

Annoter en constituants pour évaluer des analyseurs syntaxiques

Anne Vilnat(1), Laura Monceaux(2), Patrick Paroubek(1), Isabelle Robba(1),
Véronique Gendner(1), Gabriel Illouz(1), Michèle Jardino(1)

(1) LIMSI - CNRS
Université d'Orsay, BP 133
91403 Orsay CEDEX
{prénom.nom}@limsi.fr

(2) LINA
2 rue de la Houssinière, BP 92208
44322 Nantes CEDEX 03
{prénom.nom}@lina.univ-nantes.fr

Résumé - Abstract

Cet article présente l'annotation en constituants menée dans le cadre d'un protocole d'évaluation des analyseurs syntaxiques (mis au point dans le pré-projet PEAS, puis dans le projet EASY). Le choix des constituants est décrit en détail et une première évaluation effectuée à partir des résultats de deux analyseurs est donnée.

This paper focuses on constituent annotation in a syntactic parsers evaluation protocol (which was elaborated in PEAS pre-project and EASY project). The choice of the constituents is described in details, and the results of a first evaluation between two parsers are given.

Mots-clefs – Keywords

annotation en constituants, évaluation, analyseurs syntaxiques
constituent annotation, evaluation, syntactic parser

1 Introduction

Avec le développement d'Internet, l'accès à de grandes quantités de données textuelles est devenu possible ; les analyseurs syntaxiques qui, en amont d'autres outils linguistiques, traitent ces données ont alors dû se fixer de nouveaux objectifs. Il doivent ainsi être capables de traiter

ces grandes quantités en des temps raisonnables et surtout ils doivent faire face à des données très variées dans leur forme et leur nature. C'est dans ce contexte que les analyseurs robustes ou les analyseurs partiels ont été mis au point. Dans le même temps, les analyseurs complets ont continué d'être étudiés, améliorés et plus récemment des approches *mixtes* ont été élaborées. Ces dernières retournent une analyse partielle de la phrase mais dont les éléments sont analysés en profondeur quand cela est possible (Frank et al., 2003), (Aït-Mokthar et al., 2002).

L'évaluation des analyseurs permet d'une part à ceux qui les utilisent de connaître leurs forces et leurs faiblesses, et d'autre part à ceux qui les développent de disposer d'une référence à laquelle ils peuvent se confronter. Cependant, comme le soulignent (Blache et Morin, 2003) ou (Briscoe et al., 2002), l'évaluation des analyseurs syntaxiques n'est en rien une tâche facile et ce pour diverses raisons : les sorties des analyseurs varient dans leur forme et leur nature d'un analyseur à l'autre, et même en parvenant à un accord sur ces sorties, des métriques équitables et représentatives ne sont pas actuellement disponibles. Conscients de ces difficultés, nous avons malgré tout mis au point un protocole complet d'évaluation des analyseurs syntaxiques, baptisé PEAS, et dont la volonté était de n'écarter la participation d'aucun analyseur. Ce protocole PEAS (Gendner et al., 2003) a servi de pré-projet au projet EASY¹ qui rassemble 5 fournisseurs de corpus et 14 analyseurs participants. La campagne d'évaluation menée par EASY au printemps 2004 sera, dans la mesure du possible, reproductible et donc reconduite autant d'années que nécessaire. Après une présentation du protocole EASY, nous détaillons les principes adoptés pour l'annotation, nous exposons les premiers résultats d'évaluation comparative entre deux analyseurs obtenus dans le cadre de PEAS et enfin nous évoquons quelques travaux connexes.

2 Présentation du protocole d'évaluation

Le protocole mis au point dans PEAS puis EASY est articulé autour des modules 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. Et son formatage : normalisation, *tokenisation* et découpage en phrases.
2. L'annotation d'un sous-ensemble de 20 000 mots. Elle est effectuée à l'aide d'un éditeur HTML et les résultats sont transcrits dans un format XML.
3. L'analyse par les participants et la transcription des sorties de leur analyseur dans un format XML commun.
4. L'évaluation, prévue actuellement comme un calcul du rappel et de la précision sur les constituants et les relations, est toujours ouverte à de nouvelles propositions.

