

Ne nous arrêtons pas en si bon chemin :  
Améliorations de l'apprentissage global d'analyseurs en  
dépendances par transition

**Lauriane Aufrant, Guillaume Wisniewski, François Yvon**

23ème conférence sur le Traitement Automatique des Langues Naturelles

6 juillet 2016



- Étude des méthodes d'apprentissage, dans le cadre de l'analyse syntaxique en dépendances
- L'un des principes fondamentaux de l'apprentissage statistique n'est pas garanti : la similarité des exemples d'apprentissage et de test.
- Nous corrigeons ces biais en proposant trois améliorations simples de l'apprentissage.

## 1 État de l'art

- Analyse en dépendances par transition
- Oracle dynamique glouton
- Recherche en faisceau et apprentissage global
- Limites des méthodes d'apprentissage actuelles

## 2 Améliorations proposées

- Oracle dynamique global
- Reprise d'apprentissage
- Références dans l'espace sous-optimal

## 3 Validation expérimentale

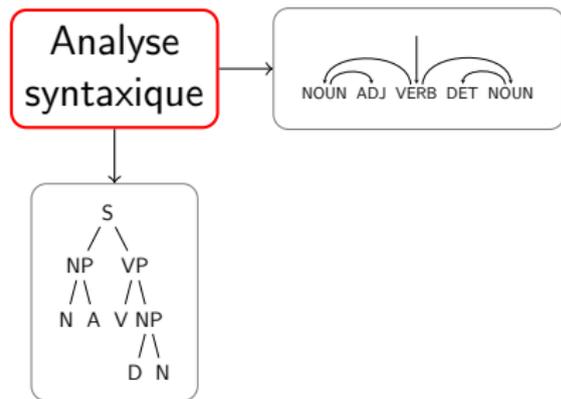
- Cadre expérimental
- Résultats
- Analyse de similarité

## 4 Conclusions

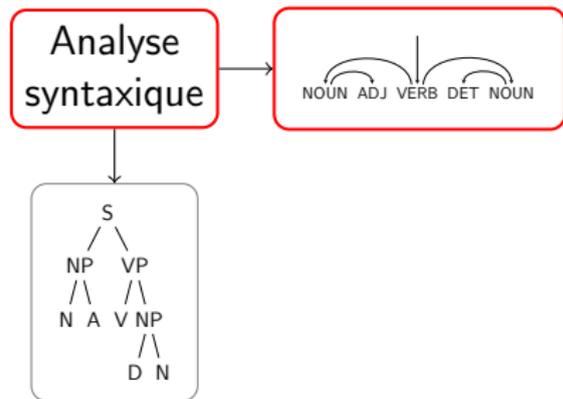
- 1 État de l'art
  - Analyse en dépendances par transition
  - Oracle dynamique glouton
  - Recherche en faisceau et apprentissage global
  - Limites des méthodes d'apprentissage actuelles
- 2 Améliorations proposées
  - Oracle dynamique global
  - Reprise d'apprentissage
  - Références dans l'espace sous-optimal
- 3 Validation expérimentale
  - Cadre expérimental
  - Résultats
  - Analyse de similarité
- 4 Conclusions

Analyse  
syntaxique

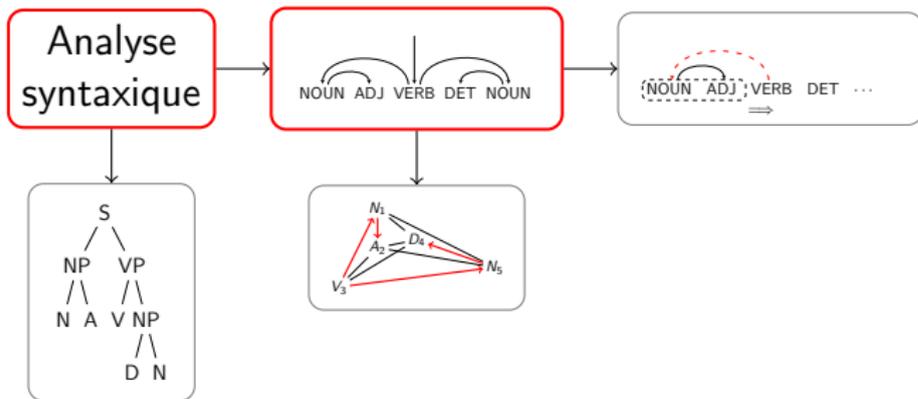
# Analyse en dépendances par transition



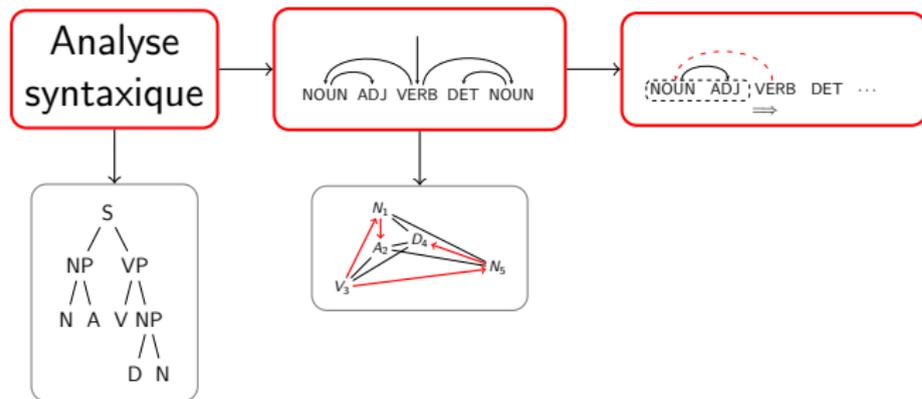
# Analyse en dépendances par transition



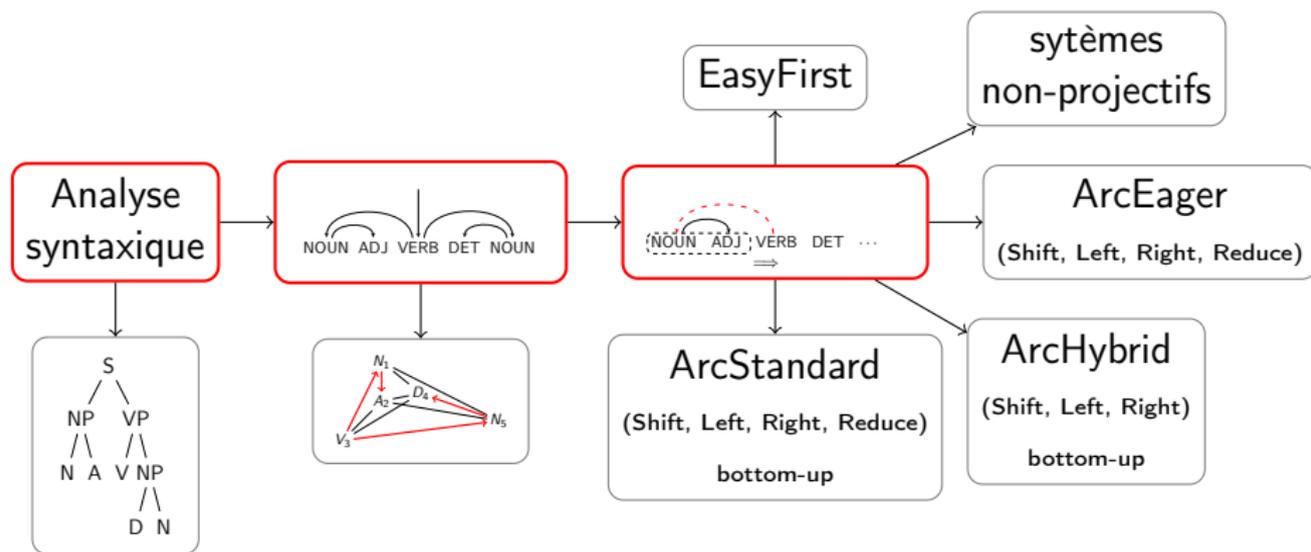
# Analyse en dépendances par transition



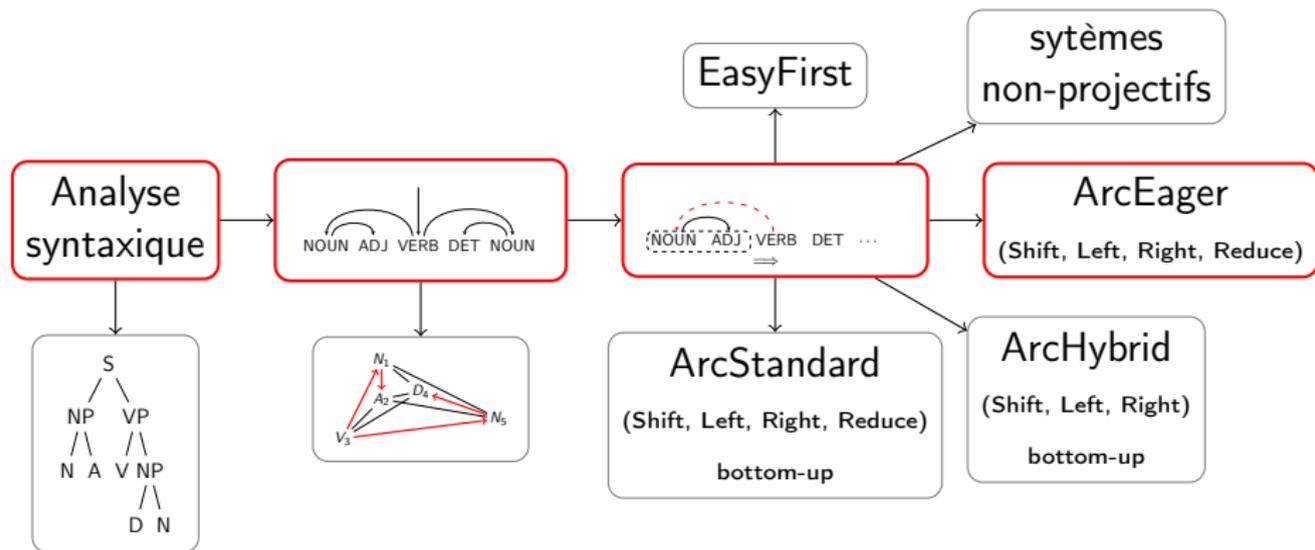
# Analyse en dépendances par transition



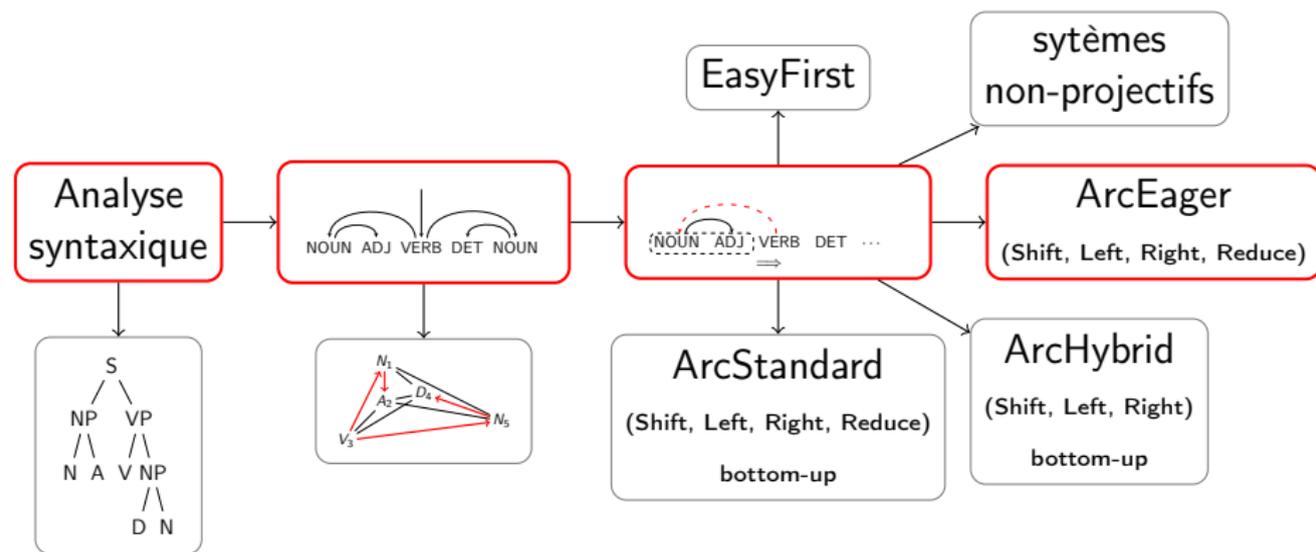
# Analyse en dépendances par transition



# Analyse en dépendances par transition



# Analyse en dépendances par transition

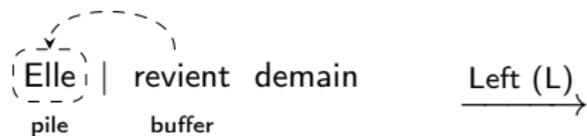


Actions prédites par un classifieur :  
perceptron multi-classes moyenné [Collins & Roark, 2004]

