

TYPOLOGIE DES SUBORDONNÉES ET DES CONNECTEURS

EN VUE DE LA DÉTECTION
AUTOMATIQUE DES
PROPOSITIONS SYNTAXIQUES
DU FRANÇAIS

Yayoi NAKAMURA-DELLOYE
Paris 7 - LATTICE (CNRS UMR 8094)
yayoi@free.fr

PLAN

- ✻ Contexte : alignement et détection des propositions
- ✻ Problèmes des étiquettes classiques des connecteurs et des subordonnées
- ✻ Typologie des subordonnées selon la position
- ✻ Réalisation et évaluation
- ✻ Perspectives
- ✻ Conclusion

Contexte :

alignement et
détection des
propositions

Problèmes

des étiquettes
classiques des
connecteurs et des
subordonnées

Typologie

des subordonnées
selon la position

Réalisation,

évaluation

Pistes

d'amélioration :
affinement des
étiquettes

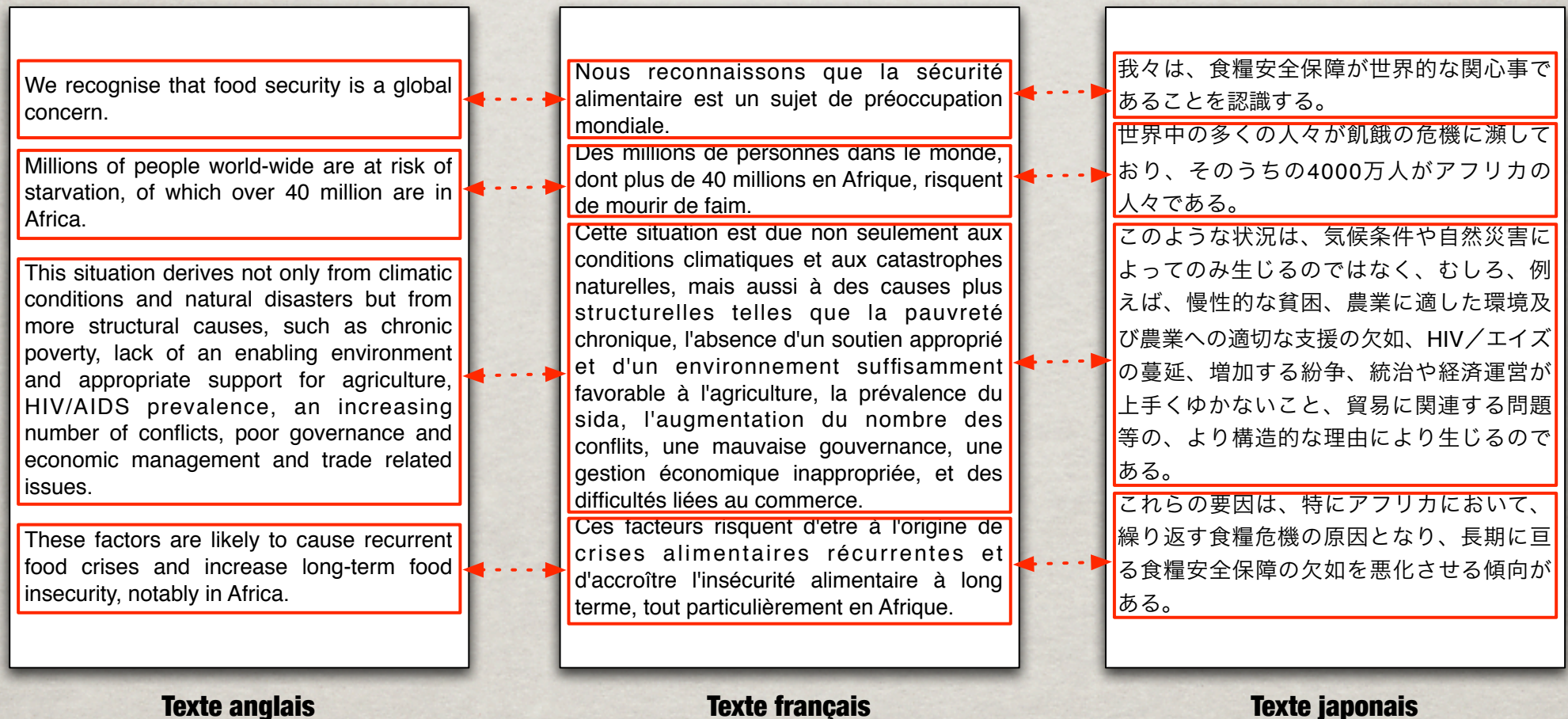
Conclusion

CONTEXTE

- ✿ **Typologie des subordonnées et des connecteurs**
en vue de la détection automatique des propositions
- ✿ **Détection automatique des propositions** dans le
cadre de :
 - la documentation technique ;
 - l'analyse discursive ;
 - ...
 - l'alignement des propositions des textes
parallèles français-japonais

