- [ils n'étaient] pas [fermés]<br>(*they were not closed*), |
| GA | - les barreaux [intacts] (*the intact bars*)<br>- la solution [retenue] fut...<br>(*the chosen solution has been...*),<br>- les enfants [désobéissants]<br>(*the disobeying children*) |
| GR | - aussi (*also*)<br>- vous n'auriez [pas] (*you would not*) |
| PV | - [pour aller] à Paris (*for going to Paris*),<br>- de vraiment bouger (*to really move*) |

Table 1: Constituent examples

which may be adjuncts or indirect objects, as between *en quelle année* and *construit* in *en quelle année a-t on construit la première automobile* (*In which year did we build the first car*);

5. modifier-verb (**MOD_V**): concerns the constituants which certainly modify the verb, and are not mandatory, as adverbs or adjunct clauses, as between *profondément* or *quand la nuit tombe* and *dort* in *Jean dort profondément quand la nuit tombe* (*Jean deeply sleeps when the night falls*);

6. complementor (**COMP**): to link the introducer and the verb kernel of a subordinate clause, as between *qu'* and *viendra* in *Je pense qu'il viendra* (*I think that he will come*); it is also used to link a preposition and a noun phrase when they are not contiguous, preventing us to annotate them as GP;

7. attribute-subject/object (**ATB_SO**): between the attribute and the verb kernel, and precising that the attribute is relative to (a) the subject as between *grand* and *est* in *il est grand* ), or (b) the object as between *étrange* and *trouve* in *il trouve cette explication étrange*;

8. modifier-noun (**MOD_N**): to link to the noun all the constituents which modify it, as the adjective, the genitive, the relative clause... This dependency is annotated between *unique* and *fenêtre* in *l'unique fenêtre* (*the unique window*) or between *de la chambre* and *la porte* in *la porte de la chambre* (*the bedroom door*);

9. modifier-adjective (**MOD_A**): to relate to the adjective the constituents which modify it, as between *très* et *belle* in *¡la très belle collection* (*the very impressive collection*) or between *de son fils* and *fière* in *elle est fière de son fils* (*she is proud of her son*);

10. modifier-adverb (**MOD_R**): the same kind of dependency than MOD_A for the adverbs, as between *très* and *gentiment* in *elle vient très gentiment* (*she comes very kindly*);

11. modifier-preposition (**MOD_P**): to relate to a preposition what modifies it, as between *peu* and *avant* in *elle vient peu avant lui* (*she comes just before him*);

12. coordination (**COORD**): to relate the coordinate and the coordinated elements, as between *Pierre*, *Paul* and *et* in *Pierre et Paul arrivent* (*Paul and Pierre are arriving*);

13. apposition (**APP**): to link the elements which are placed side by side, when they refer to the same object, as between *le député* and *Yves Tavernier* in *Le député Yves Tavernier ...* (*the Deputy Yves Tavernier...*);

14. juxtaposition (**JUXT**): to link constituents which are neither coordinate nor in an apposition relation, as in enumeration. It also links clauses as *on ne l'entendait* et *elle était* in *on ne l' entendait plus ... elle était peut-être morte* (*we did not hear her any more... perhaps she was dead*).

Some dependencies are illustrated in the two annotated sentences illutrated in figure . These anno-

