*TALN 2003, Batz-sur-Mer, 11–14 juin 2003*

# Entre syntaxe et sémantique : Normalisation de la sortie de l'analyse syntaxique en vue de l'amélioration de l'extraction d'information à partir de textes

Caroline Hagège, Claude Roux
Xerox Research Centre Europe - XRCE
6 Chemin de Maupertuis - 38240 Meylan
caroline.hagege@xrce.xerox.com, claude.roux@xrce.xerox.com

## Mots-clefs – Keywords

analyse syntaxique robuste, normalisation en vue de l'extraction d'information
robust parsing, normalization for information extraction

## Résumé - Abstract

Cet article présente la normalisation de la sortie d'un analyseur robuste de l'anglais. Nous montrons quels sont les enrichissements que nous avons effectués afin de pouvoir obtenir à la sortie de notre analyseur des relations syntaxiques plus générales que celles que nous offrent habituellement les analyseurs robustes existants. Pour cela nous utilisons non seulement des propriétés syntaxiques, mais nous faisons appel aussi à de l'information de morphologie dérivationnelle. Cette tâche de normalisation est menée à bien grâce à notre analyseur XIP qui intègre tous les traitements allant du texte brut tout venant au texte normalisé. Nous pensons que cette normalisation nous permettra de mener avec plus de succès des tâches d'extraction d'information ou de détection de similarité entre documents.

This article presents our work on the normalization of the output of a robust dependency parser for English. We show how we have enriched our grammar to yield syntactic relations that are more general than those usually obtained with other available robust parsers. In order to achieve this result, we use syntactic properties, together with derivational morphology information. This normalization task is carried out with XIP which handles all the process that transform our input text into its normalized output. We consider that this normalization will improve the result of information extraction and similarity detection process between documents.

# Introduction

L'analyse syntaxique robuste est aujourd'hui une technique couramment employée dans le traitement automatique des langues (Giguet, Vergne, 1997), (Aït-Mokhtar et al., 1998), (Järvinen, Tapanainen, 1998), (Briscoe, Carroll, 2002). De nombreux analyseurs syntaxiques, en général basés sur les grammaires de dépendances (Mel'čuk 1988) sont capables actuellement, avec plus ou moins de succès, d'extraire à partir de textes tout venant les sujets, objets, modifieurs de verbes et des noms etc. Nous proposons ici de présenter un travail en cours, dans le prolongement de l'analyse syntaxique et qui consitue une étape de normalisation de la sortie de l'analyse syntaxique. Cette étape de normalisation est intégrée dans une chaîne de traitement allant de la segmentation du texte jusqu'à une représentation syntaxique normalisée. La sortie d'une analyse de ce type devrait permettre d'effectuer avec davantage d'acuité des tâches comme l'extraction d'information à partir de documents. Dans une première partie nous présenterons l'outil que nous utilisons, puis nous décrirons le travail autour de la normalisation et enfin nous évoquerons le problème de l'évaluation.

# 1 Analyse syntaxique robuste

## 1.1 Présentation de l'analyseur

XIP (Ait-Mokhtar et al., 2002) est un analyseur dont l'objectif est d'extraire des dépendances syntaxiques de façon robuste. XIP, de par sa souplesse, permet d'adopter des approches différentes dans l'écriture des grammaires (Hagège, Roux, 2002) :

- Une approche incrémentale qui consiste à placer des règles spécifiques et fiables en tête de façon à filtrer en amont les configurations rares ou exceptionnelles et de laisser les règles placées en fin de grammaire se charger des cas les plus généraux qui n'auraient pas été captés auparavant.

- Une approche par contrainte (au sens des grammaires de contraintes de Helsinki, (Järvinen, Tapanainen, 1998)) qui, au contraire, laisserait des règles très générales s'appliquer laissant ainsi des ambiguités dans les fonctions syntaxiques extraites. Puis, des règles permettant de filtrer les analyses s'appliquent plus avant, éliminant ainsi certaines possibilités.

Le formalisme proposé par XIP nous permet d'exprimer un large éventail de règles qui vont de la désambiguisation catégorielle, en passant par la constitution de syntagmes noyaux et la construction des dépendances[1]. Nous pouvons distinguer deux grandes familles de règles :

- Les règles qui permettent de construire les syntagmes noyaux,

- Les règles qui permettent de construire des dépendances.