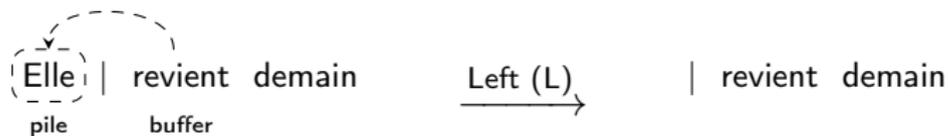
# Système ArcEager

(Elle) | revient demain  
pile      buffer

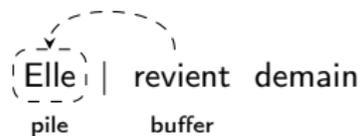
# Système ArcEager



# Système ArcEager



# Système ArcEager



Left (L) →

| revient demain

Shift (S) →

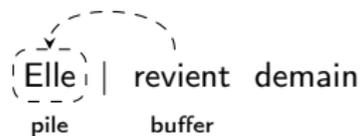
# Système ArcEager

Elle | revient demain  
pile      buffer

Left (L) → | revient demain

Shift (S) → (revient) | demain

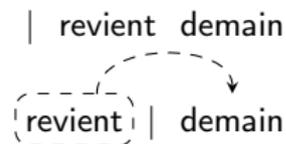
# Système ArcEager



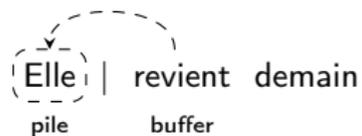
Left (L)  
→

Shift (S)  
→

Right (R)  
→



# Système ArcEager



Left (L)  
→

| revient demain

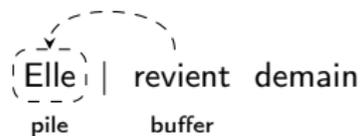
Shift (S)  
→



Right (R)  
→



# Système ArcEager



Left (L)  
→

| revient demain

Shift (S)  
→

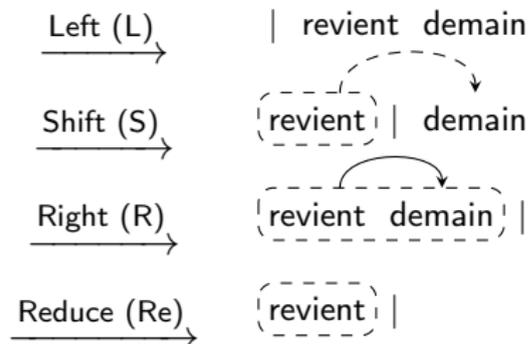
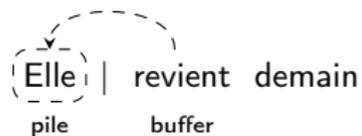


Right (R)  
→

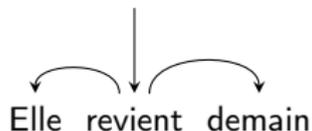
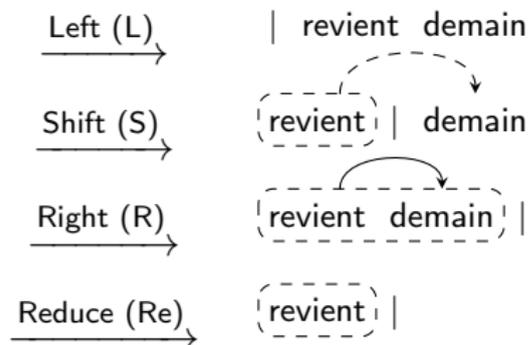
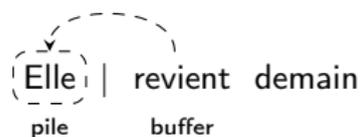


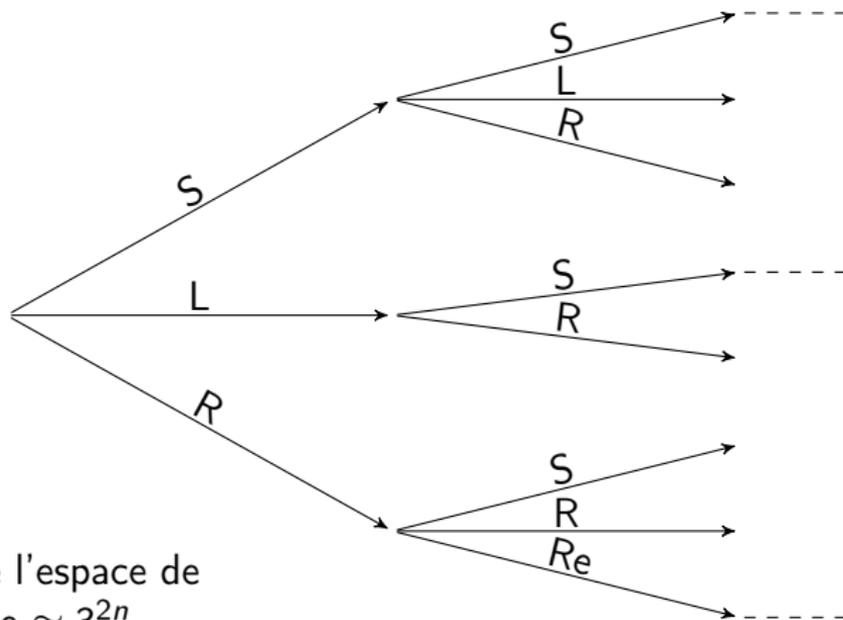
Reduce (Re)  
→

# Système ArcEager



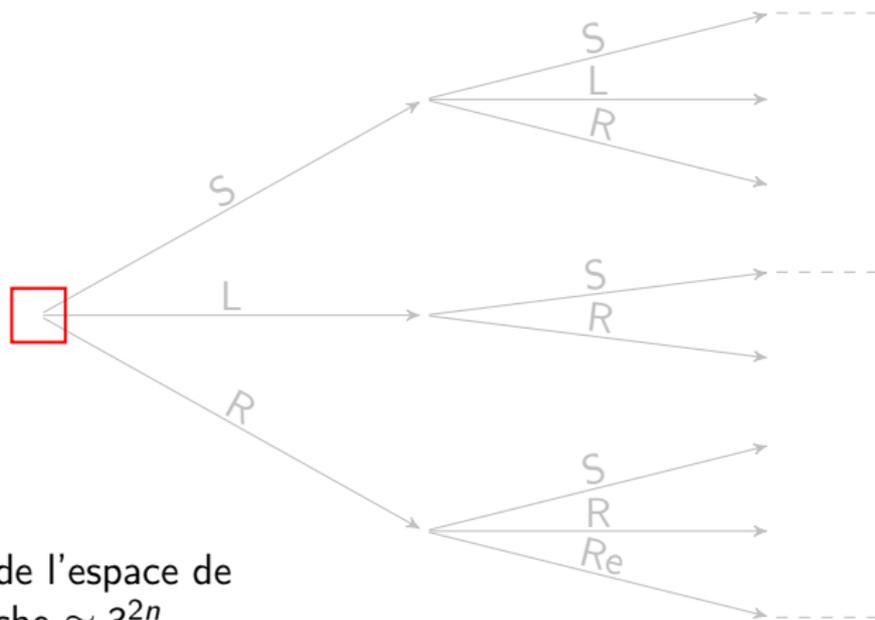
# Système ArcEager





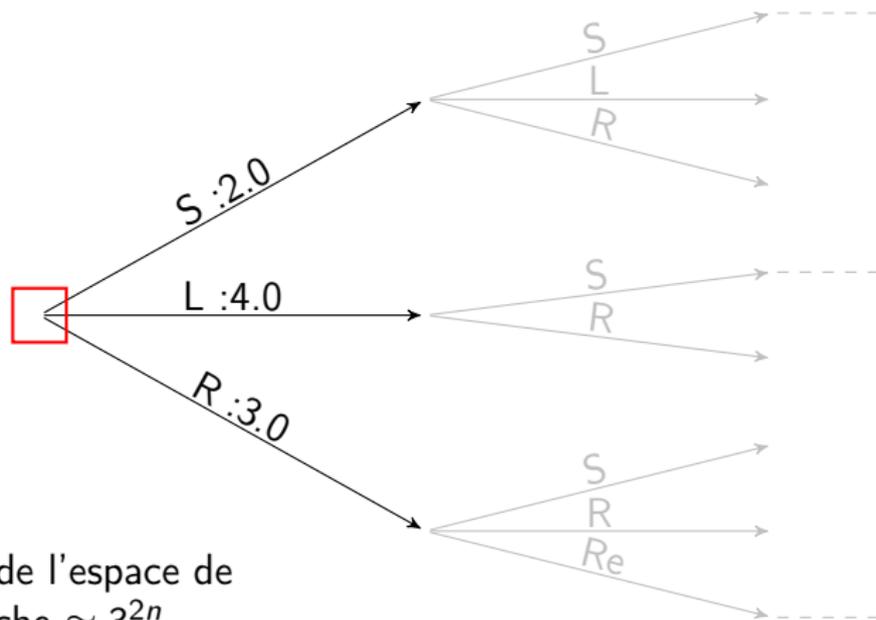
Taille de l'espace de  
recherche  $\approx 3^{2n}$