Les décisions concernant le formalisme d'annotation ont été l'objet de nombreuses discussions avec les fournisseurs de corpus d'une part (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) avec les participants d'autre part (qui doivent pouvoir disposer d'un format dans lequel il leur sera possible de transcrire les résultats de leur analyseur). Par ailleurs, le

¹Le projet EASY fait partie de la campagne d'évaluation EVALDA du programme Technolangue, qui est à l'initiative du ministère délégué à la Recherche et aux Nouvelles Technologies ; il a débuté en janvier 2003.

formalisme doit permettre une couverture la plus large possible des phénomènes syntaxiques de la langue. Notre but est aussi de constituer un corpus arboré, utilisable pour d'autres applications, constitué à la fois d'une partie annotée manuellement, et d'une partie issue de la combinaison des résultats des différents analyseurs participant à l'évaluation. Nous ne développerons pas ce point ici, pour des raisons de place. Nous avons fait le choix d'annoter deux types d'information : les constituants, avec leur catégorie et leur étendue et les relations syntaxiques ou fonctionnelles, avec les éléments en relation, qui peuvent être des mots ou des constituants.

3 L'annotation en constituants

3.1 Principe général

L'annotation que nous allons détailler dans cet article concerne les constituants, en supposant que le découpage en phrases et le découpage en mots ont été préalablement établis. Ces deux découpages sont souvent considérés comme des problèmes plus simples et quasiment résolus, mais 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 alors 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 ré-introduire 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 ci-dessous.

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.

Pour le découpage en mots, nous nous sommes inspirés du travail réalisé dans le cadre de Grace (Adda et al., 1999) pour constituer une liste de mots composés ou de locutions qui ne formeront qu'un seul mot lors de l'annotation en constituants. En revanche aucun étiquetage morphosyntaxique de ces mots n'est fait au préalable (ni imposé ou fourni aux systèmes participants à l'évaluation). Par exemple dans : *Dès que le soleil se lève, les coqs chantent., Dès_que* ne forme qu'un seul mot.

Ces deux découpages étant faits, le découpage en constituants peut alors avoir lieu. Le problème se pose de déterminer le nombre et la taille de ces constituants. Le principe général que nous

avons adopté consiste à annoter des constituants minimaux, non discontinus 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 prépositionnel (*de la voisine*), extension (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 une relation *modifieur du nom*. C'est cette solution que nous avons adoptée. Elle nous permet de ne pas rejeter les analyseurs qui ne relèvent que les constituants simples et de mieux noter les analyseurs plus précis où ces deux informations sont retrouvées. Ce parti pris nous permet également de simplifier la tâche d'annotation.

3.2 Quels constituants retenir ?

Partant d'exemples réels et de la littérature sur le domaine ((Abeillé et Blache 2000), (Marcus et al., 1993)...), nous avons déterminé une liste de six constituants. Nous en donnons une première définition, que nous illustrons sur des exemples simples². La délimitation d'un constituant GX est donnée dans une notation à la XML : `<GX> le constituant GX </GX>`.

- 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'), ainsi que 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 conjugué mais aussi participe présent (`<NV> désobéissant </NV> à leurs parents`), participe passé (`<NV> fermée </NV> à clef`) et infinitif (`<NV> ne veut </NV> pas <NV> venir </NV>`). En cas de temps composé, 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

²Plusieurs extraits sont issus du *Mystère de la chambre jaune*, de Gaston Leroux.

(*<PV> d'ébranler </PV>*). Il peut contenir aussi des modificateurs de ce verbe, comme des adverbes (*<PV> de vraiment bouger </PV>*).

3.3 Quelques problèmes

Le découpage décrit ci-dessus, semble simple à respecter, pourtant lors de l'annotation, nous avons constaté que même pour les constituants comme le groupe nominal, ce n'est pas toujours si facile. Nous allons illustrer quelques-uns des problèmes liés à l'annotation du groupe nominal, du groupe prépositionnel et du noyau verbal. Cette présentation ne prétend pas être exhaustive mais illustrer à la fois les problèmes rencontrés et les solutions qu'a permis d'apporter le choix de nos constituants minimaux.

3.3.1 Le groupe nominal GN

Pour le groupe nominal, nous détaillerons le cas du glissement d'usage d'un adjectif en nom, et l'annotation des groupes nominaux désignant des personnes. Beaucoup d'adjectifs peuvent jouer le rôle de nom et se construire de la même façon, à savoir être sujet d'une phrase et se construire avec un déterminant seul. La question est alors de savoir si on fait glisser la catégorie du mot d'adjectif en nom, ou si on considère qu'il est possible d'avoir un groupe nominal sans nom, ou encore d'avoir un groupe adjectival comme sujet. Nous n'avons pas retenu cette dernière possibilité, considérant que la fonction (sujet) et la construction avec un déterminant sont des arguments suffisamment forts pour faire de ces constituants des groupes nominaux. Il reste à trancher entre les deux autres possibilités. Ainsi dans l'exemple suivant :

<GN> La plus belle </GN> de la collection est <GN> la verte </GN>.

faut-il considérer que *belle* et *verte* deviennent des noms, ou qu'il est possible de délimiter un groupe nominal constitué d'un déterminant et d'un adjectif ? Nous avons préféré ne pas forcer la transcatégorisation des mots, qui n'a pas de réelle justification, et donc accepter qu'un groupe nominal puisse être considéré comme bien formé, même s'il ne comporte aucun nom.