[1]XIP nous permet en fait de relier par des relations des éléments linguistiques qui peuvent être des eléments lexicaux, mais aussi des élements non lexicaux correspondant à des syntagmes noyaux.

Voici par exemple une règle de la première famille permettant de construire un syntagme noyau nominal :

```
10> SN = Det, ?*, Noun.
```

Chaque noeud lexical apparaissant dans une règle peut être contraint par un jeu de traits supplémentaires, traits qui peuvent être testés, ajoutés ou retirés[2].

Une fois, les syntagmes noyaux constitués, une seconde famille de règles prend le relais pour lier certains noeuds têtes (généralement le fils le plus à droite d'un syntagme noyau) placé dans une configuration particulière. Ces règles, dites de déduction, sont composées de trois parties, la première ou la deuxième partie étant facultatives.

- La première partie définit à l'aide d'expression régulières la configuration particulière de noeuds sur laquelle la règle s'applique.
- La seconde partie pose des conditions supplémentaires (exigence de certains traits sur certains noeuds, présence ou absence préalable d'une dépendance)
- La troisième partie correspond au résultat de l'application de ces règles. Elle peut consister en la création d'une liste de dépendances construites sur des noeuds spécifiques, ou encore en la destruction ou la modification de dépendances déjà existantes.

Par exemple, la règle ci-dessous dont chacune des trois parties figure sur une ligne séparée permet d'établir une relation SUJET entre le noyau d'un syntagme nominal et un verbe.

```
| SN{?*,#1[last:+]}, ?*[verb:~], SV{?*, #2[last:+]}|
if (~SUJET(#2,#1))
SUJET(#2,#1).
```

La règle se lit comme suit :

Première ligne : L'on trouve un SN (syntagme nominal noyau) dont le dernier fils à droite (possèdant le trait *[last:+]*) est représenté par une variable #1. Ce SN est suivi par zéro ou plusieurs noeuds de même niveaux (lexicaux ou non lexicaux), ces noeuds ne pouvant en aucun cas porter le trait *[verb:+]*. Ils sont suivis à leur tour par un noeud SV dont le dernier élément est représenté par la variable #2.

Deuxième ligne : La condition stipule que entre le dernier élément du SV (le verbe) et le noyau du syntagme nominal aucune relation sujet n'a été établie auparavant.

Troisième ligne : On construit une relation sujet entre ce verbe et le noyau du syntagme nominal.

## 1.2 Evaluation de l'analyseur

L'analyseur XIP de l'anglais (première version) a été évaluée en Septembre 2002 sur un fragment du corpus SUZANE annoté à la main[3]. Nous avons obtenu les résultats suivants (toutes

[2] Une catégorie dans XIP est un ensemble de traits.

[3] 160 phrases prises arbitrairement dans le corpus.

dépendances confondues) : 79% de précision et 75% de rappel. Les performances de XIP pour la grammaire générale de l'anglais sont de 2300 mots analysés par seconde sur un Pentium à 1GHz.

# 2 Normalisation

Le travail de normalisation que nous envisageons a pour cible l'extraction d'information à partir de documents. Il nous permet d'avoir, après l'analyse, une représentation commune pour des suites de signifiants qui ne sont pas identiques. A l'heure actuelle, ce travail de normalisation s'effectue selon trois axes :

1. L'exploitation des relations syntaxiques mises en évidence lors de l'analyse générale.
2. La hiérarchisation des propositions dans une phrase.
3. L'exploitation d'information de morphologie dérivationnelle.

## 2.1 Propriétés syntaxiques

L'exploitation des relations syntaxiques mises en évidence lors de l'analyse générale se fait en deux temps. Dans un premier temps, on raffine l'analyse grammaticale afin de considérer les sujets et objets de verbes non finis, de prendre en compte les antécédants des relatives dans le calcul du sujet et de l'objet, et de normaliser la forme passive en forme active.

Dans un deuxième temps, on exploite certaines alternances verbales telles qu'elles ont été définies dans (Levin, 1993).