CONTEXTE

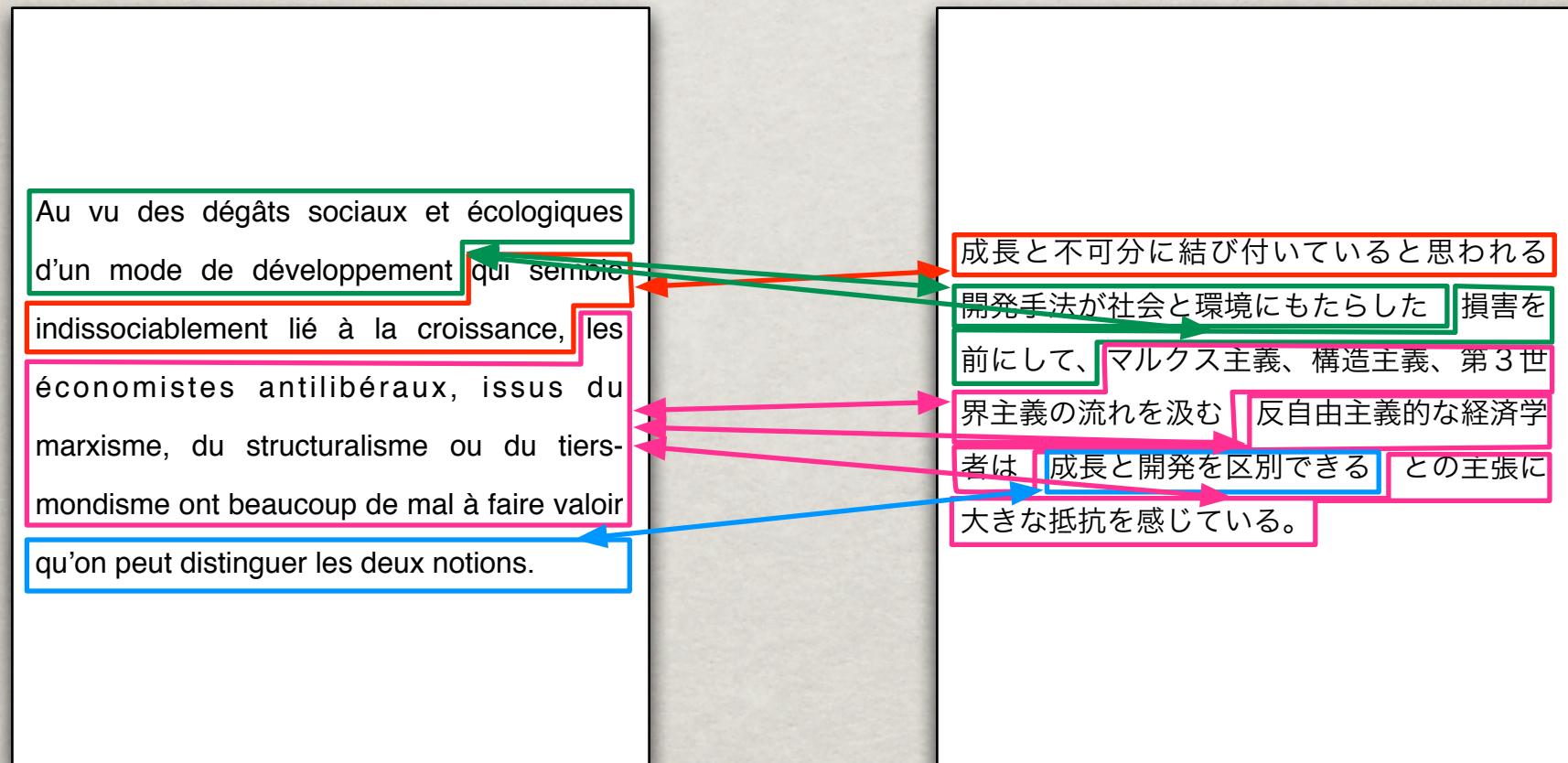
ALIGNEMENT DES TEXTES PARALLÈLES



Parallélisme

CONTEXTE

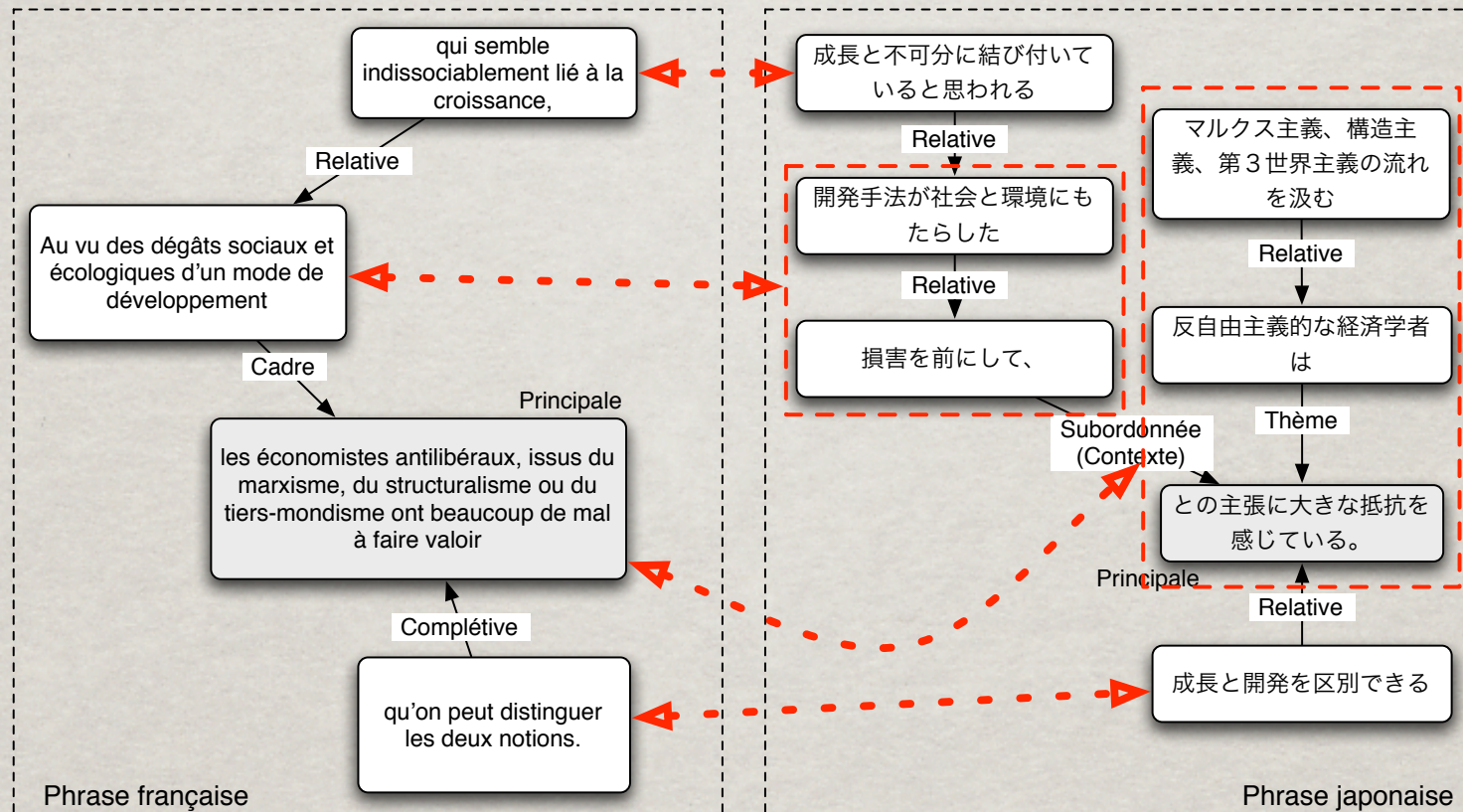
ALIGNEMENT DES TEXTES PARALLÈLES



Non-Parallélisme

CONTEXTE

ALIGNEMENT DES PROPOSITIONS



- ✿ non seulement de la détection des frontières de propositions,
- ✿ mais aussi de leur mise en relation.

CONTEXTE

NOTION PRINCIPALE

Proposition = sujet + prédicat

✻ Types de proposition :

1. Principale (ou racine)

2. Coordonnée

Mon père est professeur et ma mère travaille à la banque.

