

## **EASy : Campagne d'évaluation des analyseurs syntaxiques**

Patrick Paroubek (1), Louis-Gabriel Pouillot (2),

Isabelle Robba (1), Anne Vilnat (1)

(1)LIMSI - CNRS

Université d'Orsay, BP 133

91403 Orsay CEDEX

{prénom.nom}@limsi.fr

(2)Evaluations and Language resources Distribution Agency (ELDA)

55-57, rue Brillat Savarin 75013 Paris

pouillot@elda.org

**Mots-clefs :** Évaluation, analyseurs syntaxiques, annotation en syntaxe

**Keywords:** Evaluation, syntactic parsers, syntactic annotation

### **Résumé**

Cet article présente la campagne d'évaluation des analyseurs syntaxiques de français EASy. Nous détaillons la mise en place de la campagne, ainsi que son déroulement jusqu'à ce jour.

### **Abstract**

This paper presents EASy, the evaluation campaign of syntactic parsers of French. We detail the ongoing of the campaign, and its current state.

## **1 Introduction**

Le projet EASY, fait partie de la campagne d'évaluation EVALDA du programme Technolangue<sup>1</sup>, il a débuté en janvier 2003. Son ambition est de mettre au point un protocole complet d'évaluation des analyseurs syntaxiques, allant de la constitution d'un large corpus hétérogène et de l'annotation d'une partie de ce corpus, à l'évaluation des analyseurs participants et à la publication d'un large corpus entièrement annoté et validé (grâce à la fusion des sorties des analyseurs participants). Les réflexions autour du choix de ce qui serait annoté ont été menées avec

---

<sup>1</sup>Le programme Technolangue est à l'initiative du ministère délégué à la Recherche et aux Nouvelles Technologies

les participants et les fournisseurs de corpus, elles nous ont conduit à écrire un guide complet d'annotation disponible en ligne à l'adresse : <http://www.limsi.fr/Recherche/CORVAL/easy>.

Un des aspects importants pour nous en tant qu'organiseurs étaient de n'écarter la participation d'aucun analyseur : tout participant prêt à transcrire les sorties de son analyseur dans le formalisme retenu dans EASy devait avoir la possibilité de soumettre son analyseur pour l'évaluer.

Dans la campagne EASy, cinq fournisseurs se sont engagés à annoter en constituants et en relations 60 000 mots au total. Quatre d'entre eux sont des organismes de recherche : l'ATILF (Université Nancy 2), le DELIC (Université de Provence), le LLF (Université Paris 7) et le STIM-APHP (Paris). Le cinquième est ELDA (Agence pour l'évaluation et la distribution des ressources linguistiques) qui est co-organisateur avec le LIMSI (groupe LIR) de la campagne EASy.

Le DELIC est le seul fournisseur de corpus qui présente un article dans cet atelier. Ils ont eu la difficile tâche d'annoter des retranscriptions de l'oral, qui posent comme on le sait des problèmes spécifiques et délicats à résoudre, en particulier pour ce qui concerne la segmentation en mots et en énoncés.

Au moment de l'évaluation, treize équipes étaient à même de participer, certaines soumettant plusieurs jeux de sorties. Au final, nous avons recueilli seize ensembles de résultats. Parmi ces équipes on compte neuf laboratoires ou organismes de recherche et développement et quatre entreprises privées. Dans la liste qui suit nous avons marqué d'une étoile les équipes ayant soumis un article dans les actes de cet atelier : le CEA-LIST\*, l'ERSS\*, France Telecom R&D\*, le GREYC\*, l'INRIA-Rocquencourt (équipe ATOLL)\*, le LATL (Université de Genève)\*, le LIRMM, le LORIA\*, le LPL\*, PERTIMM, Synapse Développement, Tagmatica\*, Xerox Research Centre Europe.

La suite de l'article expose en détail la campagne EASy étape par étape et se termine par une présentation de quelques travaux en annotation de corpus et en évaluation d'analyseurs syntaxiques.