### 2.1.1 Raffinement de l'analyse grammaticale

Nous avons défini les relations de *sujet-normalisé* (SUBJ-N), *objet-normalisé* (OBJ-N) comme suit :

- Le sujet d'un verbe fini à la forme active est relié à ce verbe par la relation **SUBJ-N** et l'objet direct d'un verbe fini à la forme active est relié à ce verbe par la relation **OBJ-N**.
- Le complément d'agent d'un verbe fini à la forme passive est un sujet normalisé et le sujet d'une passive est un objet normalisé.
  *Some arrests were* **reported** *by the* **FBI** *in 1959*
  **SUBJ-N(report,FBI)**, **OBJ-N(report,arrest)**
- L'antécédant d'un pronom relatif sujet est le sujet normalisé du verbe de la relative et l'antécédant d'un pronom relatif objet est l'objet normalisé du verbe de la relative.
  *There was a unanimous vote, according to* **Republicans** *who* **attended** *the conference*
  **SUBJ-N(attend,republican)**
  *However, wave flying has its own* **problems** *which a pilot must* **understand**
  **OBJ-N(understand,problem)**

- Si une expression quantifiée est sujet d'un verbe, le sujet normalisé est la tête quantifiée par cette expression, même si syntaxiquement cette tête se trouve dans un syntagme prépositionnel dépendant du quantifieur. De manière similaire les expressions quantifiées objet d'un verbe donnent lieu à l'établissement d'une relation **OBJ-N** entre le verbe et la tête quantifiée.
  *Three of the* **boys** *were* **missing** - **SUBJ-N(miss,boy)**
  *They* **need** *three of the* **boys** - **OBJN(need,boy)**

- Les sujets et les objets de formes non finies sont récupérés comme des sujets normalisés et objets normalisés.
  *He ordered* **Peter** *to* **go** - **SUBJ-N(go,Peter)**.
  *I saw* **John leaving** *the room* - **SUBJ-N(leave,John)**, **OBJ-N(leave,room)**

- La relation existant entre un modifieur qui est une forme en *-ING* et le noyau nominal qu'il modifie est une relation de sujet normalisé.
  *Recessive host* **mutation affecting** *virus multiplication...* - **SUBJ-N(affect,mutation)**

- La tête modifiée par un participe passé est un objet normalisé de ce participe passé.
  **Measures taken** *so far...* - **OBJ-N(take,measure)**

Il est à noter que cette normalisation peut se combiner. En effet, si nous avons une relative dont le verbe est à la forme passive et que le relatif qui l'introduit a une fonction sujet, la relation entre l'antécédant de cette relative et le verbe de la relative sera **OBJ-N**. En d'autres termes pour
*The* **measures** *that have been* **taken** *last week...* on obtient une relation **OBJ-N(take,measure)**.

### 2.1.2 Alternances

Nous avons intégré dans le lexique et dans la grammaire de l'information nous permettant de prendre en compte les alternances de transitivité et quelques alternances de changement d'arguments d'un verbe. Nous avons pour cela utilisé les classes d'alternances définies par B. Levin (Levin, 1993) et avons marqué les verbes auxquels peuvent s'appliquer les alternances que nous considérons. Puis, des règles XIP nous permettent de transformer un ensemble de dépendances mettant en jeu l'un de ces verbes marqués en un autre ensemble de dépendances (voir le deuxième exemple ci-dessous).

### 2.1.3 Mise en oeuvre

L'extraction des sujets, objets normalisés d'une part, et la normalisation par les alternances d'autre part sont un prolongement naturel de la grammaire et se font par le biais des règles de déduction de XIP présentées plus haut.

Exemple1 : Extraction de l'objet normalisé correspondant au sujet de la passive[4].

[4] VDOMAIN correspond à la relation reliant le premier signifiant d'une suite verbale et le dernier signifiant d'une suite verbale, On aura donc la relation VDOMAIN(was,eat) pour la suite *was eaten* dans *The mouse was eaten by the cat*.

```
If ( SUBJ(#1,#2) & VDOMAIN[passive](#1,#3) )
OBJN(#3,#2)
```

Exemple 2 : On normalise la variante intransitive de verbes sujets à l'alternance moyenne (*Middle alternation*) en sa variante transitive. Le sujet normalisé de la forme intransitive est transformé en objet normalisé. On crée une relation **SUBJ-N** entre le verbe et un lemme variable (SOMEONE) portant le trait human:+.

```
if ( ^SUBJ-N(#1[mid_a],#2[human:~]) & ~OBJ-N(#1,#3) )
OBJ-N(#1,#2),
SUBJ-N(#1,##Pron[lemme=SOMEONE,human=+]).
```

Ainsi la phrase *The meat cut easily* produit dans un premier temps (grammaire générale) la relation **SUBJ(cut,meat)**, puis dans un deuxième temps (première phase de normalisation) la relation **SUBJ-N(cut,meat)**. Enfin, après l'application de la règle ci-dessus les relations **SUBJ-N(cut,SOMEONE)** et **OBJ-N(cut,meat)** sont créés.