3. Subordonnée

Il était déjà rentré quand je suis arrivé.

4. Incidente

Il s'en est, me semble-t-il, bien sorti.

+ Éléments extra-prédicatifs (Charolles 1997) (Combettes 1998)

L'autre jour, ...

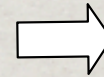
Cette affaire étant réglée, ...

En ce qui concerne X, ...

CONTEXTE PRÉ-TRAITEMENT

Texte source

Si ces chiffres peuvent susciter l'étonnement, la triste vérité est que les habitants de Reay Road et des autres poches de misère qui prolifèrent n'ont pas mieux où aller.



Pré-traitement

Tagger



Chunker

CS NP_mp VP_fin_p VP_inf
NP_ms VIRGULE NP_fs
VP_fin_s CS NP_mp PP_ms
CC P+D-mp NP_mp PP_fs
Pro_rel VP_fin_s AdvP
VP_fin_p AdvP Pro_int
VP_inf PONCT-S

Résultat

A : Condition

Si ces chiffres peuvent susciter l'étonnement,

Principale

[A] la triste vérité est [B]

B : Complétive

que les habitants de Reay Road et des autres poches de misère [C] n'ont pas mieux où aller

C : Relative

qui prolifèrent

Contexte :

alignement et
détection des
propositions

Problèmes

des étiquettes
classiques des
connecteurs et des
subordonnées

Typologie

des subordonnées
selon la position

Réalisation,

évaluation

Pistes

d'amélioration :
affinement des
étiquettes

Conclusion

PROBLÈMES DES ÉTIQUETTES CLASSIQUES DES CONNECTEURS ET DES SUBORDONNÉES

✻ Quels sont les «connecteurs» du français ?
(Wagner et Pinchon, 1991)

1. des conjonctions (**que, comme, quand, si**) et des locutions conjonctives ;
2. des adverbes et des pronoms interrogatifs (**quand, comment, où, pourquoi**) ;
3. des pronoms relatifs ;
4. des adverbes de quantité simples (**tant, tellement**).

PROBLÈMES DES ÉTIQUETTES CLASSIQUES

✻ Toutes ces distinctions de catégories sont-elles pertinentes pour l'opération de détection des propositions ?

1. difficulté de l'étiquetage ;
2. comportements syntaxiques différents de termes appartenant à la même catégorie

1. *Le gouvernement a retiré sa proposition **dont** la conformité à la Constitution avait été remise en cause.*
2. *À cette occasion, se sont réunis huit représentants **dont** notre Président.*

✻ Qu'est-ce qu'une locution conjonctive (LC)? Les LCs constituent-elles une liste fermée ?

TRAVAUX DE LE GOFFIC SUR LES TERMES EN «QU-»

(Le Goffic 1992, 1993, 2002)

☼ **Termes en «qu-»** = seuls connecteurs du français

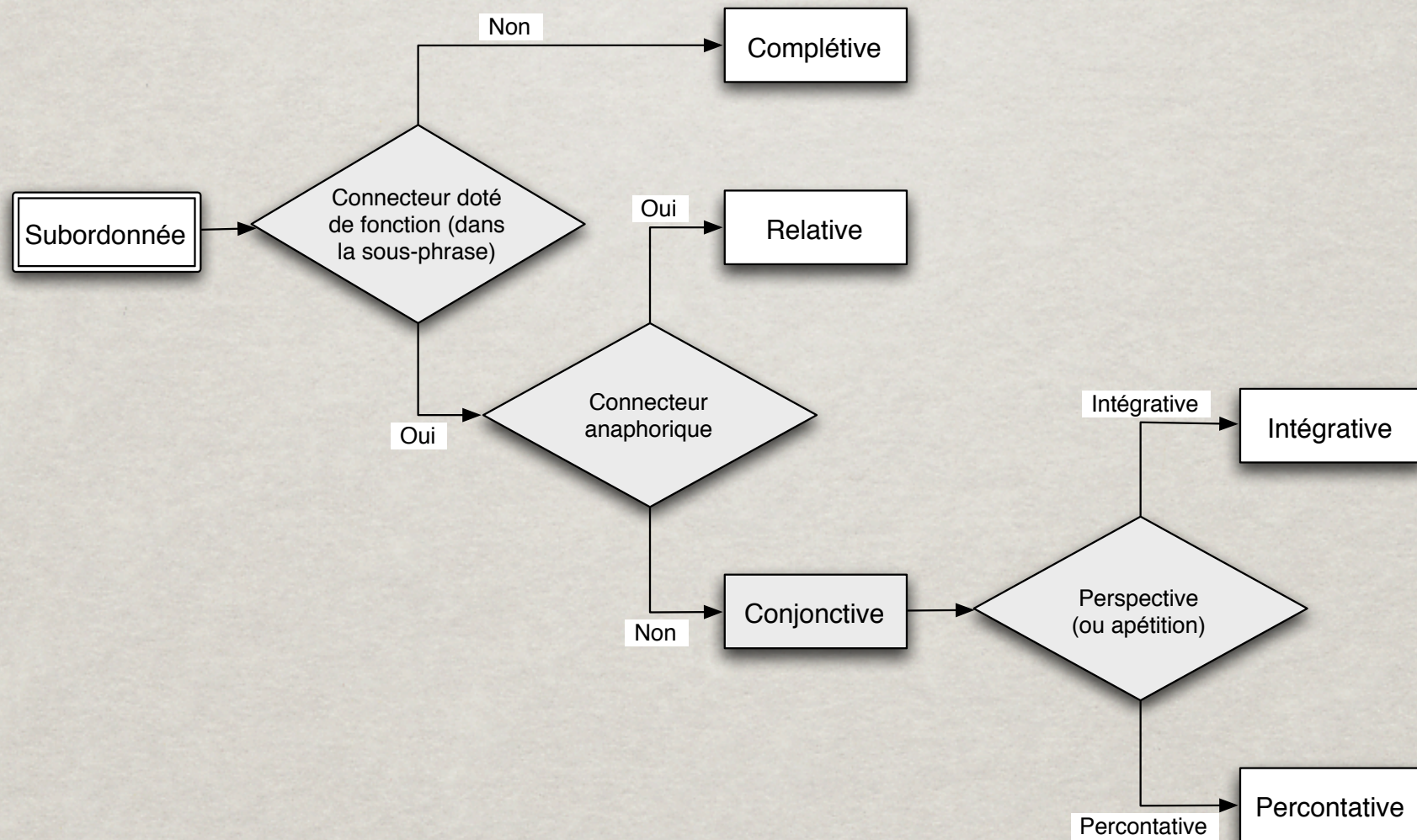
Une vieille famille indo-européenne en «*Kw-*» :

- **pronoms** : qui, que, quoi, lequel ;
- **adjectif** : quel ;
- **adverbes** : où, quand, comme, comment, combien, que, dont, pourquoi.