# Inférence gloutonne

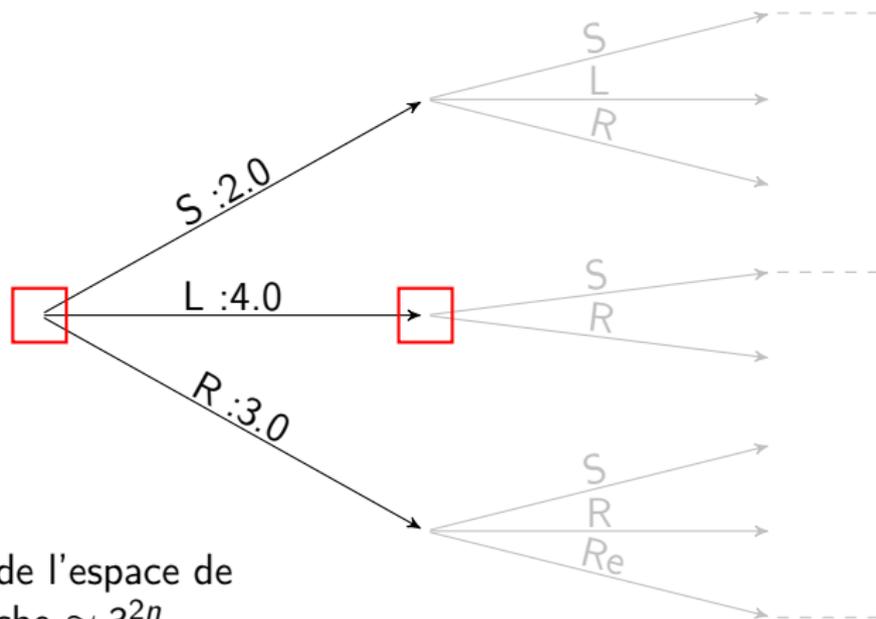


Taille de l'espace de  
recherche  $\approx 3^{2n}$

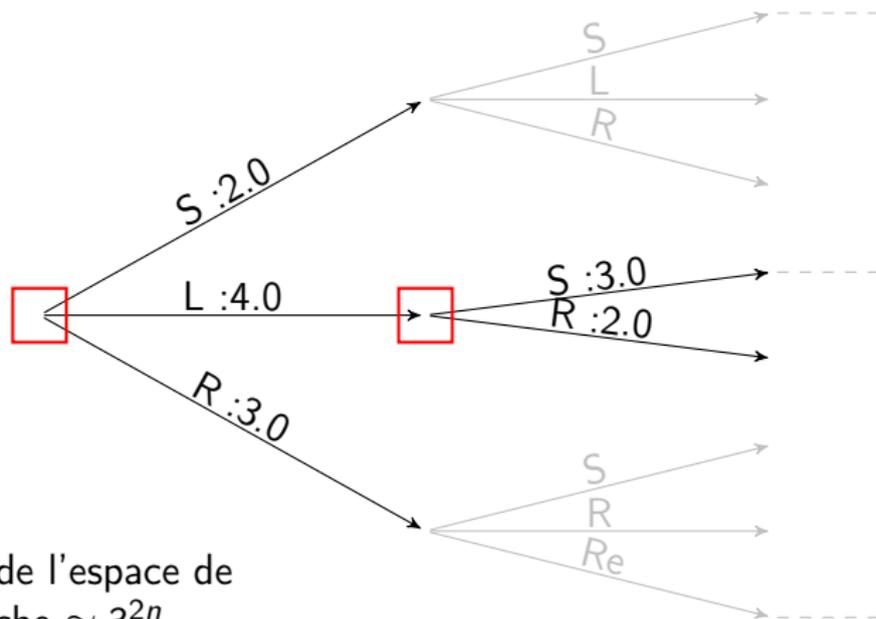
# Inférence gloutonne



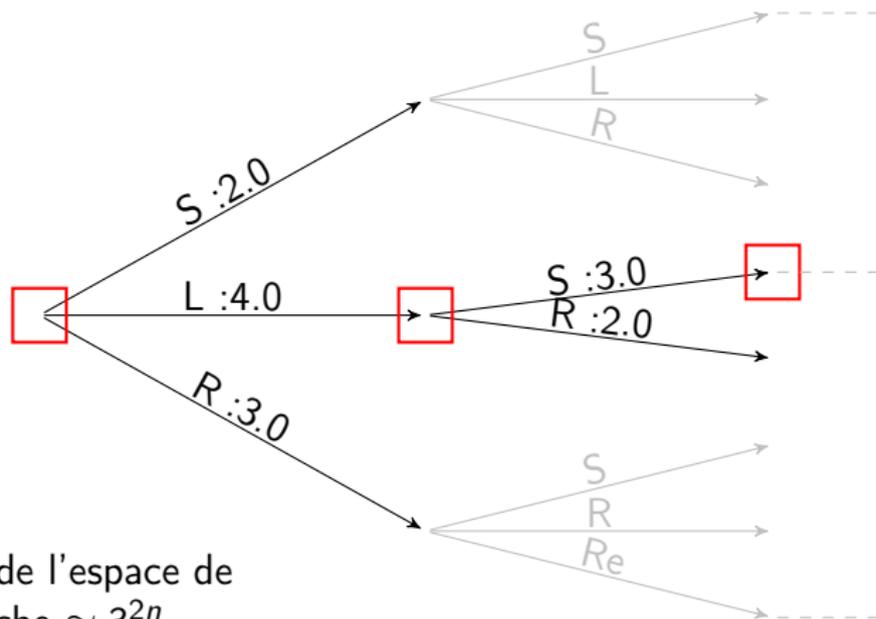
# Inférence gloutonne



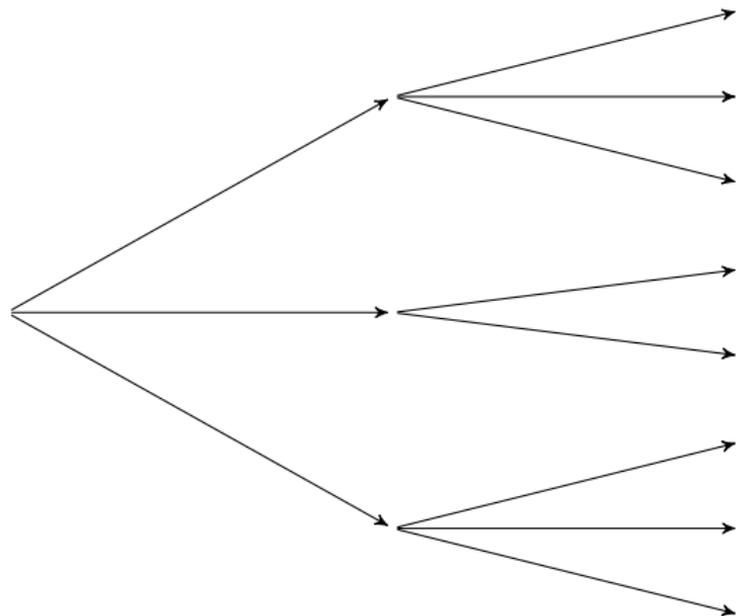
# Inférence gloutonne



# Inférence gloutonne

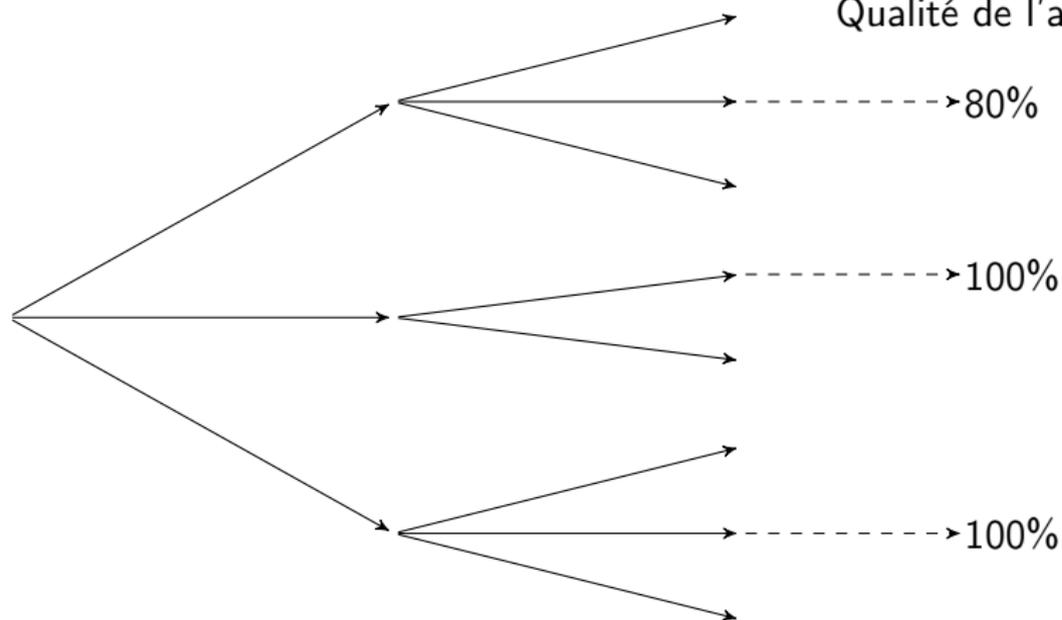


# Apprentissage glouton...

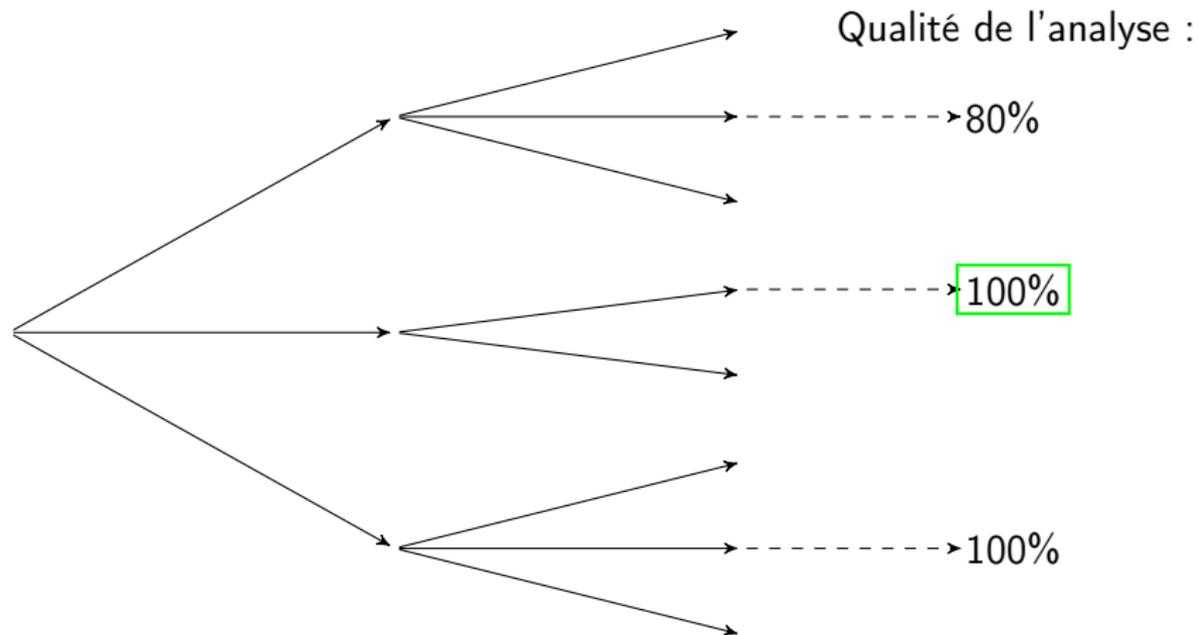


# Apprentissage glouton...

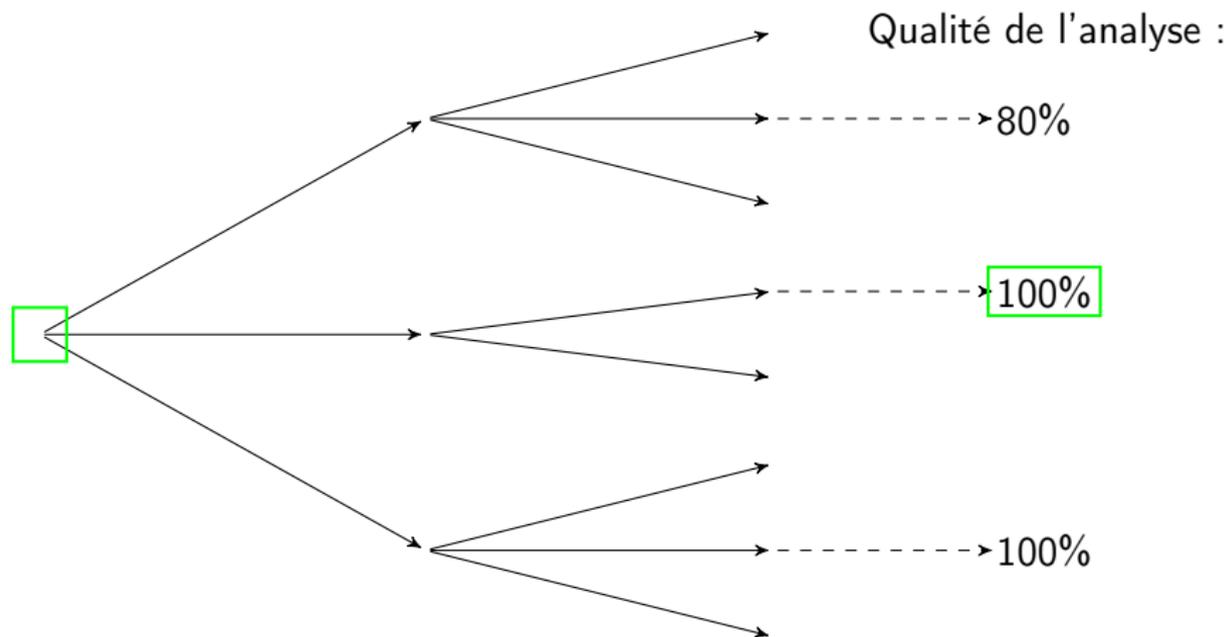
Qualité de l'analyse :



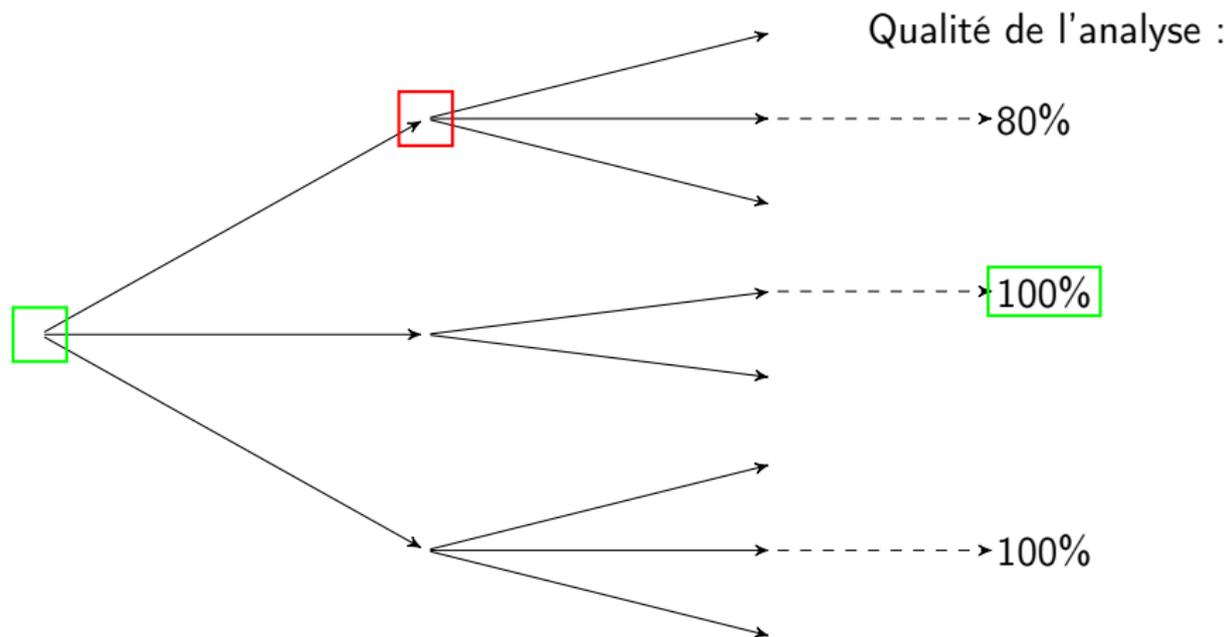
# Apprentissage glouton...