## **2 Présentation de EASy**

La campagne EASy a fait l'objet d'une étude préalable (Gendner, 2003) au sein du groupe LIR du LIMSI, elle s'est déroulée sur une période de 2 ans et a permis de proposer une infrastructure contenant les éléments suivants :

1. La constitution d'un corpus d'1 million de mots, composé de textes hétérogènes en genre : des articles de journaux, des extraits de romans, des recueils de questions, des transcriptions de l'oral, des extraits de sites Internet. Ce point est détaillé section 3.
2. Le formatage du corpus complet : normalisation, tokenisation et découpage en phrases.
3. L'annotation manuelle d'un sous-ensemble de 60 000 mots pour servir de référence. Elle est effectuée à l'aide d'un éditeur HTML, et les résultats sont transcrits dans un format XML. Les principes de cette annotation sont précisés section 4.
4. L'analyse par les participants du corpus complet et la transcription des sorties de leur analyseur dans un format XML commun.

5. L'évaluation qui consiste essentiellement en un calcul du rappel et de la précision sur les constituants et sur les relations. Cette étape est en cours de réalisation, comme nous le verrons section 5.

Les décisions concernant le formalisme d'annotation ont été l'objet de nombreuses discussions avec les participants et les fournisseurs de corpus (ils annotent de façon semi-automatique ou manuelle les corpus et doivent donc, dans la mesure du possible, être en accord avec les choix portant sur l'annotation).

Par ailleurs, le formalisme doit permettre une couverture la plus large possible des phénomènes syntaxiques de la langue. Nous avons fait le choix dans EASy d'annoter deux types d'information : les constituants, dont on annote la catégorie et l'étendue et les relations syntaxiques ou relations fonctionnelles, pour lesquelles on indique les éléments en relation, éléments qui peuvent être des mots ou des constituants. Dans ce formalisme, les constituants ne peuvent être discontinus ou imbriqués et ils sont de la taille la plus petite possible.

### **3 La constitution du corpus**

Le corpus de test de la campagne est constitué d'un million de mots issus de sources diverses, provenant de cinq fournisseurs. Ceux-ci ont produit le corpus de test ainsi que les annotations d'une partie de ce corpus qui sert de référence pour les évaluations des systèmes. Ce corpus est réparti de la façon suivante:

- ATILF : corpus de textes littéraires - 150 000 mots fournis dont 15 000 mots annotés, avec des textes de Michelet, Coppée, Theuriet et Sandeau ;
- DELIC
  - 10 fragments de dialogues transcrits extraits du Corpus du Français Parlé ; 8 000 mots fournis et annotés ; thèmes divers ;
  - 2000 courriers électroniques personnels anonymisés ; 114 000 mots fournis ; correspondance privée ;
- ELDA
  - corpus de questions : 137 000 mots dont 5 000 annotés provenant de questions de la campagne TREC traduites et de questions extraites des notices bibliographiques de la campagne AMARYLLIS ;
  - 250 courriers électroniques anonymisés : 7.000 mots fournis et annotés provenant de correspondances privées ;
  - extraits du journal Le Monde, de rapports du Sénat et de l'assemblée européenne (MLCC) : 235 000 mots dont 9 000 annotés ;
- LLF (Paris 7) : extraits du corpus Le Monde de 1992 - 15 000 mots annotés provenant d'articles généraux et économiques ;
- STIM-APHP : corpus de textes médicaux ; 100 000 mots fournis dont 5 000 annotés ; textes sur les maladie d'Alzeihmer, Parkinson, hépatite C, rapports et comptes rendus de conférences ;

La diversité des sources a pour objectif de tester les systèmes sur des structures syntaxiques très variées et sur des corpus ayant des spécificités propres. L'idée est d'évaluer sur les particularités de chaque type de corpus et non uniquement sur des phénomènes linguistiques spécifiques. Les corpus littéraires et journalistiques comportent des phrases très structurées et des repères tels que la ponctuation ou la typographie peuvent entrer dans le cadre des traitements. En revanche pour le corpus de courriers électroniques ou les transcriptions orales, la ponctuation est souvent, voire toujours, absente, la typographie (emploi de majuscules par exemple) aléatoire, l'orthographe souvent très relâchée et la structure des phrases mise à mal (Adda-Decker et al., 2003).