☼ Subordonnées introduites par une LC = des GAdv ou GPrép comprenant une subordonnée constituée par un connecteur en «qu-»

➡ un traitement unifié et homogène des types de subordonnées

TYPOLOGIE DES SUBORDONNÉES SELON LE GOFFIC



TYPOLOGIE DES SUBORDONNÉES SELON LE GOFFIC

☼ Complétive : complétive

Je crois qu'il va pleuvoir...

☼ Relative : relative avec antécédent

Le médecin qui est venu / la maison où je suis né...

☼ Intégrative :

- Pronominale : relative sans antécédent

Qui dort dîne / embrassez qui vous voulez...

- Adverbiale : circonstancielle en qu- ou si

Quand on veut, on peut / si vous avez fini, vous pouvez sortir / il est à peine sorti qu'il a commencé à pleuvoir...

☼ Percontative : interrogative/exclamative indirecte

Je sais qui a gagné / où il est allé / comment il l'a fait...

AUTRES TYPOLOGIES

☼ Typologie selon la **catégorie** du mot simple équivalent : (*Le Bon Usage*, 11ème éd.) (Biskri et Desclés 2005)

- Substantive : *Je pense qu'il viendra / Que tu m'aimes me réjouit*
- Adjective : *La femme que tu vois / la ville où j'habite*
- Adverbiale : *Il était déjà rentré quand je suis arrivé*

☼ Typologie selon la **fonction** de la subordonnée dans la principale : (Chevalier *et al.* 1964)(Grevisse 1969)(Wilmet 1997)

- Sujet : *Que je sois malade ne l'a jamais effleuré*
- Attribut : *La triste vérité est qu'il est fou*
- Objet : *Marie sait que Paul viendra*
- Circonstancielle : *Il était déjà rentré quand je suis arrivé*
- Complément de nom : *la certitude que son but était atteint*
- ... etc.

Contexte :

alignement et
détection des
propositions

Problèmes

des étiquettes
classiques des
connecteurs et des
subordonnées

Typologie

des subordonnées
selon la position

Réalisation,

évaluation

Pistes

d'amélioration :
affinement des
étiquettes

Conclusion

TYPOLOGIE DES SUBORDONNÉES

SELON LA POSITION D'APPARITION

- ☼ **Post-verbale** : (subordonnées compléments)
 - ❖ concerne les propositions substantives ;
- ☼ **Initiale/Finale** : (subordonnées circonstancielles)
 - ❖ concerne les propositions adverbiales
- ☼ **Post-nominale** : (subordonnées déterminantes)
 - ❖ concerne les propositions adjectives et adverbiales ;
- ☼ **Autres positions pouvant être occupées par un SN** : (subordonnées sujet, prépositionnelles, en prolepse)
 - ❖ concerne les propositions substantives.

POSITION POST-VERBALE

(SUBORDONNÉE COMPLÉMENT : subQ)

Subordonnées substantives :

- ✻ Complétives
 - *Je pense qu'il viendra.*
- ✻ Intégratives pronominales (relatives sans ant.)
 - *Il a le droit d'embrasser qui il veut.*
- ✻ Percontatives (ou interrogatives)
 - *Je me demande s'il est parti.*
 - *Il ne m'a pas dit quand il rentrerait.*
 - *Voyez comme c'est facile.*

Post-
verbale :
Substantives

Initiale/
Finale :
Adverbiales

Post-
nominale :
Adjectives
Adverbiales

Autres
positions
SN :
Substantives

POSITION INITIALE/FINALE (SUBORDONNÉE CIRCONSTANCIELLE : subP)

- *Quand je suis arrivé, il était déjà rentré.*
- *Si tu ne manges pas, tu ne guériras pas.*
- *Comme elle est écrite en chinois, il n'a pas pu lire cette lettre.*
- *Où il y a de la gêne, il n'y a pas de plaisir.*
- *Que le gouvernement propose une nouvelle loi, l'opposition crie au scandale.*
- *Il n'a pas pu lire cette lettre comme sa mère l'a deviné.*
- *Tu peux poser ton manteau où tu veux.*
- *Je pars, que cela vous plaise ou non.*
- *Le crocodile n'eut pas le temps de se demander ce que lui voulait ce lourdaud, que Gropopotin s'était déjà assis sur son dos.*
- *La maison est restée aussi conviviale qu'elle l'était avant.*

Post-
verbale :
Substantives

Initiale/
Finale :
Adverbiales

Post-
nominale :
Adjectives
Adverbiales

Autres
positions
SN :
Substantives

POSITION POST-NOMINALE (SUBORDONNÉE DÉTERMINANTE : subR)

✻ Adjectives : (relatives)

- *La peinture qui m'a fascinée*
- *La peinture dans laquelle notre maison était reproduite*
- *Ce à quoi je m'attendais*
- *L'idée que tout est fini.*

✻ Adverbiales : (intégratives adverbiales)

- *La déception du père quand il a entendu cette nouvelle*
- *La fille comme je l'avais imaginée*

Post-
verbale :
Substantives

Initiale/
Finale :
Adverbiales

Post-
nominale :
Adjectives
Adverbiales

Autres
positions
SN :
Substantives

AUTRES POSITION SN

(SUBORDONNÉE SUBSTANTIVE : subSN)

☼ Position Sujet :

- *Qui dort dîne.* (Intégrative pronominale)
- *Que vous avez menti me déçoit.* (Complétive)
- *Comment il a commis ce crime n'a jamais été établi.* (Percontative)

☼ Après une préposition :

- *Je voterai pour qui me promettra moins d'impôts.* (Intég.pron.)
- *Je partirai avant qu'il arrive.* (Complétive)
- *Il faudra se poser la question de pourquoi nous avons été choisis.* (Percontative)

☼ Position initiale en prolepse :

- *Comment il a fait, je vous le demande !* (Percontative)

Post-
verbale :
Substantives

Initiale/
Finale :
Adverbiales

Post-
nominale :
Adjectives
Adverbiales

Autres
positions
SN :
Substantives

DE LA TYPOLOGIE DES SUBORDONNÉES À LA TYPOLOGIE DES CONNECTEURS

		Int/Fin	post-V	post-N	Autres SN
Sub.	Intégrative pro.		△		△
	Percontative		✓		△
	Complétive		✓		△
Adj.				✓	
	Relative			✓	
Adv.	Intégrative adv.	✓		△	