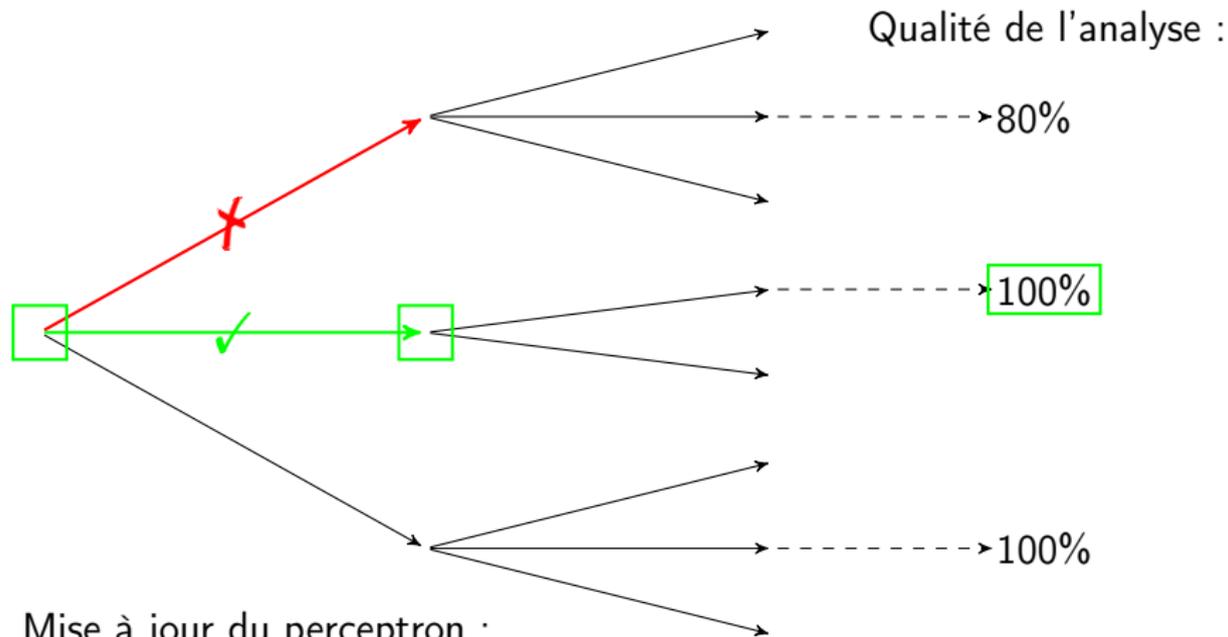
# Apprentissage glouton...



# Apprentissage glouton...



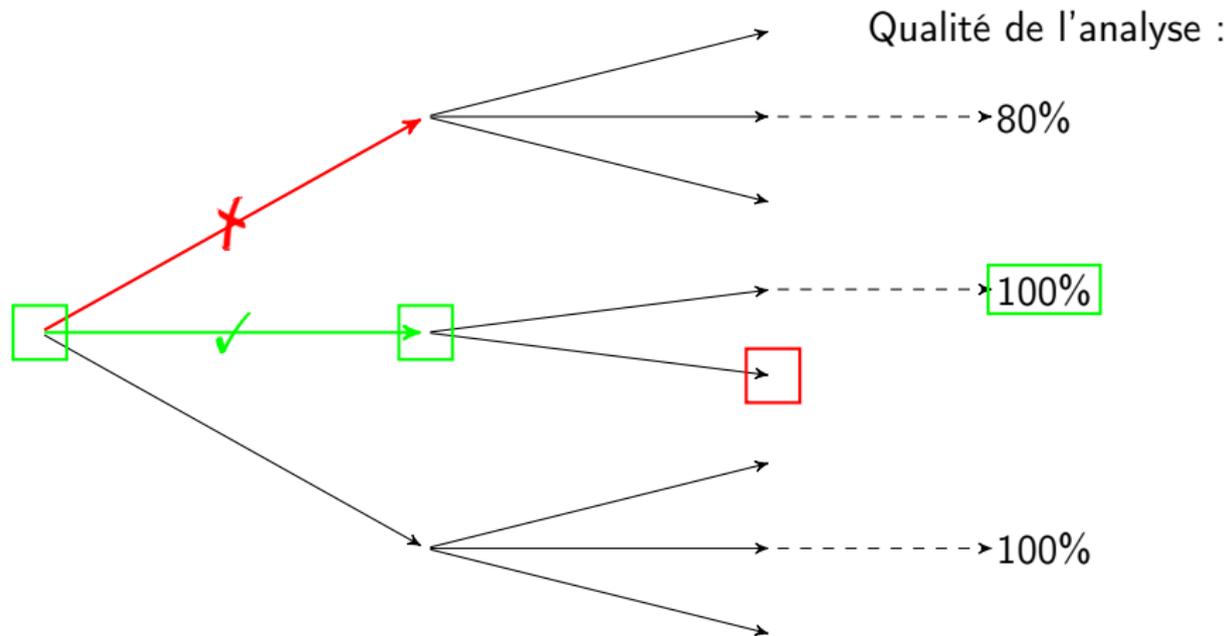
# Apprentissage glouton...



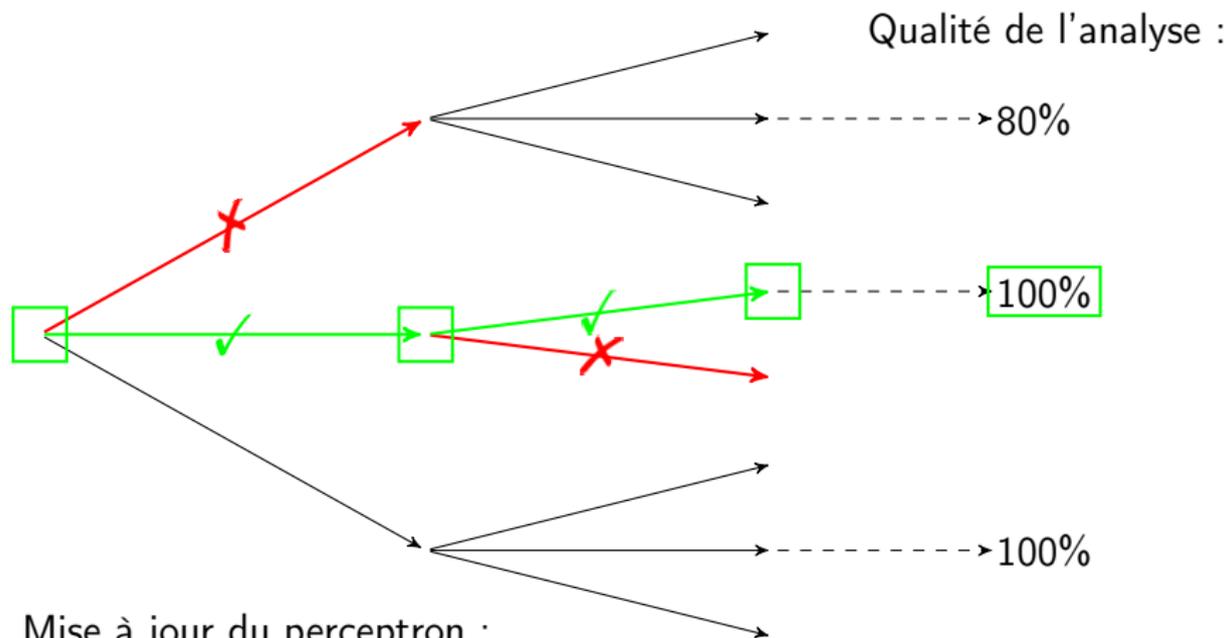
Mise à jour du perceptron :

$$\mathbf{w} \leftarrow \mathbf{w} - \phi + \phi^*$$

# Apprentissage glouton...



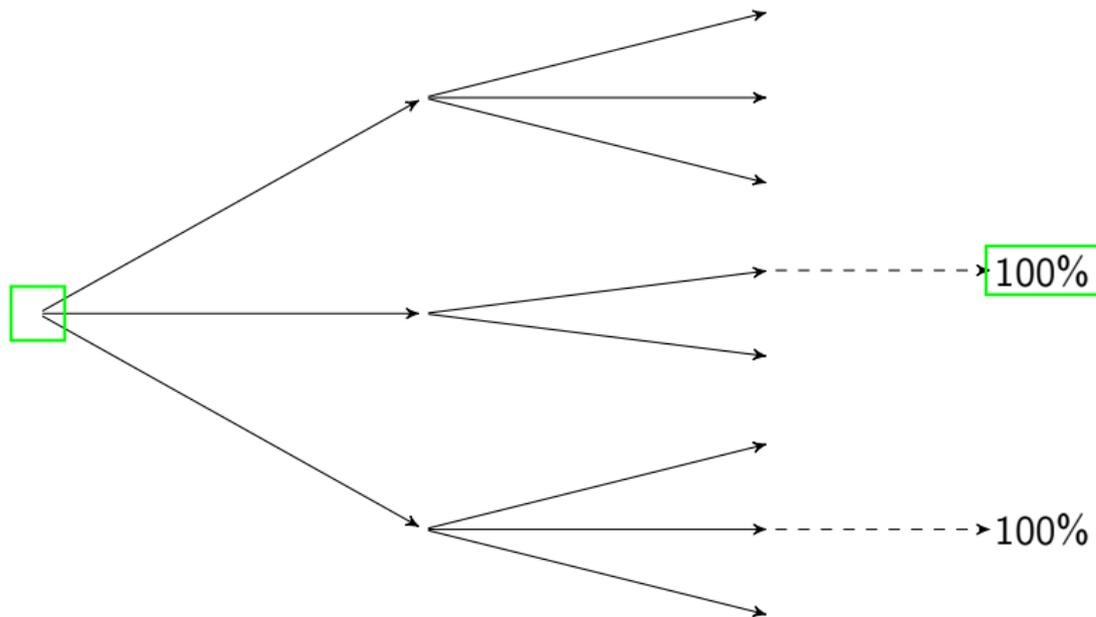
# Apprentissage glouton...



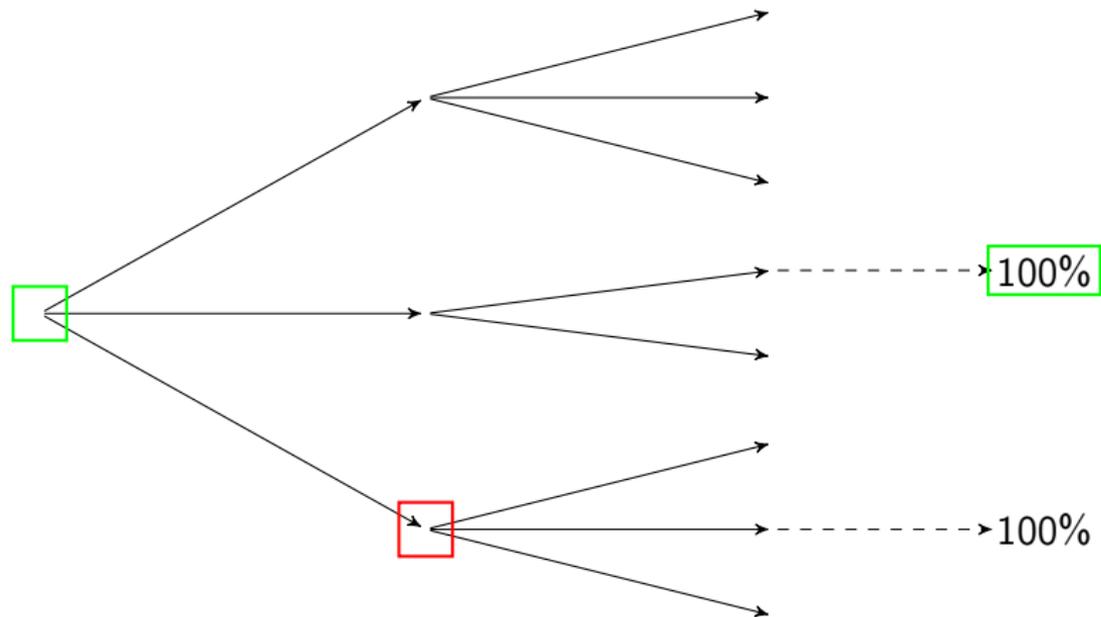
Mise à jour du perceptron :

$$\mathbf{w} \leftarrow \mathbf{w} - \phi + \phi^*$$

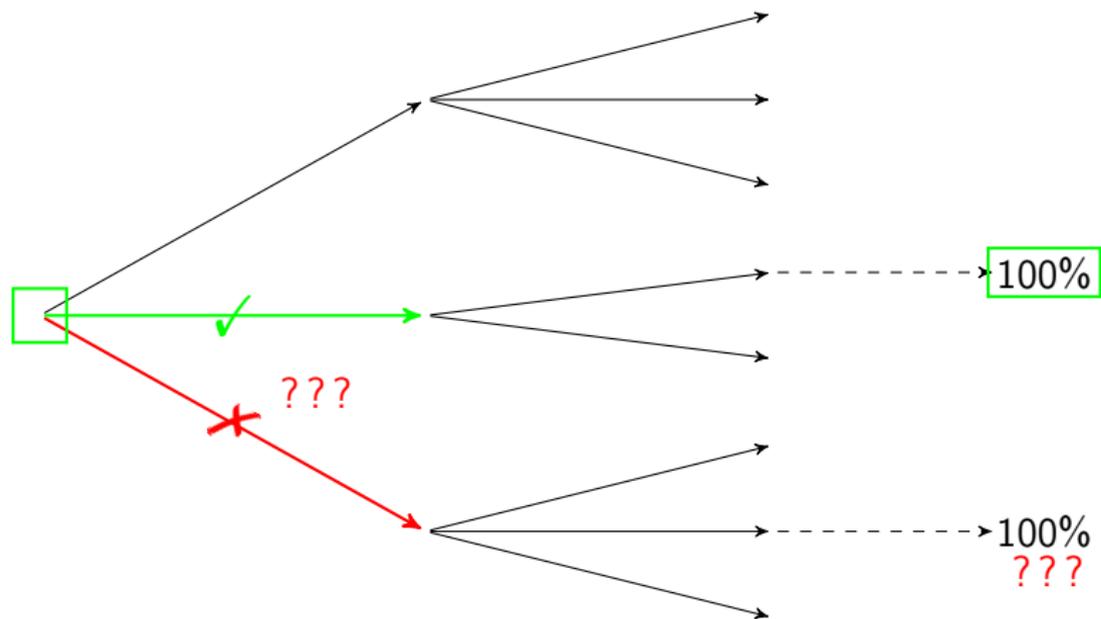
# Apprentissage glouton non-déterministe ?