Afin de permettre à chaque participant de choisir le format de corpus le plus adapté à son système, le corpus a été fourni en plusieurs formats: texte brut normalisé, texte avec la séparation d'énoncé tokenisé et non tokenisé, texte étiqueté avec la version française de l'étiqueteur de Brill (Brunet 2000) et le jeu d'étiquettes utilisé dans Grace (Adda et al., 1999). Afin de permettre une évaluation cohérente, le découpage en énoncés fourni devait être pris pour référence. Ce découpage en énoncés a été réalisé automatiquement avec des outils développés pour la campagne (et précisé ci-dessous) à l'exception des transcriptions de l'oral. En effet ce corpus ne dispose pas de repères suffisants pour un découpage cohérent. Ce découpage a donc été réalisé à la main, une fois les annotations syntaxiques effectuées.

## 4 L'annotation de la référence

### 4.1 Le découpage en phrases et en mots

La première étape de l'annotation a consisté à découper le corpus de référence en phrases et en mots. Ces deux découpages sont souvent considérés comme des problèmes simples et quasiment résolus. Pourtant, nous avons pu constater que lorsque l'on s'intéresse à des textes réels, ce n'était pas toujours le cas ! Nous ne détaillerons pas cet aspect ici, mais nous donnerons quelques exemples. Pour le découpage en phrases, le problème est évidemment rendu plus difficile quand les textes analysés sont des transcriptions d'oral, spontané ou même lu. La présence de ponctuations fortes telles que le point est soit restituée lors de la transcription, soit complètement absente. Il faut alors se fonder sur des critères tels que la durée des pauses pour essayer de réintroduire la coupure en phrases. Les textes écrits posent également un certain nombre de cas difficiles à résoudre : par exemple lorsque le texte présente de nombreuses énumérations, sous forme de "listes à puces", ou du discours rapporté, comme illustré dans les deux exemples suivants (comme l'a également montré Núria Gala Pavia (Gala Pavia, 2003) dans sa thèse).

1. *Pour brancher l'appareil, vous devez :*

- *vérifier votre installation électrique. Si vous ne respectez pas les normes, votre garantie ne fera plus effet.*
- *relier le cordon d'alimentation à votre appareil...*

2. *Le directeur affirma : "Je ne peux pas accepter une telle situation.", devant le conseil d'administration de son établissement.*

Nous avons choisi de considérer la phrase la plus longue possible, pour éviter des découpages qui risqueraient de séparer des constituants, ou d'établir des relations en franchissant la frontière

de phrase, comme dans l'exemple 2 où la fin de la phrase se rapporte à la toute première partie, avant le discours rapporté.

Concernant le découpage en mots, nous avons choisi de constituer une liste de mots composés ou de locutions qui ne formeront qu'un seul mot lors de l'annotation en constituants. Exemple :

*Dès que le soleil se lève, les coqs chantent.*

*Dès que* ne forme qu'un seul mot. Ce découpage s'est avéré peu satisfaisant, les choix étant difficiles à faire de façon cohérente : ce découpage nécessitera un réalignement pour traiter les données retournées par les participants (voir section).

## 4.2 L'annotation en constituants

Le principe général que nous avons adopté consiste à annoter des constituants minimaux et non récursifs. Ce choix est dicté par le fait que nous souhaitons proposer un cadre qui permette d'évaluer des analyseurs ayant des caractéristiques diverses, en essayant d'être le plus équitable possible envers chacun d'eux. Prenons un exemple simple :

*Le chat de la voisine*

On peut analyser cette phrase comme un groupe nominal complexe, constitué d'un groupe nominal (*le chat*) et d'un groupe et d'un groupe prépositionnel (*de la voisine*), ce dernier étant une extension (de type complément de nom) du premier. On peut aussi n'annoter que les deux constituants simples, le rattachement du groupe prépositionnel au groupe nominal étant alors noté par l'intermédiaire d'une relation de type *modifieur du nom*. C'est cette dernière solution que nous avons adoptée. Elle nous permet de ne pas rejeter les *chunkers* qui ne relèvent que les constituants simples et de mieux noter les analyseurs plus précis où ces deux informations seront retrouvées. Ce parti pris nous permet de n'avoir ni constituant récursif, ni constituant discontinu, ce qui simplifie la tâche lors de l'annotation. Partant d'exemples réels, et de la littérature sur le domaine, nous avons déterminé une liste de six constituants. Nous en donnerons une première définition, que nous illustrerons sur des exemples simples<sup>2</sup>.