✓ = fréquent ; △ = moins fréquent / rare

DE LA TYPOLOGIE DES SUBORDONNÉES À LA TYPOLOGIE DES CONNECTEURS

Connecteurs
isolés

Connecteurs
ambigus

Connecteurs
relatifs

Indicateurs de
proposition

Position	post-V				post-N				Int /Fin				pos. SN			
	I	C	P	R	I	C	P	R	I	C	P	R	I	C	P	R
qui	△		✓					✓					△		△	
que		✓				✓		✓	✓					△		
dont								✓								
où			✓		?			✓	△						△	
quand			✓		△				✓						△	
comme			✓		△				✓						△	
si			✓				△		✓						△	
quoi			✓					✓							△	
lequel			✓					✓							△	
quel			✓												△	
combien			✓												△	
comment			✓												△	
pourquoi			✓												△	

I = Intégrative, C = Complétive, P = Percontative, R = Relative

Contexte :

alignement et
détection des
propositions

Problèmes

des étiquettes
classiques des
connecteurs et des
subordonnées

Typologie

des subordonnées
selon la position

Réalisation,

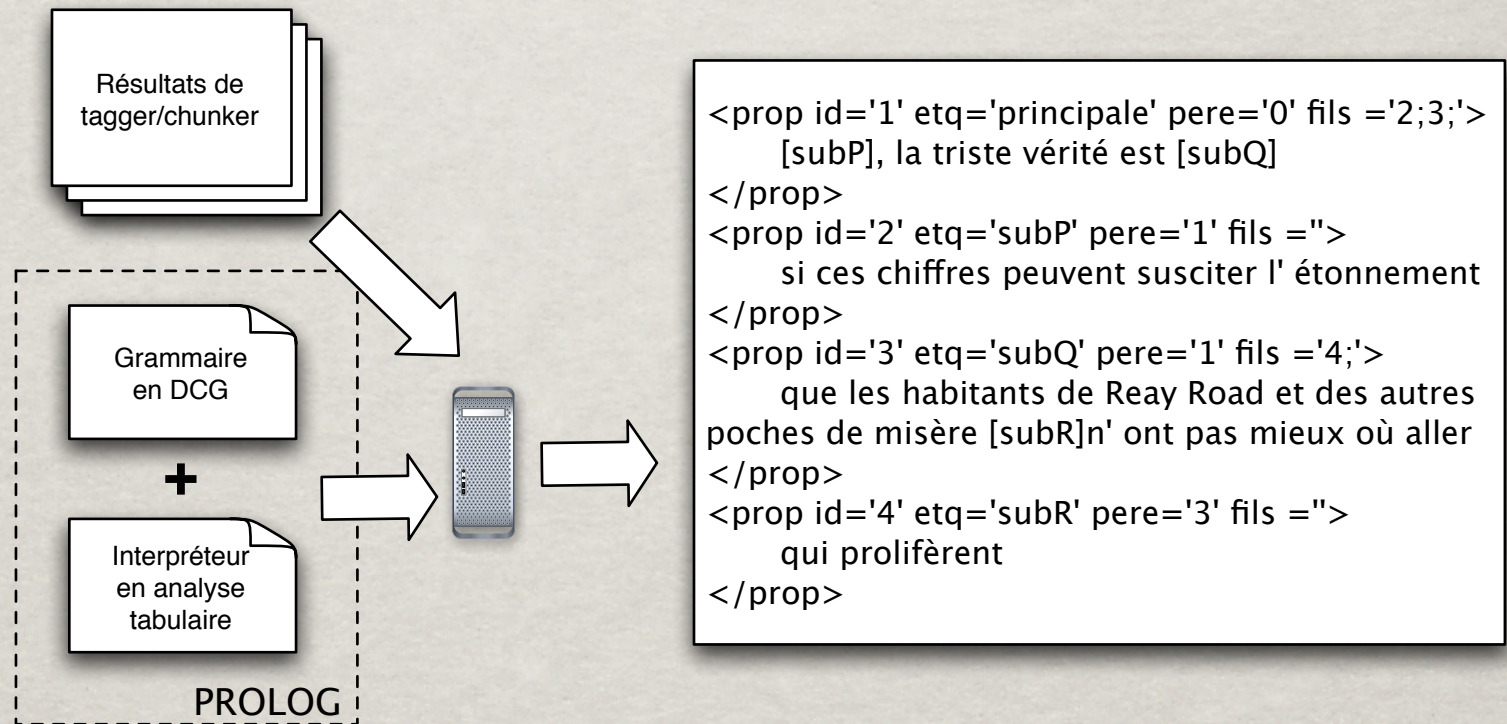
évaluation

Pistes

d'amélioration :
affinement des
étiquettes

Conclusion

RÉALISATION, ÉVALUATION



(Pereira & Shieber 1987)

ÉVALUATION

	Nb de phr	Rappel	Préc.1	Préc. 2	Préc. Totale
G8	53	96,2	98,0	100,0	98,0
Unicode	274	81,4	96,2	97,8	94,1
Zadig	1206	88,6	92,8	95,3	88,4
LMD*	1713	84,9	89,2	98,0	87,4

*Le Monde Diplomatique

A : Nombre total de phrases

B : Nombre de phrases dont l'analyse a abouti

C : Nombre de phrases dont les frontières de propositions sont correctement détectées

D : Nombre de phrases dont les relations des propositions sont correctement analysées

$$\text{Rappel (\%)} = \frac{B}{A} \times 100$$

$$\text{Précision 1 (\%)} = \frac{C}{B} \times 100$$

$$\text{Précision 2 (\%)} = \frac{D}{C} \times 100$$

$$\text{Précision T (\%)} = \frac{\text{Préc. 1} \times \text{Préc. 2}}{100} = \frac{D}{B} \times 100$$

RÉSULTAT 1/3

Si ces chiffres peuvent susciter l'étonnement, la triste vérité est que les habitants de Reay Road et des autres poches de misère qui prolifèrent n'ont pas mieux où aller.

XML

```
<prop id='1' etq='principale' pere='0' fils='2;3;'>
```

[subP], la triste vérité est [subQ]

```
</prop>
```

```
<prop id='2' etq='subP' pere='1' fils=''>
```

si ces chiffres peuvent susciter l'étonnement

```
</prop>
```

```
<prop id='3' etq='subQ' pere='1' fils='4;'>
```

que les habitants de Reay Road et des autres poches de misère [subR] n'ont pas mieux où aller

```
</prop>
```

```
<prop id='4' etq='subR' pere='3' fils=''>
```

qui prolifèrent

```
</prop>
```


RÉSULTAT 2/3

En arrivant aux frontières qui séparent l'Arabie pétérée de la Syrie, comme il passait près d'un château assez fort, des arabes armés en sortirent.