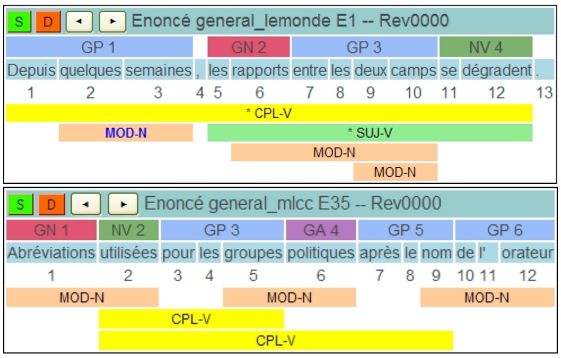


Figure 2: Example of two sentences annotations

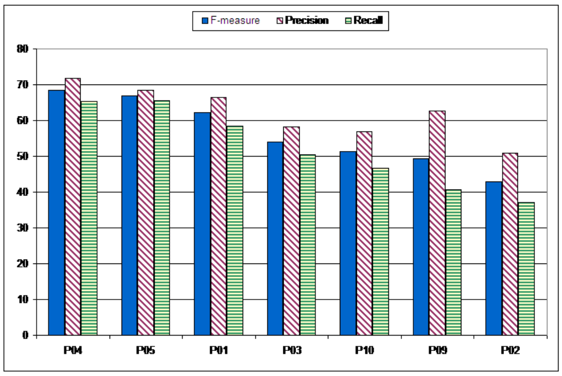


Figure 3: Overall functional relations results

tations have been made using EasyRef, a specific Web annotation tool developed by INRIA.

## 4 PASSAGE First Evaluation Campaign

### 4.1 Evalution Service

The first PASSAGE evaluation campaign was carried out in two steps. During the initial one-month development phase, a development corpus was used to improve the quality of parsers. This development corpus from the TECHNOLANGUE/EASY is composed of 40,000 sentences, out of which 4,000 sentences have been manually annotated for the gold standard. Based on these annotated sentences, an automatic WEB-based evaluation server provides fast performance feedback to the parsers' developers. At the end of this first phase, each participant indicated what he thought was his best parser run and got evaluated on a new set of 400 sentences selected from another part of the developement corpus which meanwhile had been manually annotated for the purpose and kept undisclosed.

The two phases represent a strong effort for the evaluators. To avoid adding the cost of managing the distribution and installation of the evaluation package at each developer's site, the solution of the WEB evaluation service was chosen. A few infrastructures have been already experimented in NLP, like GATE (Cunningham et al., 2002) infrastructures, but to our knowledge none has been used to provide an WEB-based evaluation service as PASSAGE did. The server was designed to manage two categories of users: parser developers and organizers. To the developers, it provides, almost in real time, confidential and secure access to the automatic evaluation of their submitted parses. To the organizers, it give access to statistics enabling them to follow the progress made by the developers, and easy management of the test phase. The evaluation server provides, through a simple WEB browser, access to both coarse and fine grain statistics to a developer's performance evaluation, globally for the whole corpus, at the level of a particular syntactic annotation or of a particular genre specific subcorpus, and also at the level of a single annotation for a particular word form.

### 4.2 Performance Results

Ten systems participated to the constituents annotation task. For most of the systems, F-measure is up to 90% and only three systems are between 80% and 90%. The trend is quite the same for Recall and Precision. Around 96.5% of the constituents returned by the best system are correct and it found 95.5% of the constituents present in gold standard. Figure 3 shows the results of the seven systems that participated to the functional relations annotation task. Performance is lower than for constituents and differences between systems are larger, an evidence that the task remains more difficult. No systems gets a performance above 70% in F-measure, three are above 60% and two above 50%. The last two systems are above 40%.

### 4.3 Systems Improvements

The higher system gets increasing results from the beginning of the development phase to the test phase for both constituents and relations. However, although the increase for relations is rather continuous, constituents results grow during the first few development evaluations, then reach a threshold from which results do not vary. This can be explained by the fact that the constituent scores are rather high, while for relations, scores are lower and starting from low scores.

Using the evaluation server, system improves its performance by 50% for the constituents and 600% for the relations, although performance vary according to the type of relation or constituent. Moreover, in repeating development evaluations, another consequence was the convergence of precision and recall.