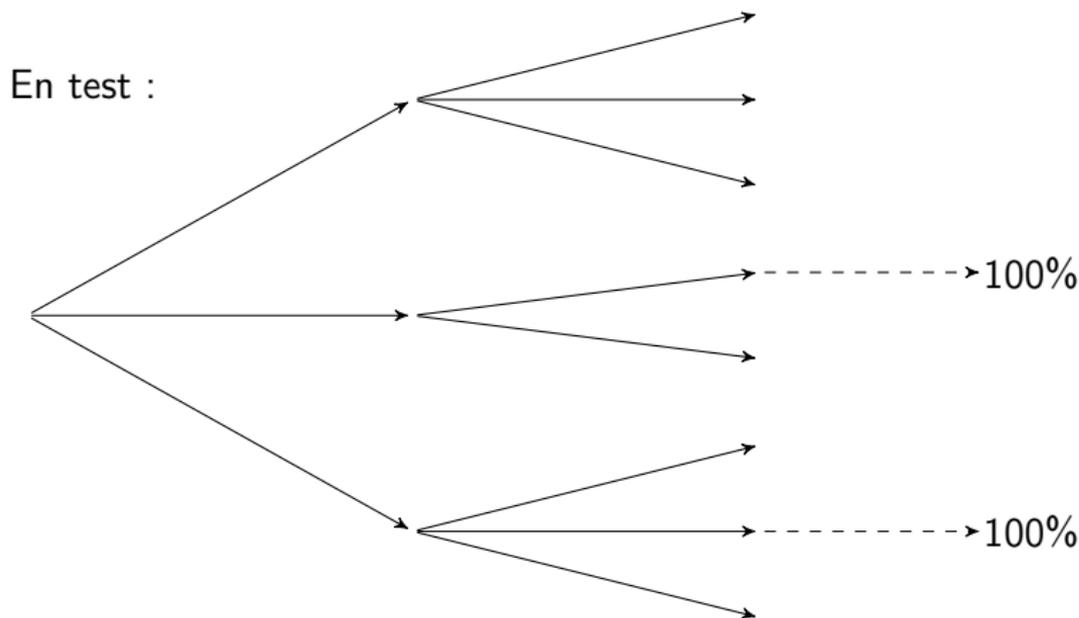
# Apprentissage glouton non-déterministe ?



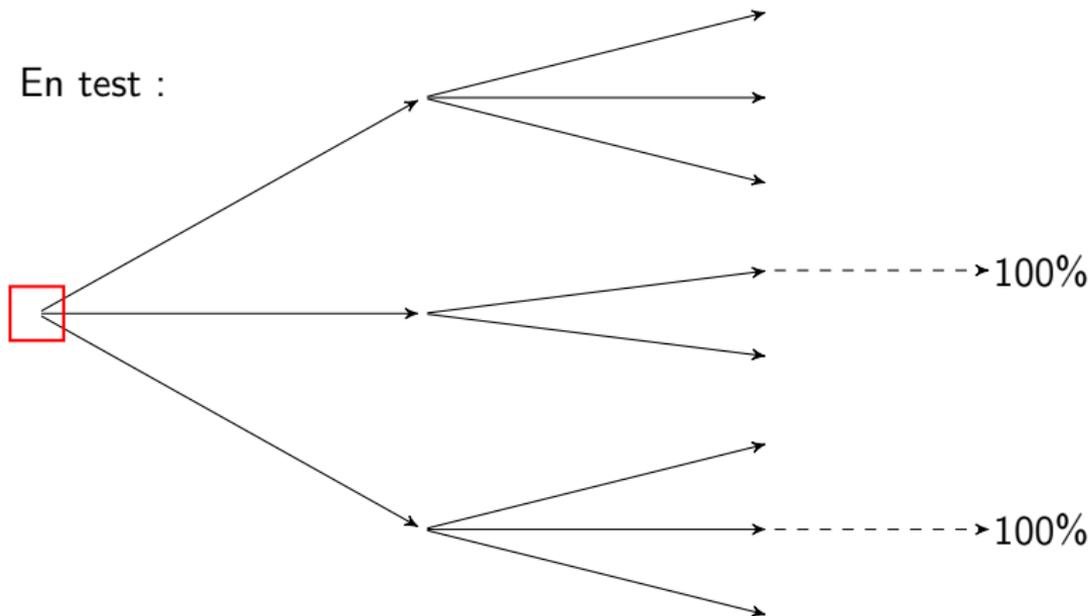
# Apprentissage glouton non-déterministe ?



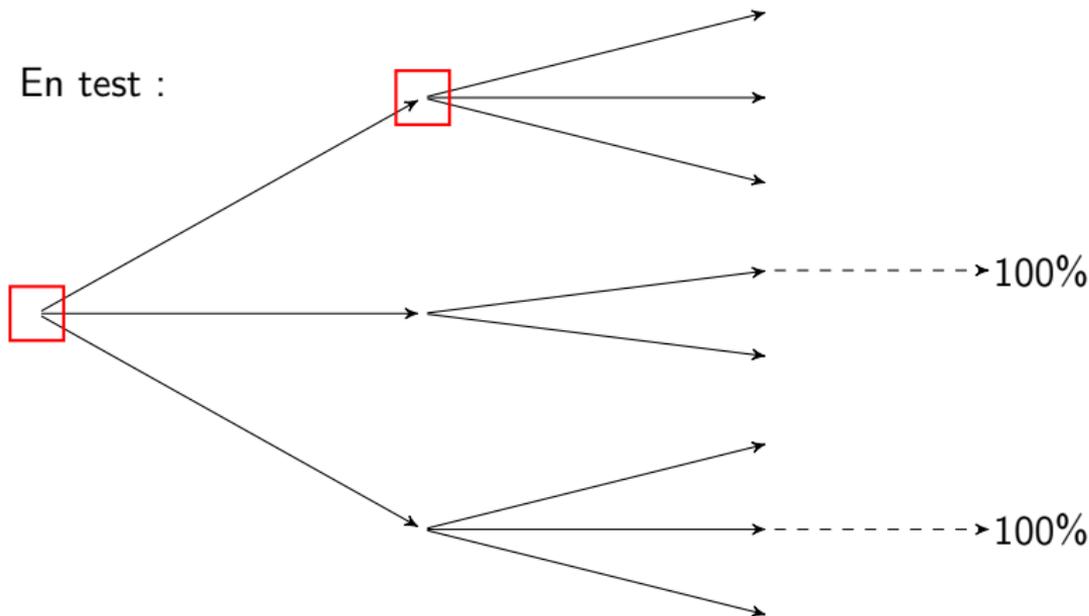
# Apprentissage glouton dans l'espace sous-optimal ?



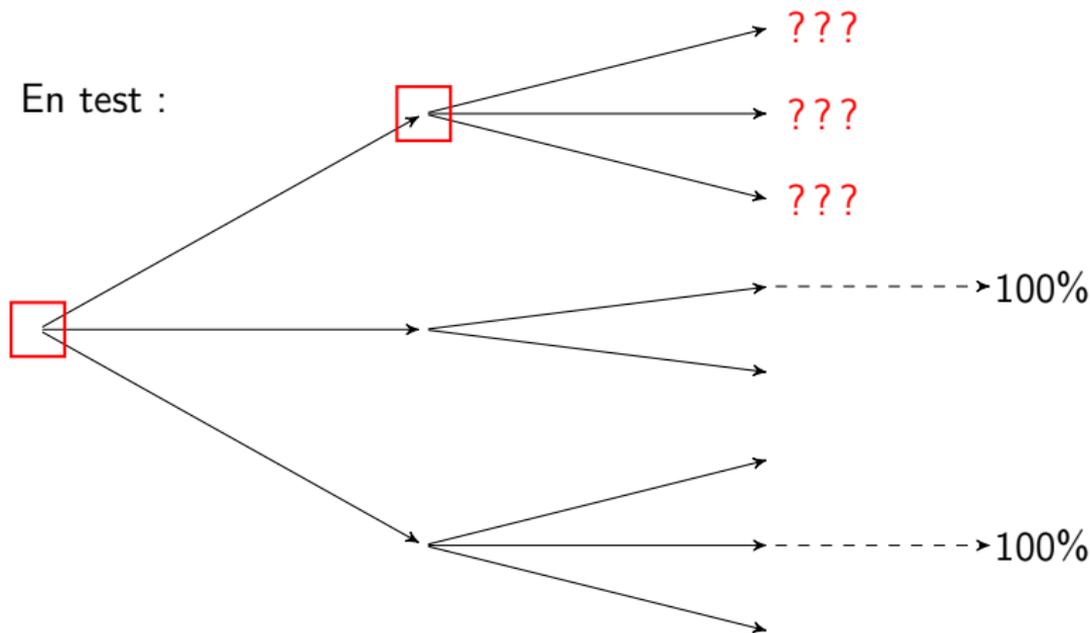
# Apprentissage glouton dans l'espace sous-optimal ?



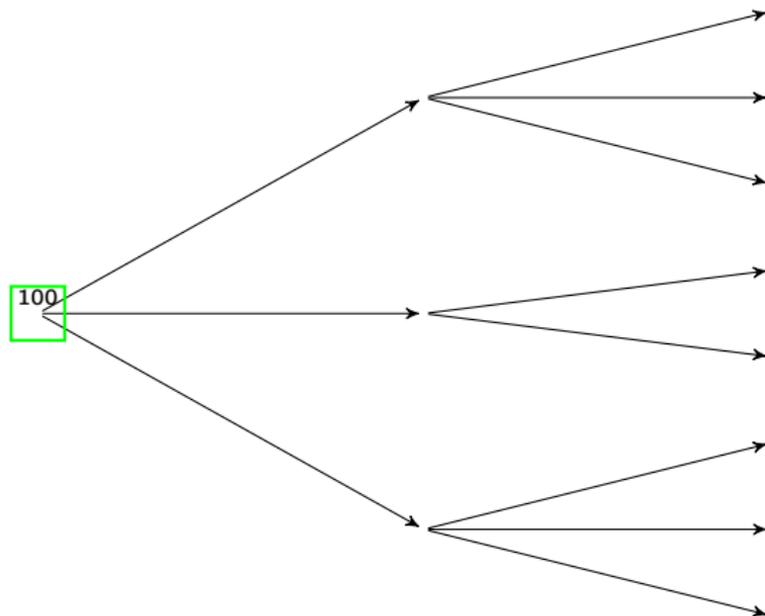
# Apprentissage glouton dans l'espace sous-optimal ?



# Apprentissage glouton dans l'espace sous-optimal ?

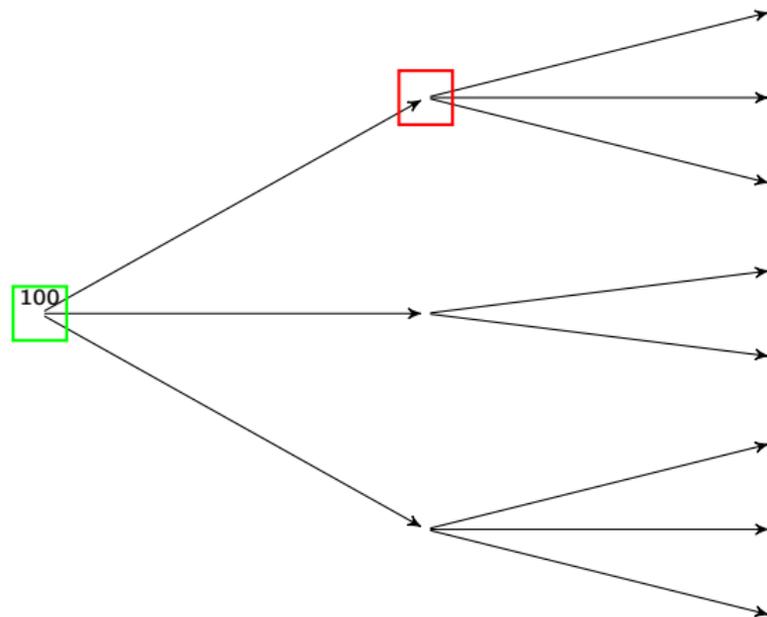


# Apprentissage glouton dynamique



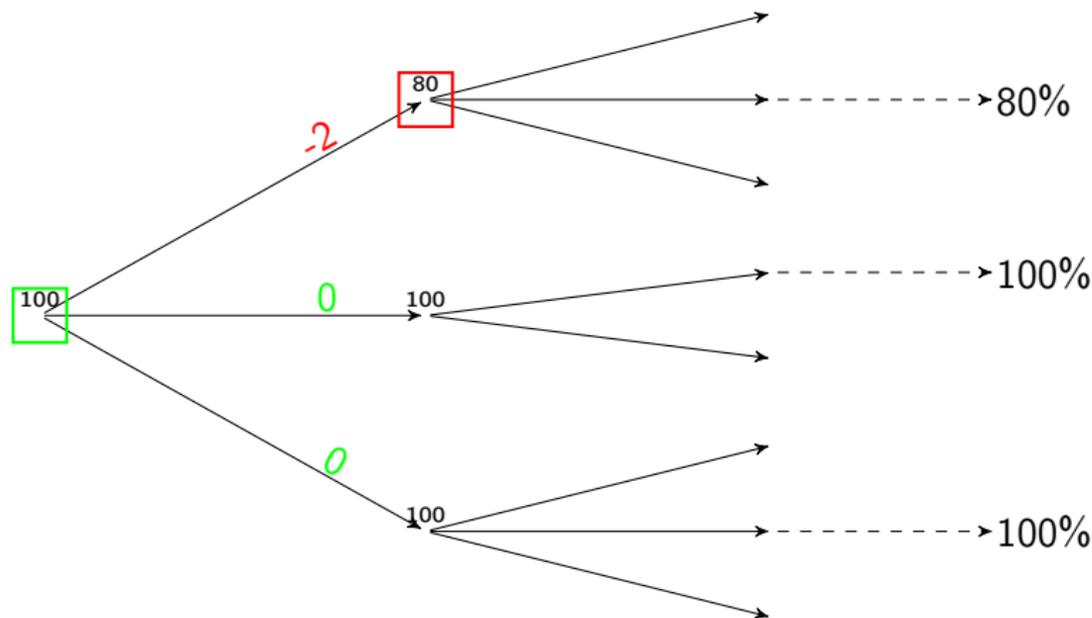
Cost(*action*) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