Une autre question s'est posée quand nous avons eu à annoter une phrase telle que :

Le président Xavier Chapuisat est présent.

Doit-il y avoir un seul, deux (*<GN> Le président </GN> <GN> Xavier Chapuisat </GN>*) ou trois groupes nominaux (*<GN> Le président </GN> <GN> Xavier </GN> <GN> Chapuisat </GN>*) ? Nous avons décidé de considérer que lorsque plusieurs noms propres se succèdent sans déterminant ni préposition, ils forment un seul GN, dans les autres cas, nous en formons plusieurs, qui seront liés par des relations de dépendance. D'où les annotations suivantes :

<GN> Le président </GN> <GN> Xavier Chapuisat </GN> est présent.

<GN> Le député </GN> <GN> PS </GN> est présent.

<GN> Le président </GN> <GP> de l'Université Paris-Sud </GP> est présent.

3.3.2 Le groupe prépositionnel GP

Il est important de noter que cette annotation en constituant GP n'est qu'un raccourci qui permet de noter à la fois l'étiquette du constituant et la relation entre le premier élément (la préposition) et ce qui suit (le GN introduit). Elle est possible dans la plupart des cas de GP, car la préposition n'est que rarement détachée du GN qu'elle introduit. C'est pourquoi nous l'avons

retenu. Cependant, on peut rencontrer des cas d'incises par exemple où il y a séparation entre ces éléments. Il faut alors revenir à l'annotation "complète", à savoir : annoter le GN comme tel, laisser la préposition non annotée et traduire la relation de dépendance entre la préposition et le GN par une relation de type *complémenteur*. Par exemple :

Il arrive <GP> en retard <GP> , avec , <GP> dans sa poche <GP> , <GN> un petit discours </GN> écrit qu' il est obligé de garder.

en retard et *dans sa poche* sont annotés simplement comme des GP. En revanche, la relation entre *avec* et *un petit discours* sera elle annotée à l'aide d'une relation de dépendance. Il faudra lors de l'évaluation considérer que ces annotations sont équivalentes.

On trouve un cas un peu similaire lors de coordinations où la préposition est "factorisée" pour deux GN.

<GN> les jeunes filles </GN> arboraient donc <GN> un fichu </GN> étroitement noué <GP> sous le menton </GP> <GP> dans les couloirs </GP> et <GN> la cour </GN> <GP> de récréation </GP>

La relation entre *dans* et *la cour* sera annotée elle aussi comme une dépendance.

3.3.3 Le noyau verbal NV

Nous allons détailler ici deux points : l'annotation des négations, et la distinction entre participes considérés comme noyau verbal et ceux considérés comme adjectifs. La particule de négation *ne* est incluse dans le NV auquel elle se rattache. Les autres adverbes de négation forment un GR car ils peuvent se trouver à distance du verbe auquel ils se rattachent. Dans la forme *ne...que*, *que* est adverbe et forme un GR, *ne* reste rattaché au verbe (dans le NV correspondant). Avec un infinitif, les adverbes de négation sont placés avant le verbe; dans ce cas ils sont inclus dans le NV ou le PV : *ma fille avait juré <PV> de ne jamais me quitter </PV>*.

La distinction entre noyau verbal et groupe adjectival n'est pas forcément claire quand il s'agit d'annoter des participes présents ou passés.

Dans le cas de temps composés, les participes passés qui suivent un auxiliaire , comme par exemple : *le coq <NV> a </NV> <NV> chanté <NV>* forment un noyau verbal. Un noyau verbal est également identifié quand on reconnaît une proposition participiale :

leurs enfants <NV> élevés <NV>, les parents entreprennent un grand voyage ces prescriptions <NV> améliorant </NV> son état, le médecin poursuit le traitement

Enfin, les participes sont considérés comme verbes quand ils sont modifiés par un complément, comme dans :

La porte de la chambre <NV> fermée </NV> à clef à l'intérieur.