## 2.2 Hiérarchisation des propositions

Les généralisations décrites ci-dessus s'accompagnent de la délimitation des propositions à l'intérieur d'une phrase. Il semble en effet intéressant de pouvoir déterminer à quel niveau d'enchâssement se trouvent les propositions analysées dans le cadre de la normalisation pour l'extraction de l'information. En effet, dans le contexte d'une application de question/réponse, le fait de connaître à quel degré d'enchâssement se trouve une réponse constitue une information qui nous permet de hiérarchiser la pertinence des réponses obtenues. (On préfèrera renvoyer une phrase dont la réponse à la question se trouve à un niveau d'imbrication faible par rapport à une phrase dont la réponse se trouve dans une proposition très imbriquée).

Nous avons défini dans la grammaire générale une relation unaire **MAIN** qui identifie le verbe de la clause principale, et une relation binaire **EMBED** qui relie deux verbes dont le premier argument est le verbe de l'enchâssée dans la proposition représentée par le premier argument. Cette relation **EMBED** est déduite de l'analyse en dépendances (On considère un enchâssement de propositions lorsqu'un verbe est dépendant d'un autre verbe fini)

Ainsi par exemple dans la phrase :
*Occasionally the stall is recognised by the pilot so that a preventative movement can be made before any serious wing-drop has had time to develop*

Les relations **MAIN[passive](recognise), EMBED(make,recognise)** et **EMBED(have,make)** sont extraites

En d'autres termes, le verbe principal de la phrase est *recognise*, dont dépend une enchâssée dont le verbe principal est *make* qui est à son tour un verbe dont dépend une nouvelle enchâssée représentée par le verbe *have*.

## 2.3 Mise en oeuvre

Ce sont également des règles de déduction qui exploitent les dépendances préalablement extraites par la grammaire générale que nous utilisons ici.

Par exemple la règle :

```
if ( MAIN(#1) & ?[sentence:~](#1,#9) & MOD[sentence,relativ](#9,#10) )
EMBED[relativ=+](#10,#1)
```

s'applique si le verbe principal (MAIN) est gouverneur d'une forme nominale (on ne spécifie pas ici quel type de relation existe entre le gouverneur et la forme nominale) et que cette forme nominale est modifiée par une relative (dans la pratique la dépendance relie la forme nominale et le verbe de la relative). Dans ce cas, alors, une relation EMBED possédant le trait [relativ:+] est créée entre le verbe principal et le verbe de la relative.

## 2.4 Utilisation de la morphologie dérivationnelle

Le troisième axe sur lequel nous basons notre travail de normalisation utilise de l'information de morphologie dérivationnelle. L'idée de l'utilisation de relation entre formes liées morphologiquement n'est pas nouvelle et plusieurs travaux pour l'apprentissage automatique de mots issus de la même famille morphologique ont été publiés (Dal, Namer, 2000) (Gaussier, 1999). Dans le cadre de travaux sur le résumé multi-document la notion de paraphrase en utilisant des familles morphologiques est également présentée dans (McKeown et al., 1999).

Nous avons opté pour l'utilisation de la ressource lexicale CELEX qui fournit un arbre de dérivations morphologiques pour le lexique de l'anglais. Nous nous sommes concentrés sur quelques familles seulement :

Verbes et nom de même famille avec suffixation +or (e.g. conduct-conductor), suffixation +er (e.g. accuse, accuser) et suffixation en +ion (accelerate-acceleration). Nous nous sommes également penchés sur la derivation +(a)n mettant en relation des noms geographiques et les adjectifs correspondants (e.g. Albania-Albanian or Djibouti-Djiboutean).

