



# Améliorer l'étiquetage de "que"

par les descripteurs ciblés et les règles

Assaf URIELI

UE TAL

CLLE-ERSS - UMR 5263  
Université de Toulouse

et

Joliciel Informatique  
Foix, France

13 octobre 2014



## Problématique

- Evolution du TAL français depuis 10 ans :
  - Avant : méthodes à base de règles et grammaires
  - Après : méthodes statistiques à base d'apprentissage
- Pour faire un étiqueteur statistique, prenez :
  - Un corpus annoté (ex. French Treebank)
  - Un classifieur (ex. SVM)
  - Un algorithme (ex. gauche à droite séquentiel)
  - Quelques descripteurs
- Atouts :
  - État de l'art facile à atteindre
  - Système facile à maintenir
- Sacrifices :
  - Dépendant du corpus d'apprentissage
  - Difficulté de regarder à l'intérieur des boîtes noires
  - Beaucoup d'efforts pour gagner 0,10 % d'exactitude
  - Fossé creusé entre experts de l'apprentissage et linguistes

## Problématique

- Evolution du TAL français depuis 10 ans :
  - Avant : méthodes à base de règles et grammaires
  - Après : méthodes statistiques à base d'apprentissage
- Pour faire un étiqueteur statistique, prenez :
  - Un corpus annoté (ex. French Treebank)
  - Un classifieur (ex. SVM)
  - Un algorithme (ex. gauche à droite séquentiel)
  - Quelques descripteurs
- Atouts :
  - État de l'art facile à atteindre
  - Système facile à maintenir
- Sacrifices :
  - Dépendant du corpus d'apprentissage
  - Difficulté de regarder à l'intérieur des boîtes noires
  - Beaucoup d'efforts pour gagner 0,10 % d'exactitude
  - Fossé creusé entre experts de l'apprentissage et linguistes

## Problématique

- Evolution du TAL français depuis 10 ans :
  - Avant : méthodes à base de règles et grammaires
  - Après : méthodes statistiques à base d'apprentissage
- Pour faire un étiqueteur statistique, prenez :
  - Un corpus annoté (ex. French Treebank)
  - Un classifieur (ex. SVM)
  - Un algorithme (ex. gauche à droite séquentiel)
  - Quelques descripteurs
- Atouts :
  - État de l'art facile à atteindre
  - Système facile à maintenir
- Sacrifices :
  - Dépendant du corpus d'apprentissage
  - Difficulté de regarder à l'intérieur des boîtes noires
  - Beaucoup d'efforts pour gagner 0,10 % d'exactitude
  - Fossé creusé entre experts de l'apprentissage et linguistes

## Problématique

- Evolution du TAL français depuis 10 ans :
  - Avant : méthodes à base de règles et grammaires
  - Après : méthodes statistiques à base d'apprentissage
- Pour faire un étiqueteur statistique, prenez :
  - Un corpus annoté (ex. French Treebank)
  - Un classifieur (ex. SVM)
  - Un algorithme (ex. gauche à droite séquentiel)
  - Quelques descripteurs
- Atouts :
  - État de l'art facile à atteindre
  - Système facile à maintenir
- Sacrifices :
  - Dépendant du corpus d'apprentissage
  - Difficulté de regarder à l'intérieur des boîtes noires
  - Beaucoup d'efforts pour gagner 0,10 % d'exactitude
  - Fossé creusé entre experts de l'apprentissage et linguistes

## Problématique : étiquetage morphosyntaxique de mots fonctionnels

- L'étiquetage morphosyntaxique atteint 97 % avec des descripteurs génériques
  - Facilement adaptable à n'importe quelle langue
  - Pas de connaissances linguistiques spécifiques → limite les gains d'exactitude possibles
- Mais les 3 % qui reste = 3 % erreurs en parsing
- Corriger des mots fonctionnels ambigus comme "que" par l'injection de connaissances linguistiques :
  - Des descripteurs ciblés
  - Des règles