Les côteaux <NV> environnant </NV> la ville sont très verts.

Dans les autres cas, ils sont notés comme des adjectifs.

3.4 Quel bénéfice ?

Pour conclure cette présentation, plusieurs questions se posent. La première est de savoir si l'annotation manuelle d'un gros corpus en suivant ces principes est possible. Pour répondre à cette question, nous avons d'abord annoté un premier corpus, ce qui nous a permis de vérifier le bien fondé de notre approche. Cette annotation a concerné aussi bien les constituants que nous avons présenté ici que les relations (non détaillées dans cet article), qui traduisent à la fois les

imbrications des groupes simples dans des groupes plus complexes comme nous l'avons montré plus haut, que les relations syntaxiques comme la relation *sujet* pour n'évoquer que celle-ci. Nous avons pour cela développé des outils d'annotation simples, ne nécessitant que l'usage d'un éditeur HTML standard. Cette étude ne nous a pas permis de détecter tous les problèmes, mais de lancer la campagne d'annotation du projet EASY, sur un corpus beaucoup plus gros. Les problèmes rencontrés par les annotateurs nous ont alors permis d'affiner nos définitions, sans remettre en cause le cadre général. On peut aussi se demander si cette annotation qui paraît simple permet de mettre en évidence des différences entre les analyseurs, même si nous souhaitons à terme procéder à une évaluation globale qui regroupe constituants et relations. C'est le but de l'étude que nous allons présenter ci-dessous.

4 Premiers résultats de l'évaluation menée dans PEAS

Dans le protocole d'évaluation PEAS, nous comparons les sorties des différents analyseurs à un corpus annoté de référence en calculant les mesures de précision et de rappel pour les différents constituants et pour chaque type de constituant présent dans le corpus de référence. Nous avons en effet élaboré un corpus de référence composé de différents types de texte : corpus journalistiques (noté dans les figures : LeMonde.num), des extraits de roman (ABU.num et L782.num), des questions et des corpus oraux (111997.num et SECODIP.num), afin de permettre une évaluation ciblée sur les phénomènes linguistiques propres à un type de texte particulier. Pour pouvoir comparer les sorties des analyseurs avec notre corpus de référence, elles devront être projetées dans le formalisme PEAS présenté section 3. Cette projection passe par l'application de règles de réécriture plus ou moins complexes. Dans la projection des deux analyseurs évalués, certaines informations ne sont pas représentées par le même type d'information (adverbe représenté par une étiquette morpho-syntaxique et non par un constituant) (Monceaux 2002).

Une fois la projection réalisée, la concordance de la segmentation en énoncés et en formes entre la référence et les sorties des analyseurs doit être vérifiée, car les analyseurs syntaxiques n'effectuent pas nécessairement des segmentations identiques à celles de la référence. Une étape de réalignement est donc nécessaire. Notons que le décompte des erreurs se fait selon la segmentation de la référence ce qui peut parfois entraîner une variation artificielle du nombre de formes (et potentiellement d'erreurs) certaines formes n'étant pas réalignées ou bien faisant l'objet d'une expansion en plusieurs tokens ; mais jusqu'à présent ce biais (variation des distances entre les points représentatifs des performances des systèmes) n'a jamais remis en cause les résultats globaux de l'évaluation, ni les différences relatives de performance entre les différents systèmes.

Dans les résultats des figures 1 et 2, on constate que l'analyseur p2 retourne plus d'informations précises et en plus grand nombre que l'analyseur p1, sauf pour le constituant PV. Mais si l'on regarde de plus près les mesures effectuées, pour chaque constituant, par type de corpus, on peut constater que dans certains cas la précision de l'analyseur p1 est légèrement plus élevée que celle de p2, comme par exemple pour le sous-corpus de questions, où la précision globale de tous les constituants est sensiblement meilleure pour l'analyseur p1.