# Apprentissage glouton dynamique



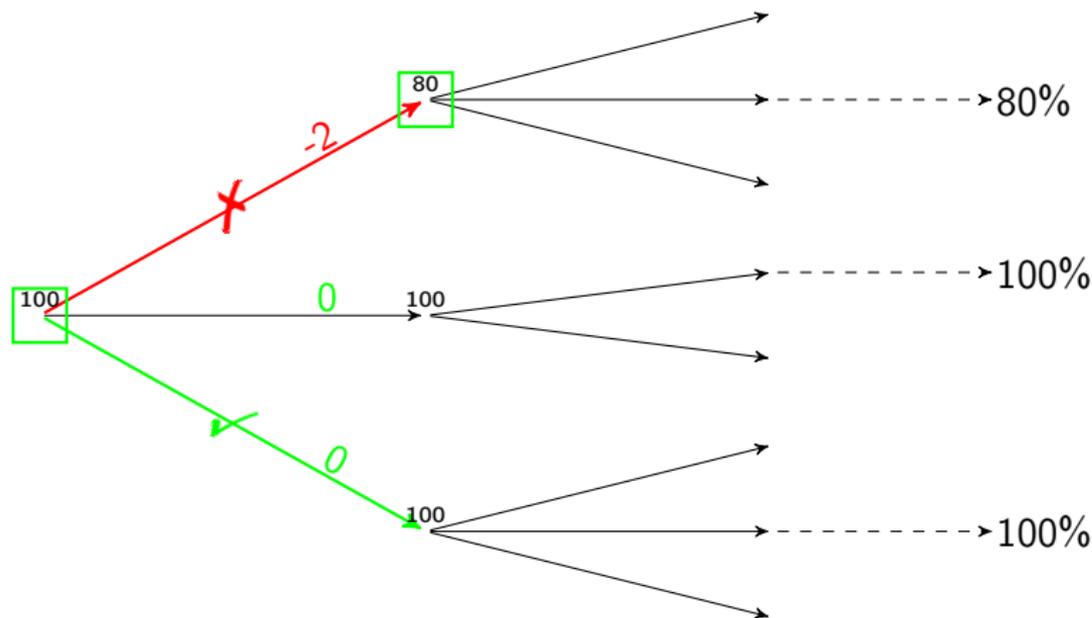
Cost(*action*) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

# Apprentissage glouton dynamique



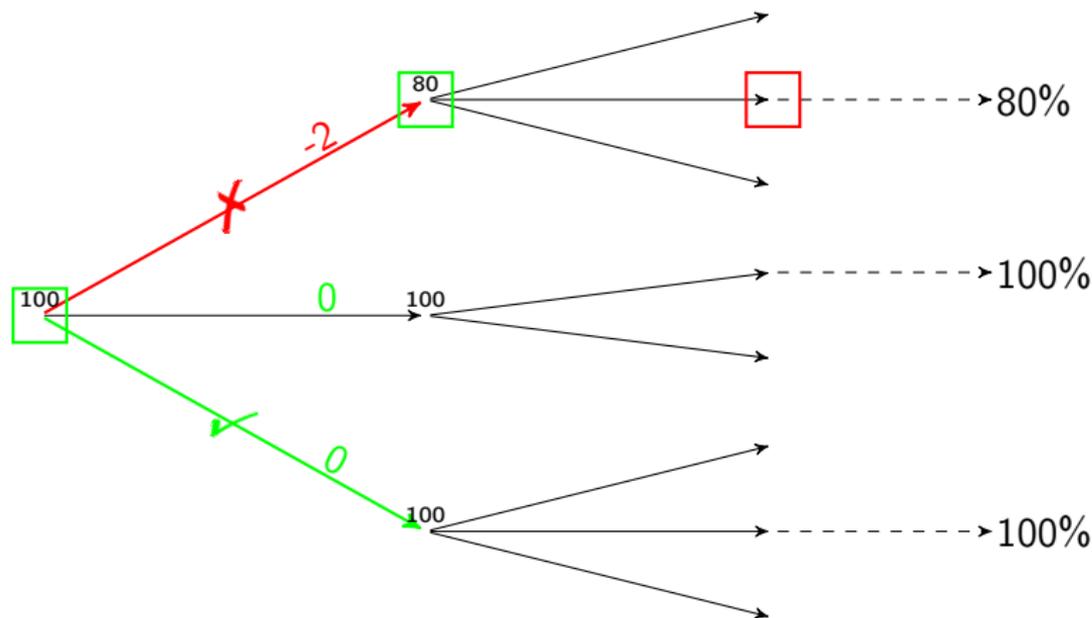
Cost(action) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

# Apprentissage glouton dynamique



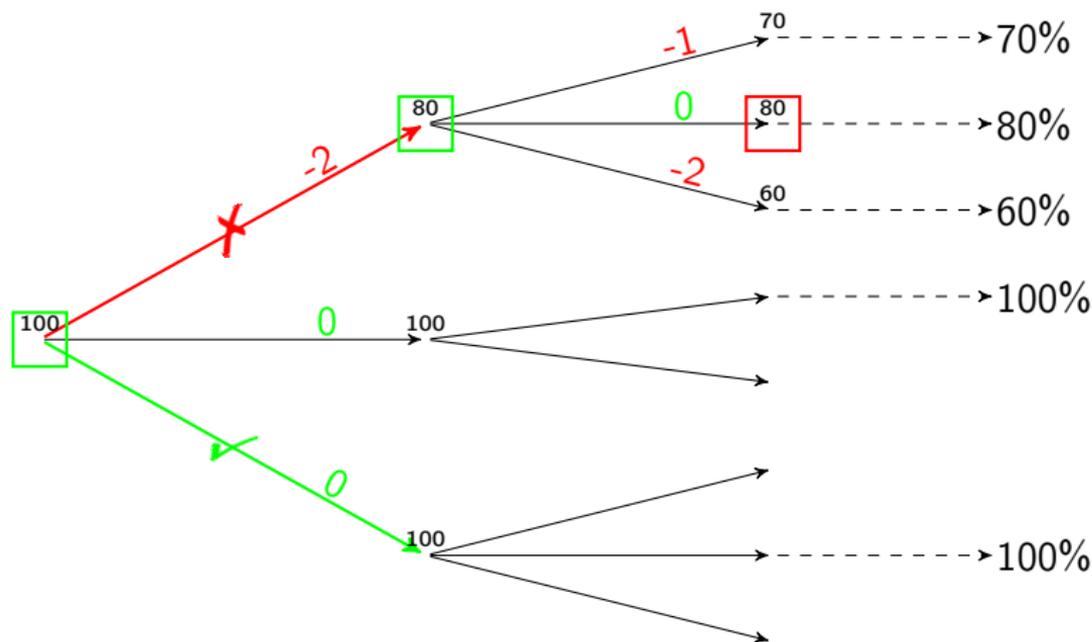
Cost(action) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

# Apprentissage glouton dynamique



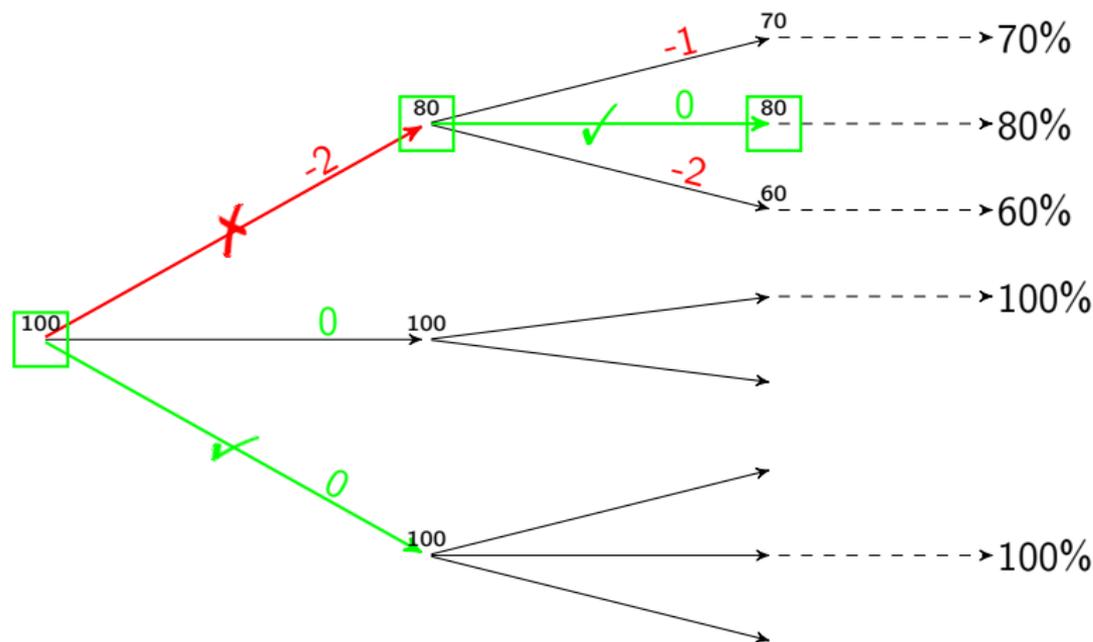
Cost(action) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

# Apprentissage glouton dynamique



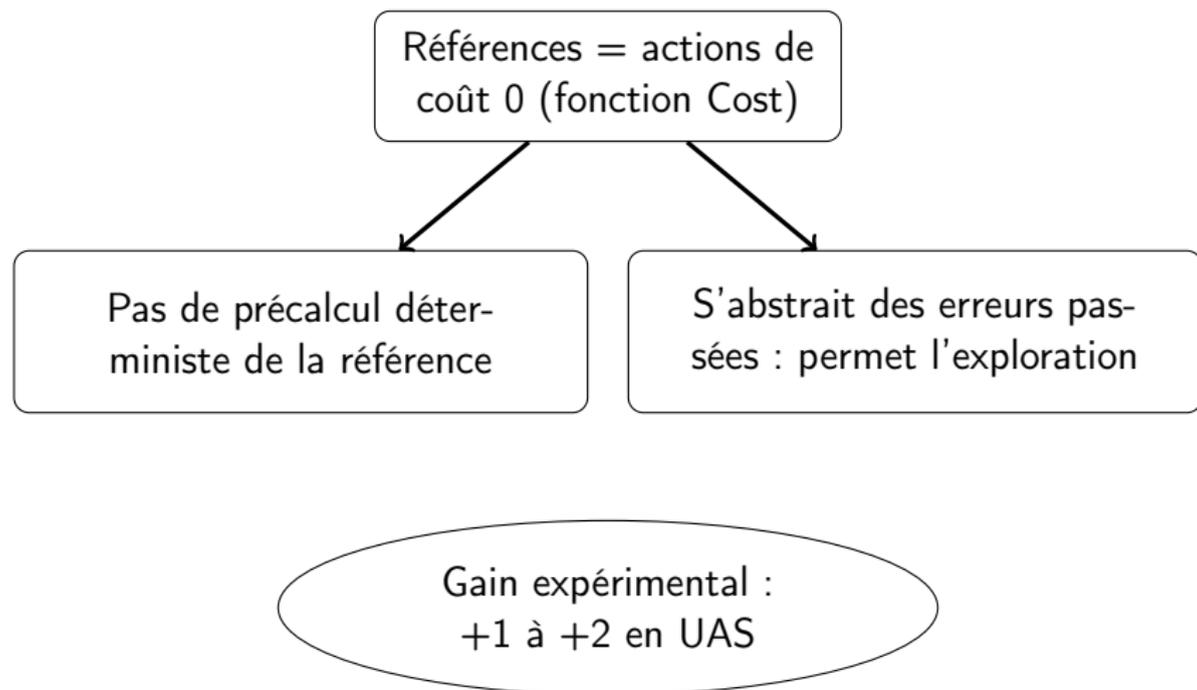
Cost(action) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