## Problématique : étiquetage morphosyntaxique de mots fonctionnels

- L'étiquetage morphosyntaxique atteint 97 % avec des descripteurs génériques
  - Facilement adaptable à n'importe quelle langue
  - Pas de connaissances linguistiques spécifiques → limite les gains d'exactitude possibles
- Mais les 3 % qui reste = 3 % erreurs en parsing
- Corriger des mots fonctionnels ambigus comme "que" par l'injection de connaissances linguistiques :
  - Des descripteurs ciblés
  - Des règles

## Problématique : étiquetage morphosyntaxique de mots fonctionnels

- L'étiquetage morphosyntaxique atteint 97 % avec des descripteurs génériques
  - Facilement adaptable à n'importe quelle langue
  - Pas de connaissances linguistiques spécifiques → limite les gains d'exactitude possibles
- Mais les 3 % qui reste = 3 % erreurs en parsing
- Corriger des mots fonctionnels ambigus comme "que" par l'injection de connaissances linguistiques :
  - Des descripteurs ciblés
  - Des règles



- ① Étiquetage morphosyntaxique
- ② Des descripteurs ou des règles ?
- ③ Les descripteurs ciblés
- ④ Les règles
- ⑤ Conclusions et perspectives

# Étiquetage morphosyntaxique

Je	n'	ai	bu	<b>que</b>	trois	verres
CLS	ADV	V	VPP	?		

- **Étiqueteur** : Talismane (Urieli, 2013)
  - <http://redac.univ-tlse2.fr/talismane.html>
- **Corpus** :
  - Entraînement : SPMRL-fr (Seddah et al, 2013)
  - Evaluation : SPMRL + Sequoia (Candito et al, 2012)
- **Tagset** : Crabbé et Candito (2008)
- **Lexique** : Lefff (Sagot, 2010)
- **Apprentissage** :
  - SVM linéaire
  - Paramètres :  $\epsilon = 0,01$ ,  $C = 0,5$ , cutoff=3
  - Exactitude de 96,58 sur SPMRL-dev, et 96,55 sur SPMRL-test

# Étiquetage morphosyntaxique

Je	n'	ai	bu	<b>que</b>	trois	verres
CLS	ADV	V	VPP	?		

- **Étiqueteur** : Talismane (Urieli, 2013)
  - <http://redac.univ-tlse2.fr/talismane.html>
- **Corpus** :
  - Entraînement : SPMRL-fr (Seddah et al, 2013)
  - Evaluation : SPMRL + Sequoia (Candito et al, 2012)
- **Tagset** : Crabbé et Candito (2008)
- **Lexique** : Lefff (Sagot, 2010)
- **Apprentissage** :
  - SVM linéaire
  - Paramètres :  $\epsilon = 0,01$ ,  $C = 0,5$ , cutoff=3
  - Exactitude de 96,58 sur SPMRL-dev, et 96,55 sur SPMRL-test

# Étiquetage morphosyntaxique

Je	n'	ai	bu	<b>que</b>	trois	verres
CLS	ADV	V	VPP	?		

- **Étiqueteur** : Talismane (Urieli, 2013)
  - <http://redac.univ-tlse2.fr/talismane.html>
- **Corpus** :
  - Entraînement : SPMRL-fr (Seddah et al, 2013)
  - Evaluation : SPMRL + Sequoia (Candito et al, 2012)
- **Tagset** : Crabbé et Candito (2008)
- **Lexique** : Lefff (Sagot, 2010)
- **Apprentissage** :
  - SVM linéaire
  - Paramètres :  $\epsilon = 0,01$ ,  $C = 0,5$ , cutoff=3
  - Exactitude de 96,58 sur SPMRL-dev, et 96,55 sur SPMRL-test

# Étiquetage morphosyntaxique

Je	n'	ai	bu	<b>que</b>	trois	verres
CLS	ADV	V	VPP	?		