- le groupe nominal (GN) : il est constitué d'un nom éventuellement précédé d'un déterminant (<GN> *la porte* </GN>) et/ou d'un adjectif antéposé accompagné de ses modifieurs (<GN> *la très grande porte* </GN>), d'un nom propre (<GN> *Rouletabille* </GN>) ou d'un pronom non clitique (<GN> *eux* </GN>, <GN> *qui* </GN>).
- le groupe prépositionnel (GP) : il est constitué d'une préposition et du GN qu'elle introduit (<GP> *de la chambre* </GP>) ou d'un déterminant et d'une préposition contractés (du, aux, des) avec le GN introduit (<GP> *du pavillon* </GP>) ou d'une préposition suivie d'un adverbe (<GP> *de là* </GP>), ou de pronoms qui remplacent des GP (<GP> *dont* </GP>, <GP> *où* </GP>),...
- le noyau verbal (NV) : il regroupe un verbe, les pronoms clitiques plus éventuellement les particules euphoniques (*-t-* et *l'*) et la particule *ne* qui lui sont rattachés (<NV> *j'entendais* </NV>, <NV> *on ne l'entendait* </NV> *plus*). Un noyau verbal peut être à différents modes : temps conjugués mais aussi participe présent (<NV> *désobéissant* </NV> *à leurs parents*), participe passé (<NV> *fermée* </NV> *à clef*) et infinitif (<NV> *ne veut*

<sup>2</sup>Nous indiquerons alors la délimitation d'un constituant de type GX en adoptant une notation à la XML : <GX> *le constituant de type GX* </GX>.

</NV> *pas* <NV> *venir* </NV>). En cas de temps composés, nous identifions un NV distinct pour chaque verbe (<NV> *ils n'étaient* </NV> *pas* <NV> *fermés* </NV>).

- le groupe adjectival (GA) : il contient un adjectif lorsqu'il n'est pas épithète antéposé au nom (*les barreaux* <GA> *intacts* </GA>) ou un participe passé (*la solution* <GA> *retenue* </GA> *fut...*) ou présent (*les enfants* <GA> *désobéissants* </GA>) employé comme adjectif.
- le groupe adverbial (GR) : il contient un adverbe, à l'exception du *ne* qui fait partie du NV (<GR> *aussi* </GR>, <GR> *encore* </GR> *vous n'auriez* <GR> *pas* </GR>).
- le groupe verbal introduit par une préposition (PV) : il correspond à un noyau verbal dont le verbe n'est pas conjugué (infinitif, participe présent, ...) et introduit par une préposition (<PV> *d'ébranler* </PV>). Il peut contenir aussi des modificateurs de ce verbe, comme des adverbes (<PV> *de vraiment bouger* </PV>).

Nous pouvons résumer ces différents exemples sur cet extrait du *Mystère de la chambre jaune* :

<GN> *la porte* </GN> <GP> *de la chambre* </GP> <NV> *fermée* </NV> <GP> *à clef* </GP> <GP> *à l'intérieur* </GP> , <GN> *les volets* </GN> <GP> *de l'unique fenêtre* </GP> <NV> *fermés* </NV> , <GN> *eux* </GN> <GR> *aussi* </GR> , <GP> *à l'intérieur* </GP> , *et* , <GP> *par-dessus les volets* </GP> , <GN> *les barreaux* </GN> <GA> *intacts* </GA> ... *et* <GN> *mademoiselle* </GN> <GN> *qui* </GN> <NV> *appelait* </NV> <GP> *au secours* </GP> ! ... *ou* <GR> *plutôt* </GR> <GR> *non* </GR> , <NV> *on ne l'entendait* </NV> <GR> *plus* </GR> ... <NV> *elle était* </NV> <GR> *peut-être* </GR> <GA> *morte* </GA> ... *mais* <NV> *j'entendais* </NV> <GR> *encore* </GR> , <GP> *au fond* </GP> <GP> *du pavillon* </GP> , <GN> *monsieur* </GN> <GN> *qui* </GN> <NV> *essayait* </NV> <PV> *d'ébranler* </PV> <GN> *la porte* </GN>

Ce découpage semble assez simple à respecter, mais lors de l'annotation, nous avons pu constater que même pour les constituants qui paraissent simples comme le groupe nominal, la délimitation n'est pas toujours facile à affectuer.