Nous avons de plus défini un ensemble de relations qui relient les paires nom-verbes extraites de CELEX mentionnés plus haut. Ces relations sont indépendantes du type de suffixe mais permettent de d'exprimer de manière plus précise les équivalences entre verbe-compléments et nom-arguments que nous utilisons pour normaliser (voir plus loin). Ces relations sont :

- S0 : Relation entre nom et verbe lorsque le nom paraphrase l'action exprimée par le verbe (ex. *acceleration-accelerate*).
- S1H : Relation entre nom et verbe lorsque le nom a le trait humain et correspond au premier actant pour l'action exprimée par le verbe (ex. *beginner,begin*).
- S1NH Relation entre nom et verbe lorsque le nom n'a pas le trait humain et correspond au premier actant pour l'action exprimée par le verbe (ex. *abbreviation,abbreviate*).
- S2 Relation entre nom et verbe lorsque le nom est le second actant pour l'action exprimée par le verbe (ex. *affirmation,affirm*).

Il faut noter qu'une même paire nom-verbe peut être reliée par plusieurs relations.

Afin d'injecter dans la grammaire cette information, XIP offre une nouvelle famille de règles qui sont un cas particulier des règles de déduction mentionnées ci-dessus. Ces règles ont la particularité d'être indexées sur un lemme.

Par exemple, pour indiquer que *inventor* est un nom relié au verbe *invent* par la relation S1H, on écrira:

```
inventor : noun = S1H(#0,##verb[lemme=invent]).
```

Cette règle est lue par XIP qui crée une relation **S1H(inventor,invent)**, relation qui pourra être manipulée dans des règles de déduction XIP au même titre que toute autre relation extraite par l'analyseur.

Ainsi, l'analyse de la suite *The inventor of the process is Smith* par la grammaire générale de l'anglais produit les relations suivantes :

DETD(inventor,the)
MODPREP(inventor,process,of)
ATTRIB(inventor,Smith)[5]

A ces relations vient s'ajouter la relation produite par la règle ci-dessus :

S1H(inventor,invent)

Enfin, par la règle de déduction XIP suivante :

```
if ( S1H(#1,#9) & MODPREP(#1,#2,?[form:fof]) & ATTRIB(#1,#3) )
OBJ-N(#9,#2), SUBJ-N(#9,#1).
```

nous obtenons enfin les relations normalisées **OBJ-N(invent,process)** et **SUBJ-N(invent,Smith)**.

Ces deux relations sont également obtenues lors de l'analyse normalisée des suites *Smith invented the process, The process was invented by Smith, Smith is the inventor of the process* ou *the invention of the process by Smith.*

## 2.5 Un exemple de normalisation

Prenons par exemple la phrase :

*You will need to connect the handset wires from the terminals.*

L'analyse générale de cette phrase nous permet d'obtenir (entre autres) les relations suivantes : **SUBJ(will,you), MAIN[modal](need), VDOMAIN(will,need), OBJ[post,nfinit](need,connect), OBJ(connect,wire), MOD(connect,terminal)**

La première étape de normalisation nous permet d'obtenir les relations : **SUBJ-N(connect,you)** (récupération du sujet de l'infinitive), **OBJ-N(connect,wire)**, **OBJ-N(connect,terminal)** (application de l'alternance réciproque[6] pour le verbe *connect*). Lors de la deuxième étape, on

[5] ATTRIB relie deux noms le premier étant sujet, le deuxième étant complément d'une forme de BE.

[6] Cette alternance indique qu'une suite SN1 V SN2 SP(Prep, SN3) est équivalente à la suite SN1 V SN2 and SN3.

obtient **EMBED[infinit](connect,need)**, qui nous indique que le verbe *connect* est dans une infinitive dépendante du verbe principal de la phrase. Enfin, l'application de la troisième étape ne donne aucune information supplémentaire pour l'analyse de cette phrase.

Prenons par ailleurs cette autre phrase :

*The connection of the terminal is needed.*

La sortie de l'analyseur général nous permet d'obtenir (entre autres) les relations : **MOD[post](connection,terminal), SUBJ(is,connection), MAIN(need), VDOMAIN(is,need)**.

Lors des première et deuxième étapes nous obtenons la relation **OBJ-N(need,connection)**. Et enfin, l'utilisation de l'information morphologique dont nous disposons, **S0(connection,connect)** nous permet de déduire par une règle XIP la relation **OBJ-N(connect,terminal)**.

Nous pouvons constater que les deux ensembles de dépendances calculés pour chacune de ces phrases par l'analyseur général n'ont pas d'intersection commune. Par contre, les deux ensembles de relations syntaxiques après normalisation présentent une relation en commun (i.e. **OBJ-N(connect,terminal)**) et une relation très similaire **OBJ-N(need,connection)** et **EMBED(connect,need)** sachant que *connect* et *connection* appartiennent à la même famille morphologique).