- **Étiqueteur** : Talismane (Urieli, 2013)
  - <http://redac.univ-tlse2.fr/talismane.html>
- **Corpus** :
  - Entraînement : SPMRL-fr (Seddah et al, 2013)
  - Evaluation : SPMRL + Sequoia (Candito et al, 2012)
- **Tagset** : Crabbé et Candito (2008)
- **Lexique** : Lefff (Sagot, 2010)
- **Apprentissage** :
  - SVM linéaire
  - Paramètres :  $\epsilon = 0,01$ ,  $C = 0,5$ , cutoff=3
  - Exactitude de 96,58 sur SPMRL-dev, et 96,55 sur SPMRL-test

# Étiquetage morphosyntaxique

Je	n'	ai	bu	<b>que</b>	trois	verres
CLS	ADV	V	VPP	?		

- **Étiqueteur** : Talismane (Urieli, 2013)
  - <http://redac.univ-tlse2.fr/talismane.html>
- **Corpus** :
  - Entraînement : SPMRL-fr (Seddah et al, 2013)
  - Evaluation : SPMRL + Sequoia (Candito et al, 2012)
- **Tagset** : Crabbé et Candito (2008)
- **Lexique** : Lefff (Sagot, 2010)
- **Apprentissage** :
  - SVM linéaire
  - Paramètres :  $\epsilon = 0,01$ ,  $C = 0,5$ , cutoff=3
  - Exactitude de 96,58 sur SPMRL-dev, et 96,55 sur SPMRL-test

# Étiquetage morphosyntaxique

Je	n'	ai	bu	<b>que</b>	trois	verres
CLS	ADV	V	VPP	?		

- **Étiqueteur** : Talismane (Urieli, 2013)
  - <http://redac.univ-tlse2.fr/talismane.html>
- **Corpus** :
  - Entraînement : SPMRL-fr (Seddah et al, 2013)
  - Evaluation : SPMRL + Sequoia (Candito et al, 2012)
- **Tagset** : Crabbé et Candito (2008)
- **Lexique** : Lefff (Sagot, 2010)
- **Apprentissage** :
  - SVM linéaire
  - Paramètres :  $\epsilon = 0,01$ ,  $C = 0,5$ , cutoff=3
  - Exactitude de 96,58 sur SPMRL-dev, et 96,55 sur SPMRL-test

## Descripteurs de base :

- $W$  : la forme lexicale
- $P$  : le pos-tag
- $L$  : le lemme
- $U$  : si le token est inconnu
- $Sfx_n / Pref_n$  : les  $n$  dernières/premières lettres
- $1st / Last$  : si le token est le premier/dernier de la phrase
- 2- et 3-grammes aux tokens en position -2, -1, 0, +1, +2

## Descripteurs (plus) "riches" :

- **Distance** < -2 ou 2 < **distance**
- Lemme < **classe d'équivalence** < pos-tag



## Descripteurs de base :

- $W$  : la forme lexicale
- $P$  : le pos-tag
- $L$  : le lemme
- $U$  : si le token est inconnu
- $Sfx_n / Pref_n$  : les  $n$  dernières/premières lettres
- $1st / Last$  : si le token est le premier/dernier de la phrase
- 2- et 3-grammes aux tokens en position -2, -1, 0, +1, +2

## Descripteurs (plus) "riches" :

- **Distance**  $< -2$  ou  $2 < \mathbf{distance}$
- Lemme  $< \mathbf{classe d'équivalence} < \text{pos-tag}$

## Étiquetage morphosyntaxique de “que”

- 1 Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- 2 Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- 3 Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- 4 Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- 5 Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- 6 Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées

## Étiquetage morphosyntaxique de “que”

- 1 Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- 2 Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- 3 Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- 4 Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- 5 Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- 6 Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées

## Étiquetage morphosyntaxique de “que”

- 1 Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- 2 Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- 3 Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- 4 Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- 5 Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- 6 Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées

## Étiquetage morphosyntaxique de “que”

- 1 Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- 2 Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- 3 Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- 4 Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- 5 Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- 6 Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées

## Étiquetage morphosyntaxique de “que”

- ➊ Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- ➋ Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- ➌ Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- ➍ Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- ➎ Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- ➏ Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées

## Étiquetage morphosyntaxique de “que”

- 1 Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- 2 Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- 3 Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- 4 Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- 5 Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- 6 Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées

## Étiquetage morphosyntaxique de “que”

- ➊ Conjonction de subordination (CS) : *Je pense **qu'**il a trop bu.*
- ➋ Pronom relatif (PROREL) : *Il boit le vin **que** j'ai acheté.*
- ➌ Pronom interrogatif (PROWH) : ***Que** buvez-vous ?*
- ➍ Adverbe négatif (ADV) : *Je n'ai bu **que** trois verres.*
- ➎ Adverbe exclamatif (ADV) : ***Qu'**il est bon, ce vin !*
- ➏ Construction comparative (CS) : *Il est plus bourré **que** moi.*

+ des clivées



## Baseline pour "que"

	ADV	CS	PROREL	PROWH	Total	Erreurs
ADV	90	<b>44</b>	4	1	139	<b>49</b>
CS	<b>37</b>	1097	<b>61</b>	0	1195	<b>98</b>
PROREL	0	<b>69</b>	244	0	313	<b>69</b>
PROWH	0	4	2	23	29	6

TABLE : Matrice de confusion de base pour *que*

- 222 erreurs pour 1 676 occurrences
- Exactitude de 86,75 %

## Des descripteurs ou des règles ?

- Descripteurs :
  - Information extraite du contexte
  - Définissent des tendances
  - Peuvent se contredire
  - S'appliquent pendant l'entraînement et l'analyse
  - Généralisent les régularités du corpus d'entraînement aux autres corpus
  - Limités au cas trouvés dans le corpus d'entraînement
- Règles :
  - Vrai/faux : si vrai, forcer ou interdire une étiquette
  - Définissent des vérités
  - Déterministes
  - S'appliquent uniquement pendant l'analyse
  - Peuvent être spécifiques à chaque corpus
  - Traduisent les connaissances de la langue ou du corpus analysé

## Des descripteurs ou des règles ?

- Descripteurs :
  - Information extraite du contexte
  - Définissent des tendances
  - Peuvent se contredire
  - S'appliquent pendant l'entraînement et l'analyse
  - Généralisent les régularités du corpus d'entraînement aux autres corpus
  - Limités au cas trouvés dans le corpus d'entraînement
- Règles :
  - Vrai/faux : si vrai, forcer ou interdire une étiquette
  - Définissent des vérités
  - Déterministes
  - S'appliquent uniquement pendant l'analyse
  - Peuvent être spécifiques à chaque corpus
  - Traduisent les connaissances de la langue ou du corpus analysé

## “Que” comme adverbe négatif : analyse d’erreurs

Dans le corpus SPMRL dev :

- ① *Mais cela **ne** représente dans cette mouture, pour un couple avec deux enfants, **qu’**/ADV une prime maximale.*
- ② *Pour cela, il **n’est pas** question **que**/CS le zloty, la monnaie polonaise, soit “l’ancre de la stabilité” de l’économie polonaise.*
- ③ *... qui, faute de volonté politique, **ne fut jamais** **que**/ADV la caricature du système français.*

## “Que” comme adverbe négatif : analyse d'erreurs

Dans le corpus SPMRL dev :

- ① *Mais cela **ne** représente dans cette mouture, pour un couple avec deux enfants, **qu'**/ADV une prime maximale.*
- ② *Pour cela, il **n'est pas** question **que**/CS le zloty, la monnaie polonaise, soit “l'ancre de la stabilité” de l'économie polonaise.*
- ③ *... qui, faute de volonté politique, **ne fut jamais** que/ADV la caricature du système français.*

## “Que” comme adverbe négatif : analyse d’erreurs

Dans le corpus SPMRL dev :

- ➊ *Mais cela **ne** représente dans cette mouture, pour un couple avec deux enfants, **qu’**/ADV une prime maximale.*
- ➋ *Pour cela, il **n’est pas** question **que**/CS le zloty, la monnaie polonaise, soit “l’ancre de la stabilité” de l’économie polonaise.*
- ➌ *... qui, faute de volonté politique, **ne fut jamais que**/ADV la caricature du système français.*