XML

```
<prop id='1' etq='principale' pere='0' fils='2;4;'>
```

```
  [ED] [subP], des arabes armés en sortirent
```

```
</prop>
```

```
<prop id='2' etq='ED' pere='1' fils='3;'>
```

```
  en arrivant aux frontières [subR],
```

```
</prop>
```

```
<prop id='3' etq='subR' pere='2' fils=''>
```

```
  qui séparent l'Arabie pétérée de la Syrie
```

```
</prop>
```

```
<prop id='4' etq='subP' pere='1' fils=''>
```

```
  comme il passait près_d' un château assez fort
```

```
</prop>
```


RÉSULTAT 3/3

Tout ce qui passe sur mes terres est à moi, dit -il, aussi bien que ce que je trouve sur les terres des autres

XML

```

<prop id='1' etq='principale' pere='0' fils='2;3;4;'>
  tout ce [subR] est à moi [Inc] aussi bien que ce [subR]
</prop>
<prop id='2' etq='subR' pere='1' fils=''>
  qui passe sur mes terres
</prop>
<prop id='3' etq='Inc' pere='1' fils=''>
  , dit -il,
</prop>
<prop id='4' etq='subR' pere='1' fils=''>
  que je trouve sur les terres des autres
</prop>
  
```


ÉVALUATION : ERREURS

ANALYSE LINÉAIRE

1. étiquettes erronées attribuées par le pré-traitement ;
2. mauvaise interprétation de connecteurs :
 introducteur de proposition
 ⇔ intr. de syntagme (structure réduite sans verbe)



*Et dire || qu'au moment de son apogée, dans les années 1950,
 Cockerill employait encore plus de 25 000 personnes, || que la ville de
 Seraing était toujours noire de fumée, de bruit, de monde, de travail.*



*Et [dire || qu'au moment de son apogée, dans les années 1950,
 Cockerill employait encore plus de 25 000 personnes, que la ville de
 Seraing] || [était toujours noire de fumée, de bruit, de monde, de travail.]*

ÉVALUATION : ERREURS

ANALYSE LINÉAIRE

1. étiquettes erronées attribuées par le pré-traitement ;
2. mauvaise interprétation de connecteurs :
3. ambiguïté de la position post-prépositionnelle
intégrative pronominale (subSN)
⇔ subordonnée déterminante (subR)

[...] le pouvoir est seulement entre les mains (1)
| de | qui détient des armes à feu (2)
|, de | qui possède les richesses (3).

Diagram illustrating the ambiguity of the post-prepositional position:

- A blue arrow points from the phrase "qui détient des armes à feu (2)" to a blue smiley face icon labeled "subSN".
- A red arrow points from the phrase "qui possède les richesses (3)." to a red frowny face icon labeled "subR".

À comparer :

il admet aussi un Être supérieur (1)
|, de qui la forme et la matière dépendent (2) ☺ subR

ÉVALUATION : ERREURS

ANALYSE STRUCTURALE 1/4

☼ Trois types d'erreurs de l'analyse structurale

1. Coordination de subordonnées sans connecteur

$$\text{☺ } 1 - [2 + 3] \quad \text{☹ } [1 - 2] + 3$$

De son côté, Taikong Corp. explique (1)

| que la firme n'a pas encore le droit de les vendre en France (2)

| , mais peut les exposer (3)

ÉVALUATION : ERREURS

ANALYSE STRUCTURALE 2/4

2. Ambiguïté de la virgule précédant une subordonnée :
 Subordonnée coordonnée ?
 ⇔ Subordonnée déterminante simple ?

$$\text{😊 } 1 - [2 + 3 + [4 - 5]] \quad \text{😞 } 1 - [2 + [3 - 4 - 5]]$$

Personne ne m'a expliqué (1)

*| qu'il s'agissait de la première étape de l'expansion prétendument
 bienveillante d'une nation nouvelle (2)*

*|, mais que cette expansion signifiait en réalité l'expulsion violente des
 Indiens de la totalité du continent (3)*

|, qu'elle serait jalonnée d'atrocités indicibles (4)

| à l'issue desquelles on parquerait les survivants dans des réserves (5)

ÉVALUATION : ERREURS

ANALYSE STRUCTURALE 3/4

3. Ambiguïté des positions

1. *C'est facile à dire* (1)
 | *quand on n'est pas concerné dans sa chair* (2)




😊 subP

☹ subQ

2. *Le pontife trouva dans son coeur* (1)
 | *que cela valait beaucoup* (2)




😊 subQ

☹ subR

ÉVALUATION : ERREURS

ANALYSE STRUCTURALE 4/4

☼ Relations ambiguës

Paris avait estimé, à l'époque, (1)

|| qu'une référence aux valeurs religieuses n'était pas acceptable (2)

|| car elle soulevait des problèmes politiques et constitutionnels en France. (3)

(Paris avait estimé X) + (car Y)

?

Paris avait estimé (une référence aux valeurs religieuses n'était pas acceptable + car Y)

ÉVALUATION FRÉQUENCE DES SUBORDONNÉES

occurrence(%) : **LMD**, **Zadig**

		Int/Fin	post-V	post-N	Autres SN
Sub.	Intégrative pro.		△ 0 0		△ 0,4 0,8
	Percontative		✓ 2 4		△ 0 0
	Complétive		✓ 12 27		△ 0,2 0
Adj.				✓ 2 0,3	
	Relative			✓ 60 57	
Adv.	Intégrative adv.	✓ 15 11		△ 0 0	