D'après les mesures effectuées, la tâche de segmentation en constituants ne semble pas être aussi simple qu'il paraît, puisque globalement l'analyseur p2 renvoie environ 20% d'informations erronées et ne détecte pas environ 20% des constituants. Dans cette évaluation, on note aussi que le constituant GA semble être le constituant posant le plus de problèmes : pour les deux

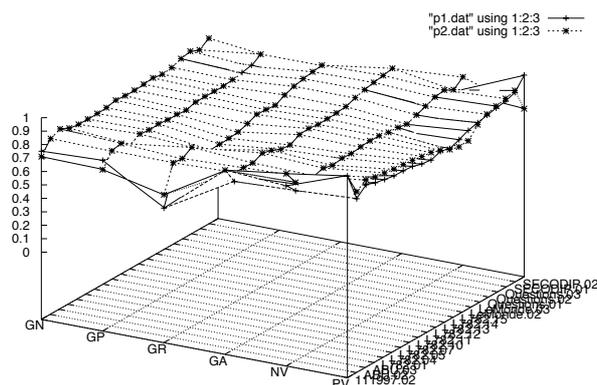


Figure 1: Résultats comparés en précision

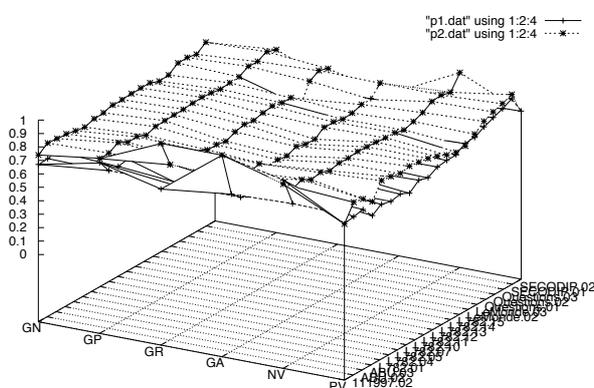


Figure 2: Résultats comparés en rappel

analyseurs, la précision et le rappel sont beaucoup plus faibles.

L'évaluation proposée par PEAS présente les comparaisons de performance en fonction du type de composant et de texte, ce qui permet d'effectuer des comparaisons fines entre les systèmes ou de calibrer au mieux les performances d'un analyseur en fonction des spécificités de l'application pour laquelle il sera utilisé.

5 Travaux en évaluation des analyseurs syntaxiques

A notre connaissance, pour les analyseurs du français, la première publication sur l'évaluation des analyseurs syntaxiques est (Abeillé, 1991) qui relate l'évaluation par un comité d'experts de plusieurs analyseurs du français. Cette évaluation portait sur le nombre d'analyses retournées et sur leur validité. Le projet TSNLP (Open et al., 1996) a tenté de capitaliser les connaissances linguistiques sous forme de jeux de tests. Il offre pour plusieurs langues européennes dont le français, des phrases de test avec des exemples positifs et négatifs, pour les principaux phénomènes syntaxiques. Les corpus arborés offrent l'avantage d'être plus représentatifs de la distribution des phénomènes linguistiques dans la langue, le premier et le plus célèbre est certainement le PennTreeBank (Marcus et al., 1993), qui contient des annotations basées sur les constituants et qui a inspiré d'autres développements, par exemple (Brants et al., 2002) ou pour le français (Abeillé et Blache 2000).

La première campagne d'évaluation comparative selon une méthodologie boîte noire est (Black et al., 1991) qui attaque le problème par le biais des mesures d'évaluation et propose de comparer les systèmes sur les frontières de constituants au moyen du taux de croisement (nombre de frontières de constituants produits par l'analyseur, qui croisent une frontière de constituant de la référence) et le taux de rappel (nombre de frontières de constituants produits par l'analyseur qui existent dans la référence). On a par la suite ajouté la précision à ces deux mesures pour former ce que l'on appelle le GEIG7 scheme (Srinivas et al., 1996) ou mesures PARSEVAL (Carroll et al., 2003). Malheureusement ces mesures ne sont applicables pratiquement que sur des constituants non étiquetés (dépourvus de catégorie), les sorties des analyseurs étant trop différentes pour essayer d'incorporer cette information dans le processus d'évaluation. Il en résulte une évaluation ne tenant compte que d'une partie des informations produites par les analyseurs et qui de plus s'applique plus facilement aux formalismes produisant une analyse en constituants.

Pour essayer d'ouvrir l'évaluation vers différents types d'analyseurs de nouvelles approches orientées sur les dépendances (Lin, 1998) ou même sur les relations grammaticales sont apparues (Carroll et al., 1998) (Briscoe et al., 2002) (Carroll et al., 2003). C'est l'approche que nous suivons dans EASY (Vilnat et al., 2003).