### 4.3 Annotation en relation

L'annotation en relation va permettre d'établir tous les liens entre les constituants minimaux décrits ci-dessus. Après de nombreuses concertations avec les participants et les fournisseurs de corpus, nous avons retenu une liste de 14 relations, que nous détaillons ci-dessous, en commençant par les relations les plus simples.

- sujet-verbe (SUJ\_V) : entre *elle* et *était* dans : <NV> *elle était* </NV>, ou entre *mademoiselle* et *appelait* dans <GN> *mademoiselle* </GN> <NV> *appelait* </NV> ;
- auxiliaire-verbe (AUX\_V) : entre *a* et *construit* dans : <NV> *on a* </NV><NV> *construit* </NV> ;
- complément d'objet direct-verbe (COD\_V) : la relation se note entre le verbe principal et le groupe nominal, comme par exemple entre *construit* et *la première automobile* dans : <NV> *on a*</NV><NV> *construit* </NV> <GN> *la première automobile* </GN>;

- complément-verbe (CPL\_V) : pour lier au verbe les autres compléments exprimés sous forme de GP ou de PV, que ce soit les circonstants ou les compléments indirects, comme par exemple entre *en quelle année* et *construit* dans <GP>En quelle année</GP><NV>a-t on</NV> <NV>construit </NV> <GN> la première automobile </GN> ;
- modifieur-verbe (MOD\_V) : cette relation concerne tous les constituants dont on peut affirmer qu'ils sont modifieurs et non compléments du verbe, comme les adverbes ou les propositions circonstancielles, comme entre *profondément* ou *quand la nuit tombe* et *dort* dans <GN> Jean </GP><NV> dort </NV> <GR> profondément </GR> quand <GN> la nuit </GN><NV> tombe </NV> ;
- complémenteur (COMP) : pour lier le complémenteur et le noyau verbal de la proposition subordonnée, comme entre *qu'* et *viendra* dans <NV> Je pense </NV> qu'<NV> il viendra </NV> ;
- attribut-sujet/objet (ATB\_SO) : entre l'attribut et le noyau verbal, en indiquant si l'attribut se rapporte au sujet (comme entre *grand* et *est* dans <NV> il est </NV><GA> grand </GA>) ou à l'objet (comme entre *étrange* et *trouve* dans <NV> il trouve </NV> <GN> cette explication </GN> <GA> étrange </GA>) ;
- modifieur-nom (MOD\_N) : on retrouve ici le lien entre les différents éléments qui composeraient un "gros" GN, comme le lien entre *unique* et *fenêtre* dans <GN> l'unique fenêtre </GN> ou entre *de la chambre* et *la porte* dans <GN> la porte </GN> <GP> de la chambre </GP> ;
- modifieur-adjectif (MOD\_A) : pour indiquer ce qui se rapporte à un adjectif comme entre *très* et *belle* dans <GN> la très belle collection </GN> ou entre *de son fils* et *fière* dans <NV> elle est </NV> <GA> fière </GA> <GP> de son fils </GP> ;
- modifieur-adverbe (MOD\_R) : même type de relation pour les adverbes comme entre *très* et *gentiment* dans <NV> elle vient </NV> <GR> très </GR> <GR> gentiment </GR> ;
- modifieur-préposition (MOD\_P) : pour annoter ce qui se rapporte à une préposition comme entre *peu* et *avant* dans <NV> elle vient </NV> <GR> peu </GR> <GP> avant lui </GP> ;
- coordination (COORD) : pour lier les trois éléments que sont la coordination et ses coordonnés, comme entre *Pierre, Paul* et *et* dans <GN> Pierre </GN> et <GN> Paul </GN> <NV> arrivent </NV> ;
- apposition (APP) : la relation d'apposition lie l'élément apposé et celui auquel il s'appose en marquant l'identité entre les référents, comme entre *le député* et *Yves Tavernier* dans <GN> Le député </GN> <GN> Yves Tavernier </GN>... ;
- juxtaposition (JUXT) : elle est utilisée pour les constituants qui ne sont ni subordonnés ni coordonnés ni apposés, comme dans les cas d'énumération. Ainsi, elle lie *on ne l'entendait* et *elle était* dans <NV> on ne l'entendait </NV> <GR> plus </GR> ... <NV> elle était </NV> <GR> peut-être </GR> <GA> morte </GA> ;