# Apprentissage glouton dynamique

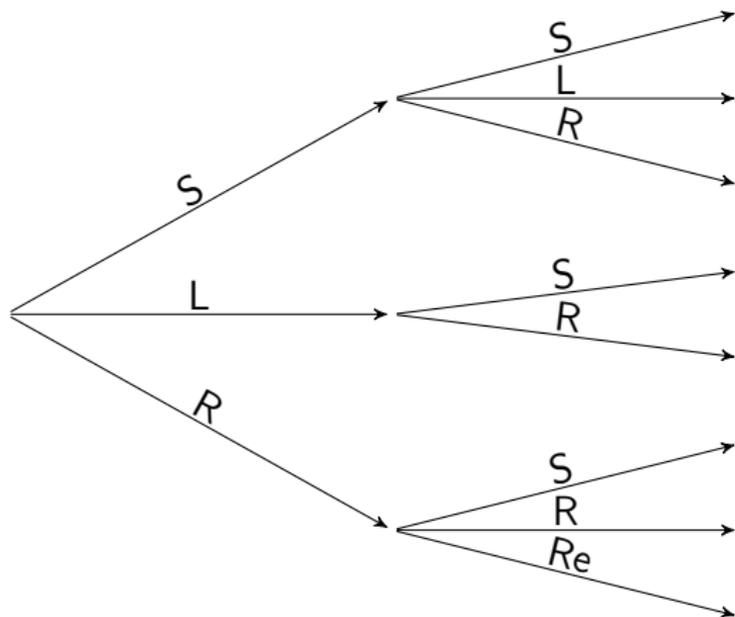


Cost(action) [Goldberg & Nivre, 2012] :  
 $\Delta$  qualité d'analyse de la phrase

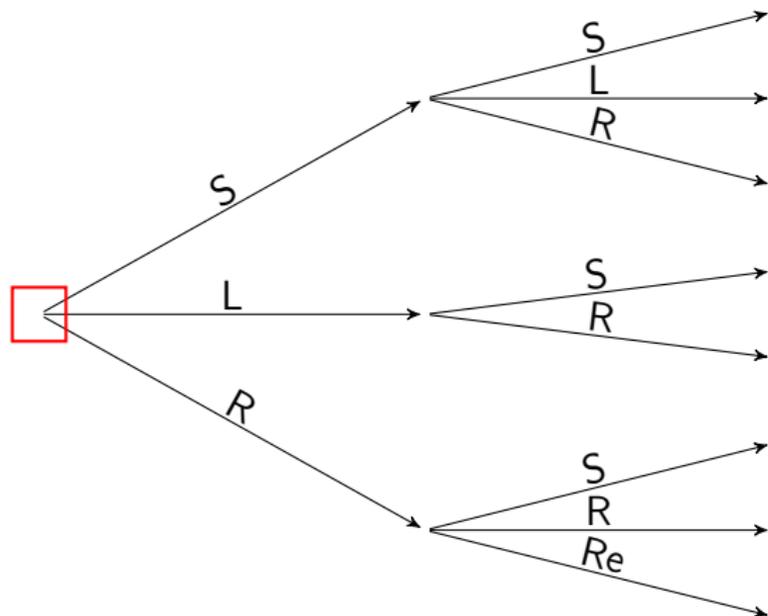
# Oracle dynamique glouton [Goldberg & Nivre, 2012]



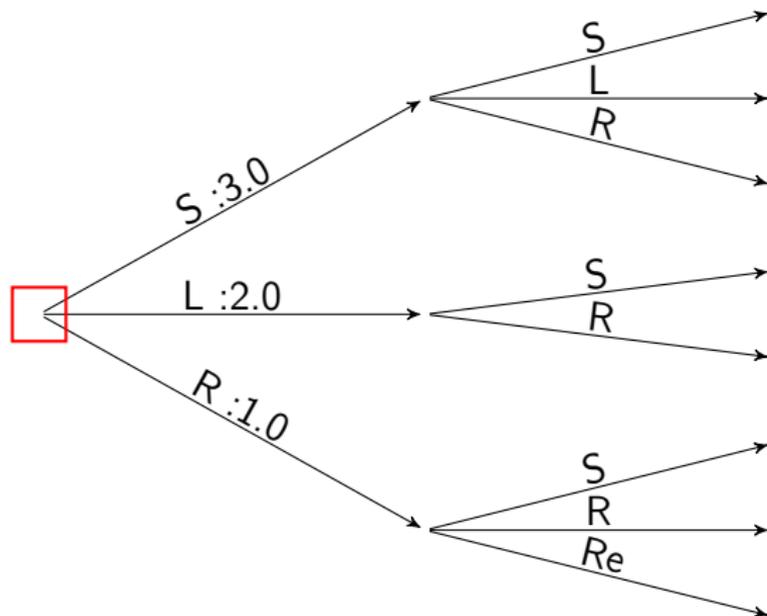
# Recherche en faisceau (*beam*)



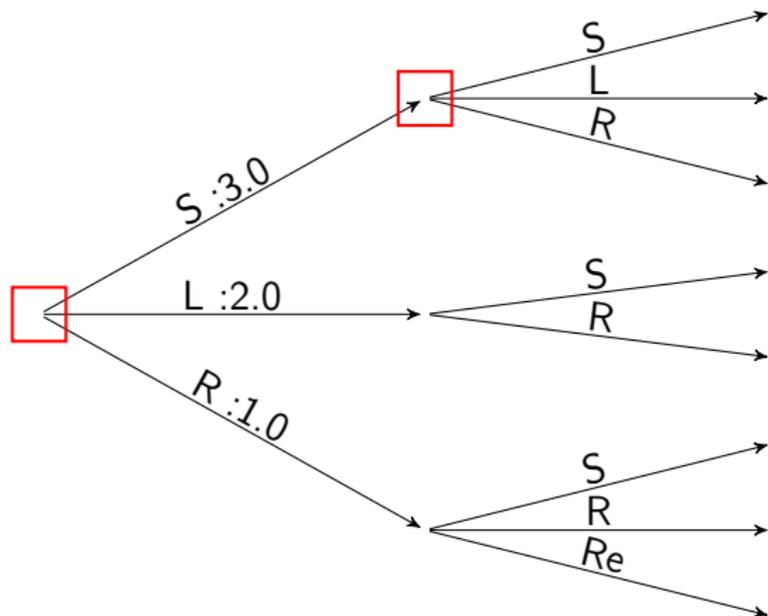
# Recherche en faisceau (*beam*)



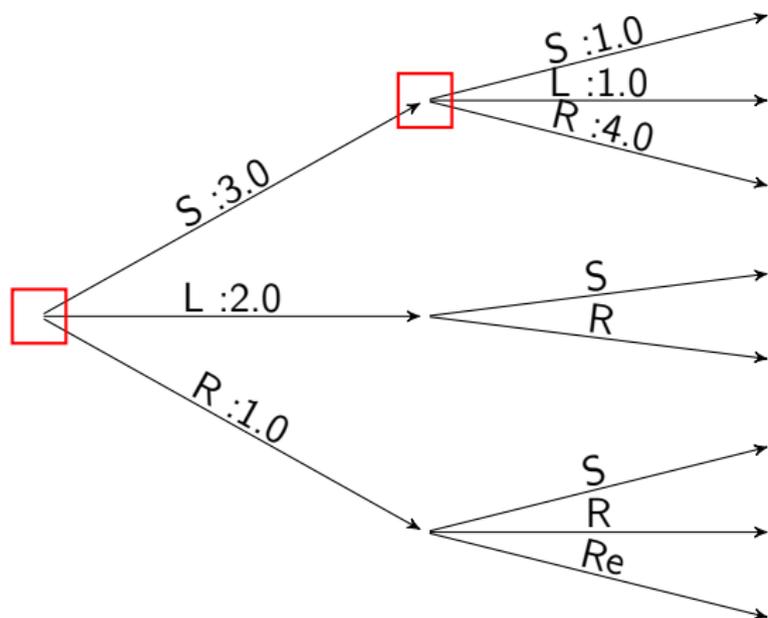
# Recherche en faisceau (*beam*)



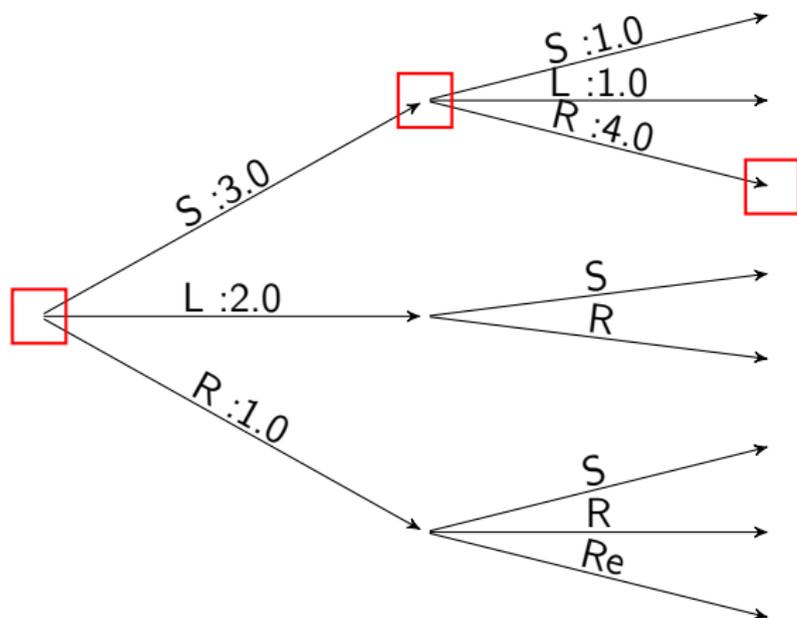
# Recherche en faisceau (*beam*)



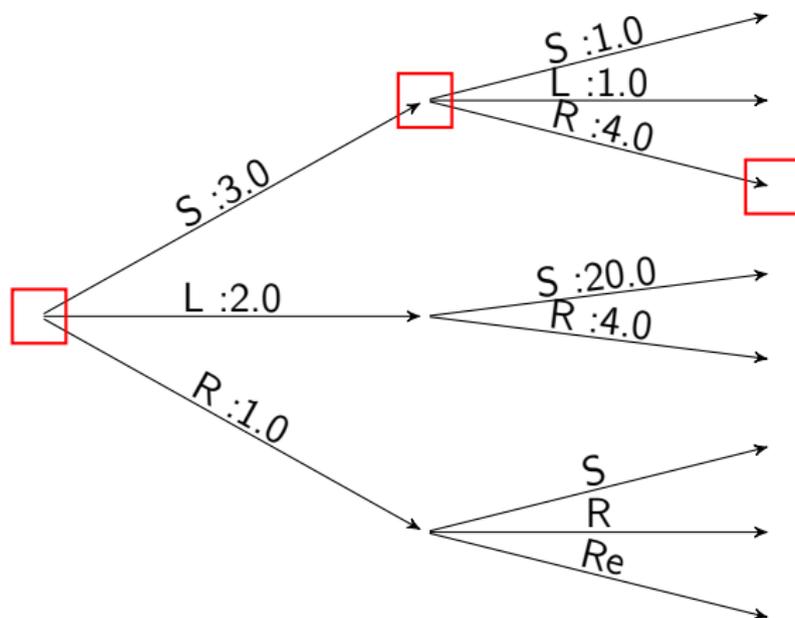
# Recherche en faisceau (*beam*)



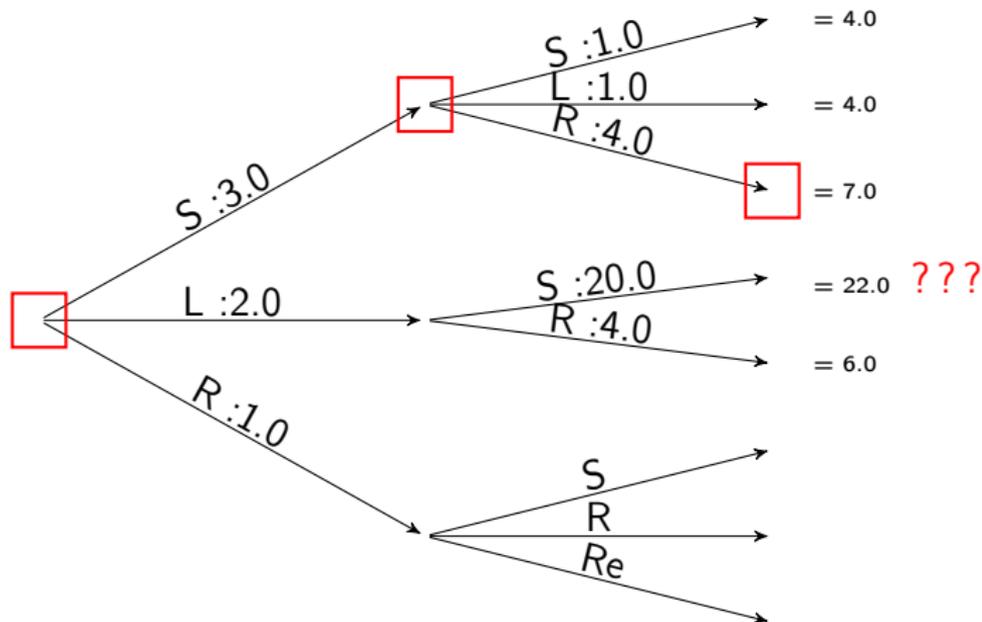
# Recherche en faisceau (*beam*)