## “Que” comme adverbe négatif : descripteurs ciblés

- Ne précèdent sans autre particule négative
- Pas de *ne* précédent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
- Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*

## “Que” comme adverbe négatif : descripteurs ciblés

- *Ne* précèdent sans autre particule négative
- Pas de *ne* précèdent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
- Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*



## “Que” comme adverbe négatif : descripteurs ciblés

- *Ne* précèdent sans autre particule négative
- Pas de *ne* précèdent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
  - Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*

## “Que” comme adverbe négatif : descripteurs ciblés

- *Ne* précèdent sans autre particule négative
- Pas de *ne* précèdent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
- Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*

## “Que” comme adverbe négatif : descripteurs ciblés

- *Ne* précèdent sans autre particule négative
- Pas de *ne* précèdent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
- Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*

## “Que” comme adverbe négatif : descripteurs ciblés

- *Ne* précèdent sans autre particule négative
- Pas de *ne* précèdent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
- Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*

## “Que” comme adverbe négatif : descripteurs ciblés

- Ne précèdent sans autre particule négative
- Pas de *ne* précèdent
- *Que* négatif possible : y a-t-il un *ne*, *ne pas* ou *ne plus* plus tôt dans la phrase, sans prendre en compte d'autres particules négatives
- Combinaison de particules négatives à courte distance

Mais :

- ① *L'Amérique, moins superficielle qu'on ne l'imagine parfois, a entrepris une réflexion sur son identité bien avant que/CS [...]*
- ② *[...] ne peuvent ainsi éviter que/CS, en la matière, l'histoire ne se repète*
- ③ *Il lui manque que/ADV le sac à main de Maggie*

## “Que” comme pronom relatif : analyse d'erreurs

- ① [...] *la Commission des opérations de bourse (COB) a annoncé le 14 janvier qu' /CS elle saisit la justice [...]*
- ② *Le gouvernement va présenter dans un délai de trois mois les dispositions qu' /PROREL il entend retenir [...]*
- ③ *Le fait qu' /CS ils aient accepté de reprendre les pourparlers est interprété de façon positive.*

## “Que” comme pronom relatif : analyse d'erreurs

- ① [...] *la Commission des opérations de bourse (COB) a annoncé le 14 janvier **qu'**/CS elle saisit la justice [...]*
- ② *Le gouvernement va présenter dans un délai de trois mois les dispositions **qu'**/PROREL il entend retenir [...]*
- ③ *Le fait **qu'**/CS ils aient accepté de reprendre les pourparlers est interprété de façon positive.*

## “Que” comme pronom relatif : analyse d'erreurs

- ① [...] *la Commission des opérations de bourse (COB) a annoncé le 14 janvier qu' /CS elle saisit la justice [...]*
- ② *Le gouvernement va présenter dans un délai de trois mois les dispositions qu' /PROREL il entend retenir [...]*
- ③ *Le fait qu' /CS ils aient accepté de reprendre les pourparlers est interprété de façon positive.*



## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots {*assurance, certitude, doute, espoir, fait, ...*} ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
- **Le verbe qui précède a un objet direct**
- **Le verbe qui suit a un objet direct**
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- **Que suivi d'un verbe subjonctif**
- ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots { *assurance, certitude, doute, espoir, fait, ...* } ?
- Verbe précédent sous catégorise avec *que* : ex. *penser*
- Le verbe qui précède a un objet direct
- Le verbe qui suit a un objet direct
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- **Que suivi d'un verbe subjonctif**
- ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots {*assurance, certitude, doute, espoir, fait, ...*} ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
  - Le verbe qui précède a un objet direct
  - Le verbe qui suit a un objet direct
  - **Que suivi directement d'un verbe**
    - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
  - **Que suivi d'un verbe subjonctif**
  - ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots { *assurance, certitude, doute, espoir, fait, ...* } ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
- **Le verbe qui précède a un objet direct**
- **Le verbe qui suit a un objet direct**
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- **Que suivi d'un verbe subjonctif**
- ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots {*assurance, certitude, doute, espoir, fait, ...*} ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
- **Le verbe qui précède a un objet direct**
- **Le verbe qui suit a un objet direct**
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- **Que suivi d'un verbe subjonctif**
- ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots { *assurance, certitude, doute, espoir, fait, ...* } ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
- **Le verbe qui précède a un objet direct**
- **Le verbe qui suit a un objet direct**
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- Que suivi d'un verbe subjonctif
- ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots { *assurance, certitude, doute, espoir, fait, ...* } ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
- **Le verbe qui précède a un objet direct**
- **Le verbe qui suit a un objet direct**
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- **Que suivi d'un verbe subjonctif**
- ...