Pour plus de détails sur l'ensemble de l'annotation, on peut se reporter au guide d'annotation qui a été rédigé pour servir à la fois aux annotateurs et aux participants, et qui est disponible à l'adresse suivante : <http://www.limsi.fr/Recherche/CORVAL/easy>

## **5 Évaluations prévues**

Dans le protocole d'évaluation, nous allons effectuer plusieurs mesures, tenant compte à la fois de la diversité des corpus à annoter et de la variété des annotations. Pour les corpus, nous donnerons une évaluation par type de données (journalistique, littéraire, oral, écrit "relâché").

Ensuite nous donnerons les résultats des participants sur la reconnaissance des constituants et de leurs frontières, mais en distinguant les types de constituants. En effet, tous les constituants ne posent pas des problèmes de même niveau. Les mesures mises en œuvre sont de classiques mesures de rappel et de précision. Ce type d'évaluation a déjà été pratiqué dans une pré-version de l'évaluation EASy, permettant déjà de montrer des variations entre deux analyseurs utilisés pour ces tests (Vilnat et al., 2004).

Pour évaluer la reconnaissance des relations, nous distinguerons également des sous-ensembles de relations : d'une part les relations entre constituants minimaux telles que les modificateurs du nom, de l'adjectif et de la préposition ou auxiliaire-verbe. Ensuite nous nous intéresserons aux relations correspondant à des relations syntaxiques "de base" telles que Sujet-Verbe ou COD-Verbe. Nous évaluerons dans un autre ensemble les relations réputées plus complexes, telles que l'apposition ou la juxtaposition. Notre but est de fournir des résultats détaillés, permettant d'avoir une évaluation précise des analyseurs, déclinée selon les différents phénomènes et les types de textes, plutôt qu'une évaluation trop globale. Nous espérons ainsi être en mesure de donner à des utilisateurs éventuels les moyens de choisir l'outil le mieux adapté à leurs besoins, et fournir un retour utile aux développeurs de systèmes.

L'état actuel de l'évaluation est de rendre possible les mesures prévues pour comparer les résultats des analyseurs et la référence. Pour cela, la concordance de la segmentation en énoncés et en formes de la référence et des sorties des analyseurs doit être vérifiée, car les analyseurs syntaxiques n'effectuent pas nécessairement des segmentations identiques à celles de la référence. Même si le corpus a été fourni sous forme brute et sous forme segmentée en énoncés et en formes (section 3), tous les participants n'ont pas choisi de travailler sur les données segmentées. Une étape de réaligement est donc en cours.

## **6 Travaux connexes en annotation de corpus et évaluation**

Tandis que la construction du Penn Treebank, (Marcus et al., 1993), se poursuit pour la langue anglaise, les travaux en annotation de corpus se développent depuis quelques années pour une grande majorité des langues européennes. Les 3 ateliers intitulés "Workshop on Treebanks and Linguistic Theories" (TLT 2002, 2003, 2004) nous informent sur les langues étudiées (le suédois, le bulgare, ou le basque par exemple) et sur les méthodes employées, de même le recueil d'Abeillé (Abeillé, 2003), qui présente la construction de treebanks dans plusieurs langues et aussi leur utilisation à des fins d'évaluation. Généralement les corpus constitués sont de grande taille, leur annotation concerne des aspects morphologiques, syntaxiques et plus rarement des aspects plus sémantiques (comme le traitement de la co-référence).