✓ = fréquent ; △ = moins fréquent / rare

ÉVALUATION

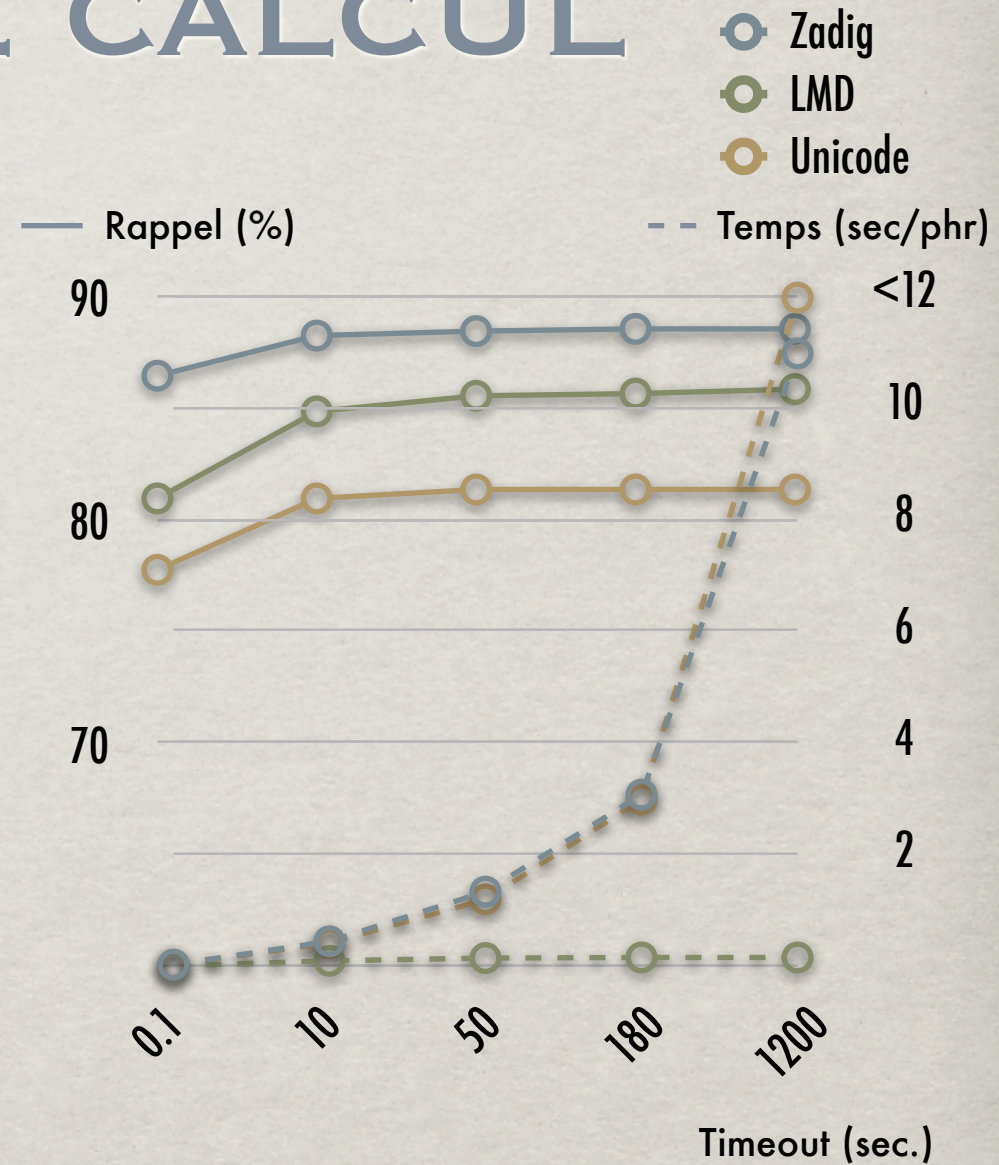
TEMPS DE CALCUL

- Analyse tabulaire descendante

Analyse impossible sans interpréteur

- Fonction de contrôle de temps

- Évaluation avec le résultat à timeout 180 s ;
- Résultat déjà intéressant avec le timeout à 0,1s ;
- Bon équilibre avec le timeout à 10 s.



Contexte :