## “Que” comme pronom relatif : descripteurs ciblés

- **Structure coordonnée** : chercher l'étiquette du *que* précédent.
- **Après nom explicatif** : Le *que* suit-il un des mots { *assurance, certitude, doute, espoir, fait, ...* } ?
- **Verbe précédent sous catégorise avec que** : ex. *penser*
- **Le verbe qui précède a un objet direct**
- **Le verbe qui suit a un objet direct**
- **Que suivi directement d'un verbe**
  - *L'exemple que/PROREL fournit Sombart est particulièrement éclairant*
- **Que suivi d'un verbe subjonctif**
- ...



## Descripteurs ciblés : résultats

	ADV	CS	PROREL	PROWH	Total	Erreurs
ADV	133 (+43)	6 (-38)	0 (-4)	0 (-1)	139	6 (-43)
CS	10 (-27)	1135 (+38)	50 (-11)	0 (-1)	1195	60 (-38)
PROREL	0	52 (-17)	261 (+17)	0	313	52 (-17)
PROWH	0	0 (-4)	4 (+2)	25 (+2)	29	4 (-2)

TABLE : Matrice de confusion pour *que* avec les descripteurs ciblés

- 139 nouvelles corrections pour 29 nouvelles erreurs
- 122 erreurs pour 1 676 occurrences
- Exactitude de 86,75 % à 92,72 %
- Taux d'erreur -45 %

## Les règles

- CS : après *s'attendre à ce que*, *veiller à ce que*, *n'empêche que*...
- CS : pour toute expression de type "*être ADJ que*", tel que "*il est probable que*".
- !PROREL : si *que* suit un verbe directement, ou séparé du verbe uniquement par un commentaire entouré par des virgules
  - Sequoia/Europarl : *Je sais, Madame la Présidente, que/CS vous êtes déjà intervenue [...]*
- ...

## Les règles

- CS : après *s'attendre à ce que*, *veiller à ce que*, *n'empêche que*...
- CS : pour toute expression de type "*être ADJ que*", tel que "*il est probable que*".
- !PROREL : si *que* suit un verbe directement, ou séparé du verbe uniquement par un commentaire entouré par des virgules
  - Sequoia/Europarl : *Je sais, Madame la Présidente, que/CS vous êtes déjà intervenue [...]*
- ...

## Les règles

- CS : après *s'attendre à ce que*, *veiller à ce que*, *n'empêche que*. . .
- CS : pour toute expression de type "*être ADJ que*", tel que "*il est probable que*".
- !PROREL : si *que* suit un verbe directement, ou séparé du verbe uniquement par un commentaire entouré par des virgules
  - Sequoia/Europarl : *Je sais, Madame la Présidente, que/CS vous êtes déjà intervenue [. . .]*
  - . . .

## Les règles

- CS : après *s'attendre à ce que*, *veiller à ce que*, *n'empêche que*. . .
- CS : pour toute expression de type "*être ADJ que*", tel que "*il est probable que*".
- !PROREL : si *que* suit un verbe directement, ou séparé du verbe uniquement par un commentaire entouré par des virgules
  - Sequoia/Europarl : *Je sais, Madame la Présidente, que/CS vous êtes déjà intervenue [. . .]*
- . . .