Une autre approche, toute récente, de l'annotation de corpus mérite aussi d'être citée, cette fois l'annotation ne porte que sur peu de phrases (très peu même pour l'instant) mais elle vise des aspects plus sémantiques allant jusqu'à la pragmatique. Un premier atelier a eu lieu l'an dernier et le suivant se tiendra pendant la conférence ACL fin juin 2005 (Frontiers in Corpus Annotation

2004, 2005).

En ce qui concerne l'évaluation des analyseurs, la méthode la plus largement utilisée a sans doute été la méthode Parseval (Black et al., 1991). Cette méthode est fondée uniquement sur la comparaison des frontières des constituants, celle-ci étant effectuée entre les constituants du corpus arboré de référence et ceux de l'analyseur se soumettant à l'évaluation. Les mesures effectuées sont les habituelles mesures de précision et de rappel. La méthode Parseval a fait l'objet de plusieurs critiques qui sont détaillées dans (Carroll et al., 2003) et auparavant dans (Carroll, 2002) ou (Briscoe et al., 2002). Parmi ses détracteurs, Lin (Lin, 1998) propose de fonder l'évaluation sur la structure en dépendances et non sur les frontières des constituants. Plus récemment, Carroll et al. (Carroll et al., 2003) ont mis en œuvre une méthode d'évaluation elle aussi fondée sur les relations grammaticales : ils ont annoté 10 000 mots à l'aide d'une douzaine de relations et calculé le rappel et la précision entre cette référence et les résultats d'un analyseur. Dans un esprit proche, Lin (Lin, 2003) a évalué son analyseur en dépendances Minipar en utilisant un corpus de plus grande taille, mais contenant moins de relations. Ces différents auteurs concluent de la même façon : trouver un ensemble commun d'éléments sur lequel mener l'évaluation n'est pas toujours immédiat. Mais l'approche fondée sur les relations leur paraît plus pertinente, car un analyseur qui aura construit les bons constituants d'une phrase pourra malgré tout être très loin d'en avoir une bonne représentation sémantique.

## 7 Conclusion

Pour la campagne EASY (Vilnat et al., 2003), (Vilnat et al., 2004) l'annotation du corpus s'est voulue exhaustive, les constituants et les relations grammaticales ont tout deux été annotés et peu de phénomènes ont été écartés. Au moment de l'évaluation, treize participants étaient finalement prêts à soumettre leur analyseur. Ce qui constitue sans aucun doute une bonne représentation des équipes travaillant sur les analyseurs du français. Parmi les systèmes présentés, certains sont le fruit de plusieurs années de travail, tandis que d'autres sont très récents (parfois même encore en développement). Qu'ils soient récents ou anciens, tous ne calculent pas constituants et relations. Pour ces raisons, il devenait naturel de calculer rappel et précision sur les constituants et sur les relations. En outre, tous les résultats seront différenciés : par type de corpus, par type de relations et de constituants ; nous permettant ainsi de dessiner une image plus précise et juste des performances de chacun.

## Références

Treebanks : Building and Using Parsed Corpora, 2003, Dordrecht Kluwer, 406 p.

Adda G., Mariani J., Paroubek P., Rajman M., Lecomte J., 1999, L'action GRACE d'évaluation de l'assignation de parties du discours pour le français, revue *Langues*, Vol. 2, No. 2, pp 119-129.

Adda-Decker M., Habert B., Barras C., Adda G., Boula de Mareuil P. et Paroubek P., 2003, A disfluency study for cleaning spontaneous speech automatic transcripts and improving speech language models, *Proceedings of the International Conference on Disfluency in Spontaneous Speech (DISS)*, Septembre 2003, Göteborg, pp 67-70.

Aït-Mokthar S., Chanod J., Roux C., 2002, Robustness beyond shallowness : incremental deep parsing, *Journal of Natural Language Engineering*, Vol. 8, No. 3-2.