# 3 Comment évaluer ?

Un des problèmes importants que nous rencontrons dans ce travail est celui de son évaluation. En effet, nous ne disposons pas de corpus annoté avec le type de relations que nous avons définies. La seule possibilité semble être une évaluation indirecte, que nous n'avons pas encore effectuée et qui permettrait de vérifier l'influence de ces traitements pour l'accomplissement d'une tâche particulière d'extraction d'information. Nous envisageons de faire cette expérience dans le contexte d'une application de question/reponse.

# Conclusion

Cet article nous a permis de présenter un travail en cours de réalisation : la normalisation de la sortie d'un analyseur syntaxique en vue d'améliorer des applications d'extraction d'information. D'autres enrichissements que ceux actuellement mis en place sont envisagés notamment l'utilisation d'autres alternances, ainsi qu'un élargissement des familles morphologiques considérées. Il est intéressant de souligner que l'outil que nous utilisons, XIP, nous a permis d'intégrer tous les traitements présentés ici et que les nouvelles relations normalisées extraites viennent s'ajouter (ou peuvent se substituer) aux relations syntaxiques de surface calculées par l'analyseur de base.

# Références

Abney S. (1991), Parsing by chunks *Principled-Based Parsing* R. Berwick, S. Abney, and C. Tenny, editors. Kluwer Academic Publishers, Dordrecht.

Aït-Mokhtar S., Chanod J-P., Roux C. (1998), *Multi-Input Dual-Entry Point Dependency Parser*, IWPT 2001, Beijing.

Aït-Mokhtar S., Chanod J-P., Roux C. (1998), *Robustness beyond shallowness: incremental dependency parsing. Special issue of the NLE Journal*,

Briscoe, T., Carroll, J. (2002) *High Precision Extraction of Grammatical Relations* penultimate draft to appear in Bunt, H., Carroll, J. and Satta, G. (eds) Proc. of Int. Workshop on Parsing Technologies, Kluwer

Dal G., Namer F. (2000) *Génération et analyse automatiques de ressources lexicales construites utilisables en recherche d'information. Traitement automatique des langues pour la recherche d'information, numéro spécial de la revue T.A.L*, Ch. Jacquemin éd., Paris.

Gaussier E. (1999) *Unsupervised learning of derivational morphology from inflectional lexicons. Workshop on Unsupervised methods for natural language processing*, ACL 1999.

Giguet E., Vergne J. (1997) *From Part of Speech Tagging to Memory-based Deep Syntactic Analysis*, Fifth International Workshop on Parsing Technologies, Boston.

Hagège C., Roux C. (2002), *A Robust and Flexible Platform for Dependency Extraction*, LREC 2002, Las Palmas (Spain).

Jacquemin B. (2003) *Construction d'une structure sémantique de l'information dans une base documentaire pour son interrogation en langage naturel* Thèse de Doctorat, Université Paris III. A soutenir

Järvinen T., Tapanainen P. (1998) *Towards an implementable dependency grammar* Proceedings of the COLING-ACL'98. Workshop on the Processing of Dependency-based Grammars, 1998.

Levin B. *English Verb Classes and Alternations - A Preliminary Investigation* The University of Chicago Press, 1993.

McKeown K., Klavens J., Hatzivassiloglou V., Barzilay R., Eskin E. (1999) *Towards Multidocument Summarization by Reformulation: Progress and Prospects*, AAAI/IAAA, 1999.

Mel'čuk I., (1988) *Dependency Syntax*. State University of New York, Albany, 1988.

Tapanainen P, Jarvinen T. (1997) *A non-projective dependency parser*, Proceedings of the 5th Conference on Applied Natural Language Processing (ACL), Washington, D.C., April 1997.

Tesnière L, (1959) *Eléments de syntaxe structurale*, Klincksiek, Paris

Trouilleux F. (2002) *Insertions et interprétation des expressions pronominales*, Actes de la 9ème Conférence Annuelle sur le Traitement Automatique des Langues TALN 2002, Nancy, France.

Voutilainen A., Heikkila J., (1994) *An English constraint grammar (EngCG): a surface syntactic parser of English*, Fries, Tottie and Schneider (eds.), Creating and using English language corpora. Rodopi.