alignement et
détection des
propositions

Problèmes

des étiquettes
classiques des
connecteurs et des
subordonnées

Typologie

des subordonnées
selon la position

Réalisation,

évaluation

Pistes

d'amélioration :
affinement des
étiquettes

Conclusion

PISTES D'AMÉLIORATION

■ Introduction de plus d'informations

➡ risque de multiplication des calculs

■ Affinement des étiquettes

attribution d'étiquettes syntactico-sémantiques

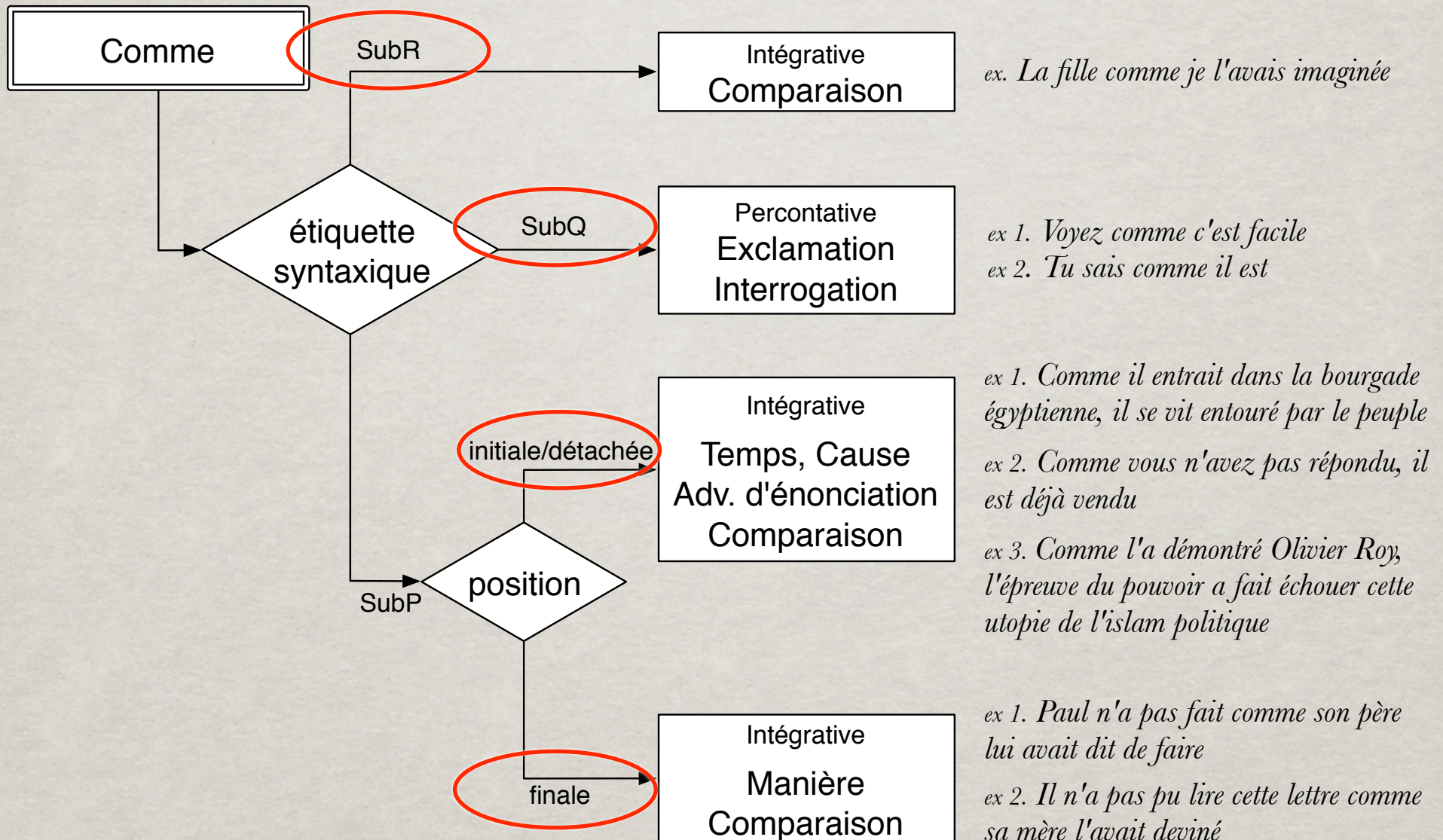
PISTES D'AMÉLIORATION

AFFINEMENT DES ÉTIQUETTES

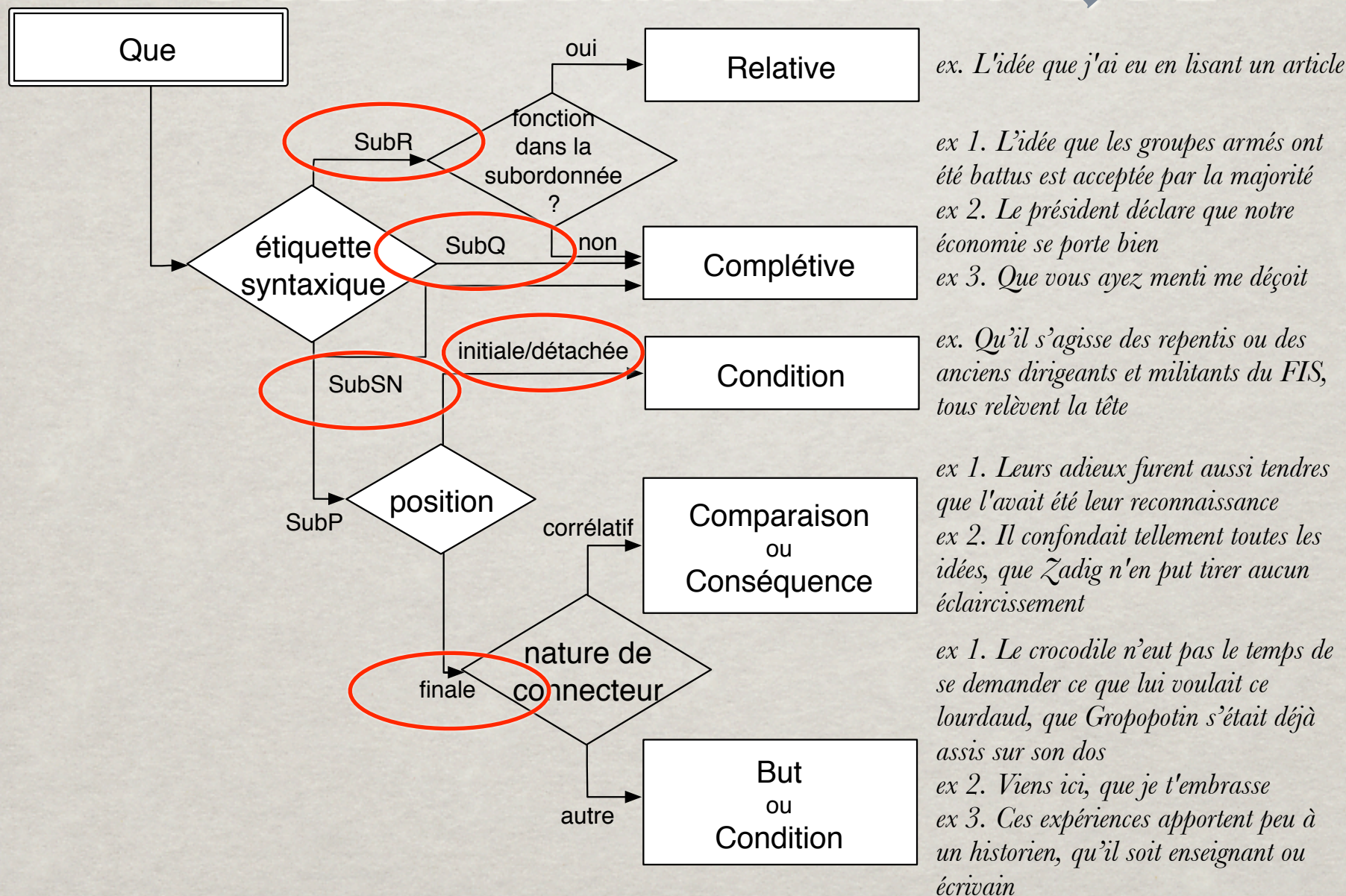
Position	post-V				post-N				Int /Fin				pos. SN			
	I	C	P	R	I	C	P	R	I	C	P	R	I	C	P	R
qui	△		✓					✓					△		△	
que		✓				✓		✓	✓					△		
dont								✓								
où			✓		?			✓	△						△	
quand			✓		△				✓						△	
comme			✓		△				✓						△	
si			✓				△		✓						△	
quoi			✓					✓							△	
lequel			✓					✓							△	
quel			✓												△	
combien			✓												△	
comment			✓												△	
pourquoi			✓												△	

I = Intégrative, C = Complétive, P = Percontative, R = Relative

AFFINEMENT DES ÉTIQUETTES : SUBORDONNÉES EN COMME



AFFINEMENT DES ÉTIQUETTES : SUBORDONNÉES EN QUE



Contexte :

alignement et
détection des
propositions

Problèmes

des étiquettes
classiques des
connecteurs et des
subordonnées

Typologie

des subordonnées
selon la position

Réalisation,

évaluation

Pistes

d'amélioration :
affinement des
étiquettes

Conclusion

CONCLUSION

- Nécessité d'une remise en question de **certaines habitudes classiques** ;
- Utilité d'un recours à des **travaux linguistiques pointus** sur des sujets connexes, et ce pour la conception non seulement d'un **système complet** mais aussi d'un **petit outil** ;
- Alignement des propositions français-japonais :
 - ✓ fournir des données utiles aux études contrastives ;
 - ✓ contribuer au développement des traitements multilingues entre les langues à structure très différentes.

MERCI

YAYOI NAKAMURA-DELLOYE

Paris 7 - LATTICE (CNRS UMR8094)

yayoi@free.fr