- Black E., Abney S., Flickenger D., Gdaniec C., Grishman R., Harison P., Hindle D., Ingria R., Jelineck F., Klavans J., Liberman M., Marcus M., Roukos S., Santorini B. et Strzalkozski T., 1991, A Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars, *Proceedings of the 4th DARPA Speech and Natural Language Workshop*, Morgan Kaufman, Pacific Grove California, pp. 306-311, Février 1991.
- Briscoe E, Carroll J., Graham J. et Copestake A., 2002, Relational evaluation schemes, *Proceedings of the Workshop Beyond PARSEVAL - Toward improved evaluation measures for parsing systems*, Third International Conference on Language Resources and Evaluation (LREC), Las Palmas, Gran Canaria, May 2nd, 2002.
- Brunet E., 2000, Qui lemmatise dilemme attise, *Lexicometra*, No 2.
- Carroll J., Briscoe T. et Sanfilipo A., 1998, Parser Evaluation: a Survey and a New Proposal, actes de First International Conference on Language Resources and Evaluation (LREC), Granada, May 1998, vol. 1 pp. 447-454 .
- Beyond Parseval - Towards improved evaluation measures for parsing systems. *Atelier de la conférence LREC*, Las Palmas Espagne, John Carroll editor, 2002
- Carroll J., Minnen G. et Briscoe E., 2003, Parser evaluation using a grammatical relation annotation scheme, in A. Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*, Dordrecht Kluwer, pp. 299-316, 2003.
- Gala Pavia, Núria, 2003, Un Modèle d'analyseur syntaxique robuste fondé sur la modularité et la lexicalisation de ses grammaires, Thèse Université Paris-Sud, LIMSI-CNRS.
- Gendner V., Illouz G., Jardino M., Paroubek P., Monceaux L., Robba I. et Vilnat A., 2003, Proposition de protocole d'évaluation des analyseurs syntaxiques, *Atelier sur l'évaluation des analyseurs syntaxiques de la conférence TALN*, pp 87-94
- Lin D., 1998, Dependency-Based Evaluation of MINIPAR *Proceedings of the Workshop on Evaluation of Parsing Systems*, First International Conference on Language Resources and Evaluation (LREC), Granada, May 1998,
- Lin D., 2003, Dependancy-based Evaluation of Minipar, in A. Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*, Dordrecht Kluwer, pp. 299-316, 2003.
- Marcus M., Santorini B. et Marcinkiewicz M., 1993, Building a large annotated corpus of English : The Penn treebank, *Computational Linguistics*, 19:313-330
- Monceaux L., 2002, *Adaptation du niveau d'analyse des interventions dans un dialogue. Application à un système de question-réponse*. Thèse de l'univeristé Paris 11, Décembre 2002
- Oepen S., Netter N. et Klein J., 1996, Test Suites for Natural Language Processing, *Linguistic Databases, Nerbonne J. editor*, Center for the Study of Language and Information, CSLI Lecture Notes, 1996.
- Srivinas B., Doran C., Hockey B.A., Joshi A.K., " An approach to Robust Partial Parsing and Evaluation Metrics ", in *Proceedings of the Workshop on Robust Parsing*, ESSLI, Prague, August 1996.
- Vilnat A., Paroubek P., Monceaux L., Robba I. Gendner V., Illouz G. et Jardino M., 2003, EASY or How Difficult Can It Be to Define a Reference Treebank for French, *Proceedings of the 2nd Workshop on Treebanks and Linguistic Theories (TLT)* , Växjö, Sweden, pp. , November 14th-15th, 2003.
- Vilnat A., Monceaux L., Paroubek P., Robba I. Gendner V., Illouz G. et Jardino M., 2004, Annoter en constituants pour évaluer des analyseurs syntaxiques. Actes de TALN 2004, Fés, Maroc, pp. 467-476.
- Workshop "Frontiers in Corpus Annotation", <http://nlp.cs.nyu.edu/meyers/frontiers/2005.html>  
<http://nlp.cs.nyu.edu/meyers/frontiers/2004.html>
- Workshop on Treebanks and Linguistic Theories, <http://www.sfs.uni-tuebingen.de/tlt04/>,  
<http://w3.msi.vxu.se/rics/TLT2003/>, <http://www.bultreebank.org/TLT2002.html>