

Que, la valse des étiquettes

Marie-Paule Jacques

Équipe de Recherche en Syntaxe et
Sémantique, UMR 5610

Maison de la Recherche, Université
Toulouse le Mirail

Un module dédié à l'analyse de *que*

Module de l'analyseur SYNTEX (Bourigault et al.) :

- prend en entrée un texte étiqueté par TreeTagger, produit une analyse syntaxique en dépendance ;
- divers modules se succèdent pour déterminer les relations de dépendances des différents mots de la phrase ;
- les plus tardifs peuvent s'appuyer sur les relations placées par les modules précédents.

Le module d'analyse de *que*

Se prononce sur deux types d'étiquettes :

- catégorie morphosyntaxique
- relations de dépendance avec autres mots de la phrase : recteur, régi



Principe : mener de front une analyse syntaxique et des décisions sur la catégorie morphosyntaxique, en fondant celles-ci sur la syntaxe.

Catégorie morphosyntaxique

Pour TreeTagger, trois étiquettes possibles :

– adverbe

Que cela nous semble alors loin !

Les geysers n'entrent en activité que la nuit et au petit matin.

– pronom relatif

Le volcan est connu pour les risques qu'il constitue pour les populations avoisinantes.

– conjonction de subordination

Les raisons invoquées proviennent du fait que la médecine n'est pas une science exacte.

Quelques difficultés

*On appelle compétence du courant la **possibilité** qu'il a de transporter des matériaux.*

*J. et M. semblent exclure la **possibilité** que des opportunités demeurent non perçues.*

*Cependant on **n'est** plus persuadé **que** cette action rasante soit entièrement responsable des formes de champignons constatées dans les déserts*

*Le volcan **n'émet plus**, par de nombreux points du cratère, **que** de la vapeur d'eau surchauffée*

Quelques difficultés

Une solution française semble dans ce cas plus difficile à envisager qu'une reprise par un groupe étranger.

Il n'en reste pas moins difficile à expliquer que les crêtes et sillons pré littoraux ne soient pas toujours parallèles à la côte.

Méthode d'analyse

Principes :

- analyse en deux temps fondée sur des règles
 1. exploitation d'informations contextuelles
 2. exploitation d'informations de structure
- procède au réétiquetage de *que* **en fonction des constructions identifiées**, si aucune construction n'est positivement identifiée, conservation de l'étiquette attribuée par TreeTagger.

Etape 1 : informations locales

Un étiquetage erroné dégrade l'analyse syntaxique.

*Aussi ne peut-on **que** s'inquiéter de voir l'Afghanistan désormais sous la coupe de ces soldats séminaristes archaïques et brutaux.*

Donc, nécessité de réviser l'étiquetage morphosyntaxique très tôt dans la chaîne.

Mais, peu d'informations de structure.

Etape 1 : informations locales

Analyse très locale :

- rattachement des auxiliaires et des adverbes
- liste de noms (*fait, possibilité, probabilité, hypothèse, etc.*) + que
- liste de verbes (*penser, croire, estimer, considérer, espérer, etc.*) + que
- liste d'adverbes « comparatifs » (*plus, moins, davantage, aussi, autant, etc.*) [+ Adj.] + que

il n' a été exploré qu' à la fin du XIXème siècle

il forme une montagne de plus de 9 000 mètres, plus grande que l'Everest !

Etape 2 : informations structurelles

Permettent analyse à grande distance

Mêmes éléments et règles que précédemment
+ toutes relations de dépendance

*Un « champ électrique » implique une activité qui **ne** peut être exprimée d'une façon univoque **qu'**en fonction du temps et de deux ou trois dimensions.*

*Les procédures s'appliquent **aussi bien** à une interface à débit de base **qu'**à une interface à débit primaire.*

Exemple

Il apparaît au contraire à Paul Amiell que la qualité de ses produits et leur suivi de fabrication exigent une production en France.

A gauche de *que*, Nom Propre, régi par une Prep, régie par un verbe. Test de conditions sur le verbe :

- présence seule de *ne* : négation exceptive → Adv ;
- présence des adverbes : *plus, moins, davantage, autant, mieux, tellement...* → CSub ;
- appartenance du verbe à la liste de ceux qui prennent *que* comme objet direct → CSub + recherche du verbe de la subordonnée introduite par *que*.

Résultats

Étiquetage morphosyntaxique

Étape	Précision	
Étiquetage initial	75 %	
Étape 1	89%	+14%
Étape 2	92%	+3%

Résultats

Relations de dépendance syntaxique

Etape	Rappel		Précision	
Étape 1	60%		97%	
Étape 2	93%	+33%	94%	-3%

Conclusion

Stratégie payante : amélioration de l'étiquetage de *que*, analyse syntaxique convenable, mais...

Différencier selon le genre ?

Impact sur le reste de l'analyse ?

Analyse de surface *vs* interprétation des relations

Adopter la même méthode pour d'autres mots multicatégoriels ? (*comme, si...*)