## 5 Parser's outputs combination

The idea to combine the output of systems participating to an evalauation campaign in order to obtain a combination with better performance than the best one was invented to our knowledge by J. Fiscus (Fiscus, 1997) in a DARPA/NIST speech recognition evaluation (ROVER/Reduced Output Voting Error Reduction). By aligning the output of the participating speech transcription systems and by selecting the hypothesis which was proposed by the majority of the systems, he obtained better performances than these of the best system. The idea gained support in the speech processing community(Lööf et al., 2007) and in general better results are obtained with keeping only the output of the two or three best performing systems, in which case the relative improvement can go up to 20% with respect to the best performance (Schwenk and Gauvain, 2000). For text processing, the ROVER procedure was applied to POS tagging (Paroubek, 2000) and machine translation (Matusov et al., 2006).

In our case, we will use the text itself to realign the annotations provided by the various parser before computing their combination, as we did for our first experiments with the EASY evaluation campaign data (Paroubek et al., 2008). Since it is very likely taht the different parsers do not use the same word and sentence segmentation, we will realign all the data along a common word and sentence segmentation obtained by majority vote from the different outputs.

But our motivation for using such procedure is not only concerned with performance improvement but also with the obtention of a confidence measure for the annotation since if all systems agree on a particular annotation, then it is very likely to be true.

At this stage many options are open for the way we want to apply the ROVER algorithm, since we have both constituents and relations in our annotations. We could vary the selection order (between constituents and relations), or use different comparison functions for the sources/targets of constituents/relations(Patrick Paroubek, 2006), or perform incremental/global merging of the annoations, or explore different weightings/thresholding strategies etc. In passage, ROVER experiments are only beginning and we have yet to determine which is the best strategy before applying it to word and sentence free segmentation data. In the early experiment we did with the "EASy classic" PASSAGE track which uses a fixed word and sentence segmentation, we measured an improvement in precision for some specific subcorpora and annotations but improvement in recall was harder to get.

## 6 Conclusion

The definition of a common interchange syntactic annotation format is an essential element of any methodology aiming at the creation of large annotated corpora from the cooperation of parsing systems to acquire new linguistic knowledge. But the formalism aquires all of its value when backed-up by the deployment of a WEB-based evaluation service as the PASSAGE examples shows. 167 experiments were carried out during the development phase (around 17 experiments per participant in one month). The results of the test phase were available less than one hour after the end of the development phase. The service proved so successful that the participants asked after the evaluation, that the evaluation service be extended to support evaluation as a perennial service

## References

Cahill, Aoife, Michael Burke, Ruth O'Donovan, Josef Van Genabith, and Andy Way. 2004. Long-distance dependency resolution in automatically acquired wide-coverage pcfg-based lfg approximations. In *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume*, pages 319–326, Barcelona, Spain, July.

Carroll, J., D. Lin, D. Prescher, and H. Uszkoreit. 2002. Proceedings of the workshop beyond parseval - toward improved evaluation measures for parsing systems. In *Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC)*, Las Palmas, Spain.

Cunningham, Hamish, Diana Maynard, Kalina Bontcheva, and Valentin Tablan. 2002. Gate: an architecture for development of robust hlt applications. In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 168–175, Morristown, NJ, USA. Association for Computational Linguistics.

Declerck, T. 2006. Synaf: towards a standard for syntactic annotation. In *In proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)*, Genoa, Italy, May. ELRA.

Fiscus, Jonathan G. 1997. A post-processing system to yield reduced word error rates: recognizer output voting error reduction (rover). In *In proceedings of the IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 347–357, Santa Barbara, CA.

Francopoulo, G., T. Declerck, V. Sornlertlamvanich, E. de la Clergerie, and M. Monachini. 2008. Data category registry: Morpho-syntactic and syntactic profiles. Marrakech. LREC.

Francopoulo, Gil. 2008. Tagparser: Well on the way to iso-tc37 conformance. In *In proceedings of the International Conference on Global Interoperability for Language Resources (ICGL)*, pages 82–88, Hong Kong, January.