6 Conclusion

Les travaux initiés dans PEAS et en cours de développement dans la campagne EVALDA-EASY du programme TECHNOLOGUE ont permis de valider comme support d'évaluation, un formalisme d'annotation syntaxique pour le français, à la fois sur les phénomènes linguistiques effectivement rencontrés en corpus et vis-à-vis de la communauté des développeurs de systèmes. Ce résultat est concrétisé dans le guide d'annotation³, décrivant comment utiliser le formalisme. Nous avons aussi présenté un premier ensemble de mesures de précision et de rappel sur les frontières de constituants, réalisé sur le corpus PEAS avec le concours de deux participants témoins. Bien que les différences de performance observées soient relativement faibles, le calcul des résultats par type de constituant et de corpus permet d'avoir une perception fine des différences de comportement entre les deux systèmes. Ces résultats seront complétés dans un avenir très proche par une comparaison avec un algorithme de simple identification des constituants, une étude de pertinence statistique des différences observées, et une étude de l'effet sur les résultats d'un relâchement de contraintes sur les frontières de constituants. Bien entendu, les résultats prennent toute leur signification lorsque les mesures auront été effectuées dans la campagne EASY, sur un corpus de taille plus importante, avec plus de participants, et surtout accompagnées de mesures sur les dépendances.

Remerciements

Nous tenons ici à remercier les membres de ELDA qui participent à l'organisation de EASY : Khalid Choukri et Kevin Mc Tait ; ainsi que les premiers annotateurs, Rémy Bonnet et Pierre Zweigenbaum, et l'équipe de Jean Véronis (LPL à Aix) qui nous ont permis par leurs questions pertinentes et précises de compléter d'exemples et d'explications notre guide d'annotation.

³<http://www.limsi.fr/Recherche/CORVAL/easy/>

Références

- ABEILLÉ A. (1991), Analyseurs syntaxiques du français, *Bulletin Semestriel de l'ATALA (Association pour le Traitement Automatique des Langues)*, Vol. 32, No. 21, pp. 107-120.
- ABEILLÉ A., BLACHE P. (2000), Grammaires et Analyseurs syntaxiques, *Ingénierie des Langues sous la direction de J.-M. Pierrel*, Ch. 2, pp. 51-76, Hermes, Paris.
- 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, *Langues*, Vol. 2, No. 2, pp 119-129.
- AÏT-MOKTHAR S., CHANOD J., ROUX C. (2002), Robustness beyond shallowness : incremental deep parsing, *Journal of Natural Language Engineering*, Vol. 8, No. 3-2.
- BLACHE P., MORIN J.-Y. (2003), Une grille d'évaluation pour les analyseurs syntaxiques, *Atelier sur l'évaluation des analyseurs syntaxiques - TALN*, pp 77-86
- 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., STRZALKOZSKI T. (1991), A Procedure for Quantitatively Comparing the Syntactic Coverage of English Grammars, *4th DARPA Speech and Natural Language Workshop*, Morgan Kaufman, pp. 306-311.
- BRANTS T., DIPPER S., HANSEN S., LEZIUS W., SMITH G. (2002), The TIGER Treebank, *1rst Workshop on Treebanks and Linguistic Theories (TLT)*, Sozopol, Bulgaria.
- BRISCOE E, CARROLL J., GRAHAM J., COPESTAKE A. (2002), Relational evaluation schemes, *Workshop Beyond PARSEVAL - Toward improved evaluation measures for parsing systems*, 3rd LREC, Las Palmas, Gran Canaria.
- CARROLL J., BRISCOE T., SANFILIPPO A. (1998), Parser Evaluation: a Survey and a New Proposal, *1st LREC*, Granada, vol. 1 pp. 447-454
- CARROLL J., MINNEN G., 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
- FRANK A., BECKER M., CRYSMANN B., KIEFER B., SCHÄFER U. (2003) Integrated shallow and deep parsing : TopP meets HPSG , *41st ACL*, pp. 104-111, 2003
- GENDNER V., ILLOUZ G., JARDINO M., PAROUBEK P., MONCEAUX L., ROBBA I., 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 *Workshop on Evaluation of Parsing Systems*, 1st LREC, Granada.
- MARCUS M., SANTORINI B., 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'université Paris 11, Décembre 2002.
- OEPEN S., NETTER N., 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.
- SRIVINAS B., DORAN C., HOCKEY B.A., JOSHI A.K. (1996), An approach to Robust Partial Parsing and Evaluation Metrics, *Workshop on Robust Parsing*, ESSLI, Prague.
- VILNAT A., PAROUBEK P., MONCEAUX L., ROBBA I., GENDNER V., ILLOUZ G., JARDINO M. (2003), EASY or How Difficult Can It Be to Define a Reference Treebank for French, *2nd Workshop on Treebanks and Linguistic Theories (TLT)* , Växjö, Sweden.