## Les règles : résultats

	ADV	CS	PROREL	PROWH	Total	Erreurs
ADV	134 (+1)	5 (-1)	0	0	139	5 (-1)
CS	10	1149 (+14)	36 (-14)	0	1195	46 (-14)
PROREL	0	48 (-4)	265 (+4)	0	313	48 (-4)
PROWH	0	0	2 (-2)	27 (+2)	29	2 (-2)

TABLE : Matrice de confusion pour *que* avec les règles

- 21 nouvelles corrections pour 0 nouvelles erreurs
- 101 erreurs pour 1 676 occurrences
- Exactitude de 86,75 % à 93,97 %
- Taux d'erreur -55 %

## Les règles : corpus non annotés

200 000 mots de chacun des corpus suivants :

- Est Républicain : le journal régional Est Républicain de l'année 2003
- Leximedia : campagne présidentielle 2007 (CLLE-ERSS)
- Frantext : des textes littéraires français du 20ème siècle
- Revues.org : une collection d'articles scientifiques dans les sciences sociales

Résultats :

- Une différence / 8 500 mots
- 46 corrections pour 5 erreurs dans les 51 premières
- *Je conteste, en tant que père de famille, que/CS l'on vienne me dire que l'argent est le corollaire du succès.*

## Les règles : corpus non annotés

200 000 mots de chacun des corpus suivants :

- Est Républicain : le journal régional Est Républicain de l'année 2003
- Leximedia : campagne présidentielle 2007 (CLLE-ERSS)
- Frantext : des textes littéraires français du 20ème siècle
- Revues.org : une collection d'articles scientifiques dans les sciences sociales

Résultats :

- Une différence / 8 500 mots
- 46 corrections pour 5 erreurs dans les 51 premières
- *Je **conteste**, en tant que père de famille, **que**/CS l'on vienne me dire que l'argent est le corollaire du succès.*



## Conclusions et perspectives

- Descripteurs ciblés : taux d'erreur -45 %
- + règles : taux d'erreur -55 %
- Complexité de maintenance ?
- Vitesse d'analyse ? (de 9 à 5 millions de mots par heure)
- D'autres mots fonctionnels ("de/des/du", "soit", ...)
- Appliquer au parsing

## Conclusions et perspectives

- Descripteurs ciblés : taux d'erreur -45 %
- + règles : taux d'erreur -55 %
- Complexité de maintenance ?
- Vitesse d'analyse ? (de 9 à 5 millions de mots par heure)
- D'autres mots fonctionnels ("de/des/du", "soit", ...)
- Appliquer au parsing

## Conclusions et perspectives

- Descripteurs ciblés : taux d'erreur -45 %
- + règles : taux d'erreur -55 %
- Complexité de maintenance ?
- Vitesse d'analyse ? (de 9 à 5 millions de mots par heure)
- D'autres mots fonctionnels ("de/des/du", "soit", ...)
- Appliquer au parsing

## Conclusions et perspectives

- Descripteurs ciblés : taux d'erreur -45 %
- + règles : taux d'erreur -55 %
- Complexité de maintenance ?
- Vitesse d'analyse ? (de 9 à 5 millions de mots par heure)
- D'autres mots fonctionnels ("de/des/du", "soit", ...)
- Appliquer au parsing

## Conclusions et perspectives

- Descripteurs ciblés : taux d'erreur -45 %
- + règles : taux d'erreur -55 %
- Complexité de maintenance ?
- Vitesse d'analyse ? (de 9 à 5 millions de mots par heure)
- D'autres mots fonctionnels ("de/des/du", "soit", ...)
- Appliquer au parsing

## Conclusions et perspectives

- Descripteurs ciblés : taux d'erreur -45 %
- + règles : taux d'erreur -55 %
- Complexité de maintenance ?
- Vitesse d'analyse ? (de 9 à 5 millions de mots par heure)
- D'autres mots fonctionnels ("de/des/du", "soit", ...)
- Appliquer au parsing