Gendner, Véronique, Gabriel Illouz, Michèle Jardino, Laura Monceaux, Patrick Paroubek, Isabelle Robba, and Anne Vilnat. 2003. Peas the first instanciation of a comparative framework for evaluating parsers of french. In *Proceedings of the 10$^{th}$ Conference of the European Chapter fo the Association for Computational Linguistics*, pages 95–98, Budapest, Hungary, April. ACL. Companion Volume.

Hockenmaier, Julia and Mark Steedman. 2007. Ccgbank: A corpus of ccg derivations and dependency structures extracted from the penn treebank. *Computational Linguistics*, 33(3):355–396.

Ide, N. and L. Romary. 2002. Standards for language ressources. Las Palmas. LREC.

Lööf, J., C. Gollan, S. Hahn, G. Heigold, B. Hoffmeister, C. Plahl, D. Rybach, R. Schlüter, , and H. Ney. 2007. The rwth 2007 tc-star evaluation system for european english and spanish. In *In proceedings of the Interspeech Conference*, pages 2145–2148.

Matusov, Evgeny, N. Ueffing, and Herman Ney. 2006. Automatic sentence segmentation and punctuation prediction for spoken language translation. In *Proceedings of the International Workshop on Spoken Language Translation (IWSLT)*, pages 158–165, Trento, Italy.

de la Clergerie, Eric, Christelle Ayache, Gaël de Chalendar, Gil Francopoulo, Claire Gardent, and Patrick Paroubek. 2008a. Large scale production of syntactic annotations for french. In *In proceedings of the First Workshop on Automated Syntactic Annotations for Interoperable Language Resources at IGCL'08*, pages 45–52, Hong Kong, January.

de la Clergerie, Eric, Olivier Hamon, Djamel Mostefa, Christelle Ayache, Patrick Paroubek, and Anne Vilnat. 2008b. Passage: from french parser evaluation to large sized treebank. In ELRA, editor, *In proceedings of the sixth international conference on Language Resources and Evaluation (LREC)*, Marrakech, Morroco, May. ELRA.

Miyao, Yusuke, Takashi Ninomiya, and Jun'ichi Tsujii. 2004. Corpus-oriented grammar development for acquiring a head-driven phrase structure grammar from the penn treebank. In *In Proceedings of the First International Joint Conference on Natural Language Processing (IJCNLP-04)*.

Paroubek, Patrick, Isabelle Robba, Anne Vilnat, and Christelle Ayache. 2008. Easy, evaluation of parsers of french: what are the results? In *Proceedings of the 6$^{th}$ International Conference on Language Resources and Evaluation (LREC)*, Marrakech, Morroco.

Paroubek, Patrick. 2000. Language resources as by-product of evaluation: the multitag example. In *In proceedings of the Second International Conference on Language Resources and Evaluation (LREC2000)*, volume 1, pages 151–154.

Patrick Paroubek, Isabelle Robba, Anne Vilnat Christelle Ayache. 2006. Data, annotations and measures in easy - the evaluation campaign for parsers of french. In ELRA, editor, *In proceedings of the fifth international conference on Language Resources and Evaluation (LREC 2006)*, pages 315–320, Genoa, Italy, May. ELRA.

Sagot, Benoît and Pierre Boullier. 2006. Efficient parsing of large corpora with a deep lfg parser. In *In proceedings of the sixth international conference on Language Resources and Evaluation (LREC)*, Genoa, Italy, May. ELDA.

Schwenk, Holger and Jean-Luc Gauvain. 2000. Improved rover using language model information. In *In proceedings of the ISCA ITRW Workshop on Automatic Speech Recognition: Challenges for the new Millenium*, pages 47–52, Paris, September.

Vilnat, A., P. Paroubek, L. Monceaux, I. Robba, V. Gendner, G. Illouz, and M. Jardino. 2004. The ongoing evaluation campaign of syntactic parsing of french: Easy. In *Proceedings of the 4$^{th}$ International Conference on Language Resources and Evaluation (LREC)*, pages 2023–2026, Lisbonne, Portugal.