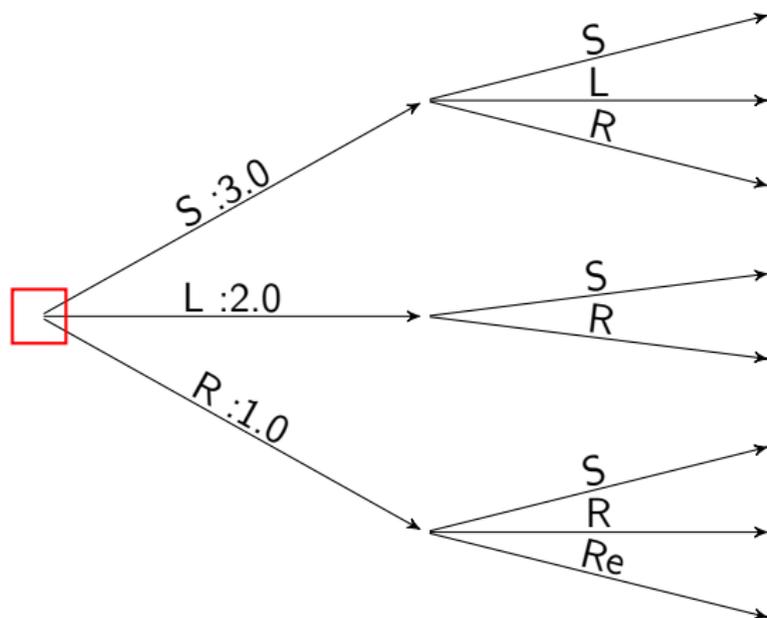
# Recherche en faisceau (*beam*)



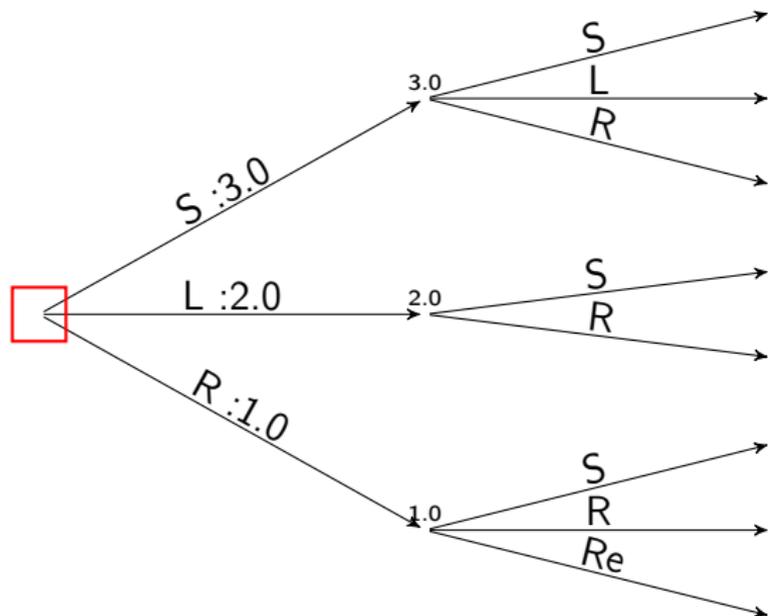
# Recherche en faisceau (*beam*)



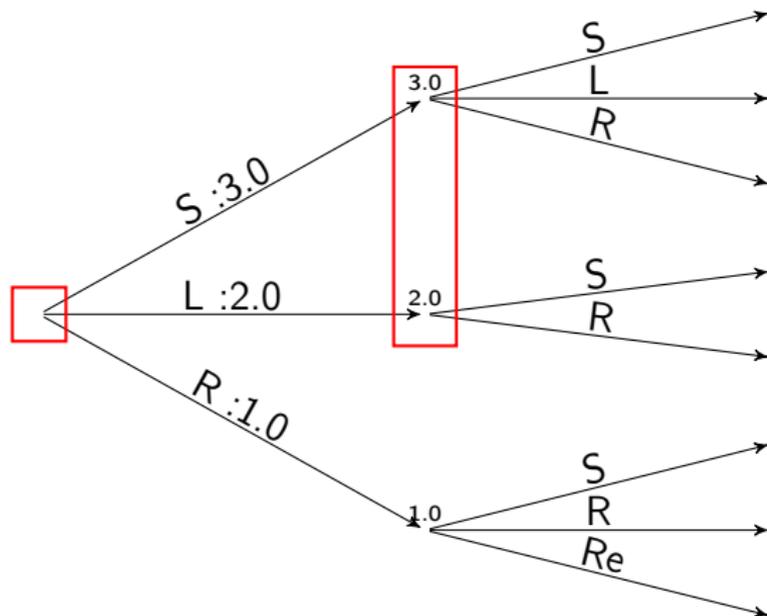
# Recherche en faisceau (*beam*)



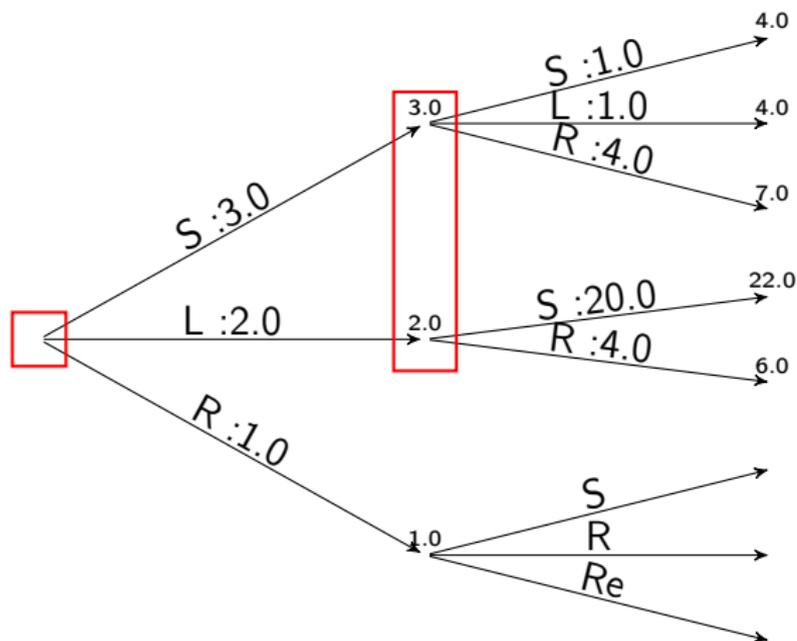
# Recherche en faisceau (*beam*)



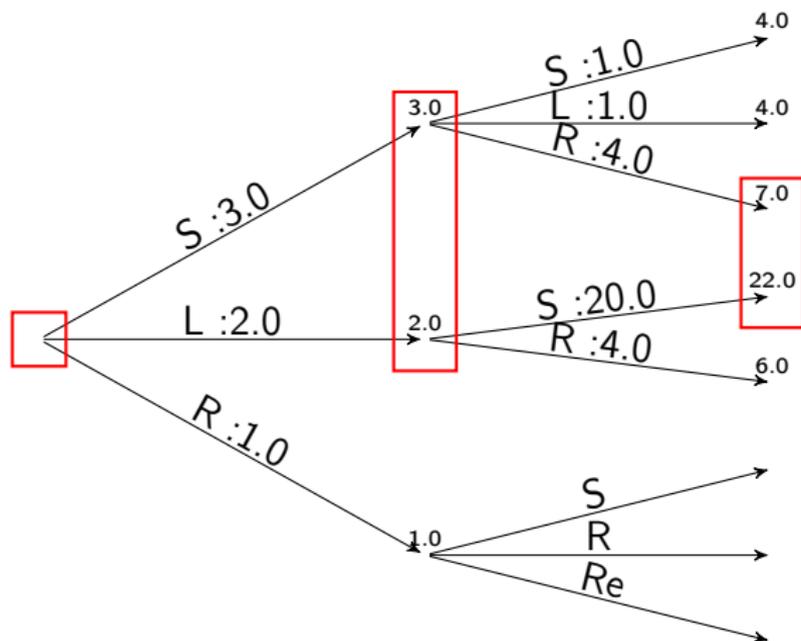
# Recherche en faisceau (*beam*)



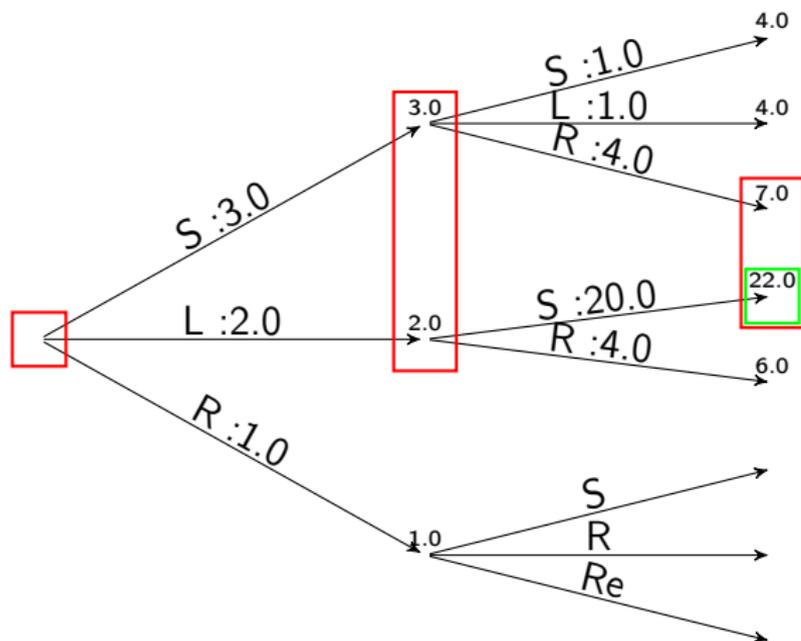
# Recherche en faisceau (*beam*)



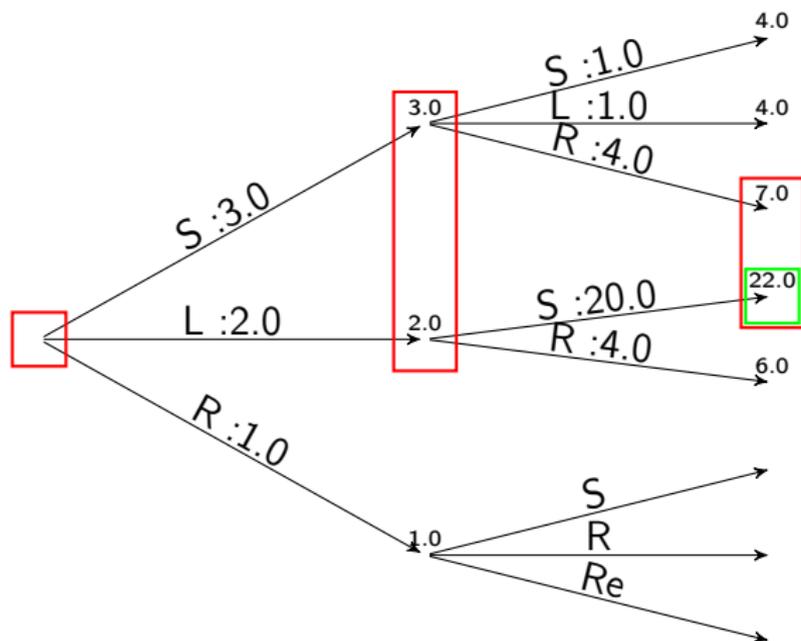
# Recherche en faisceau (*beam*)



# Recherche en faisceau (*beam*)



# Recherche en faisceau (*beam*)

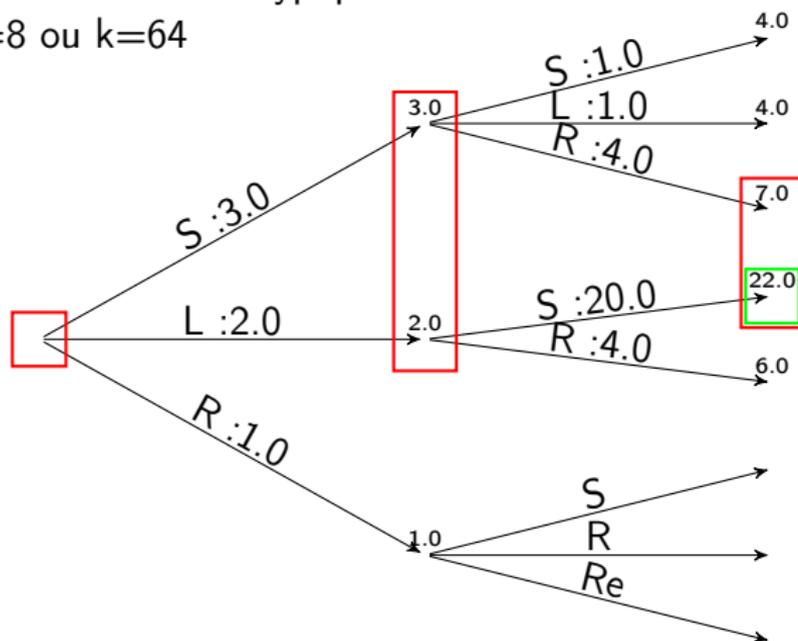


Score global :

$$\Phi_{global} = \sum \phi_{local}$$

# Recherche en faisceau (*beam*)

Taille de faisceau typique :  
k=8 ou k=64

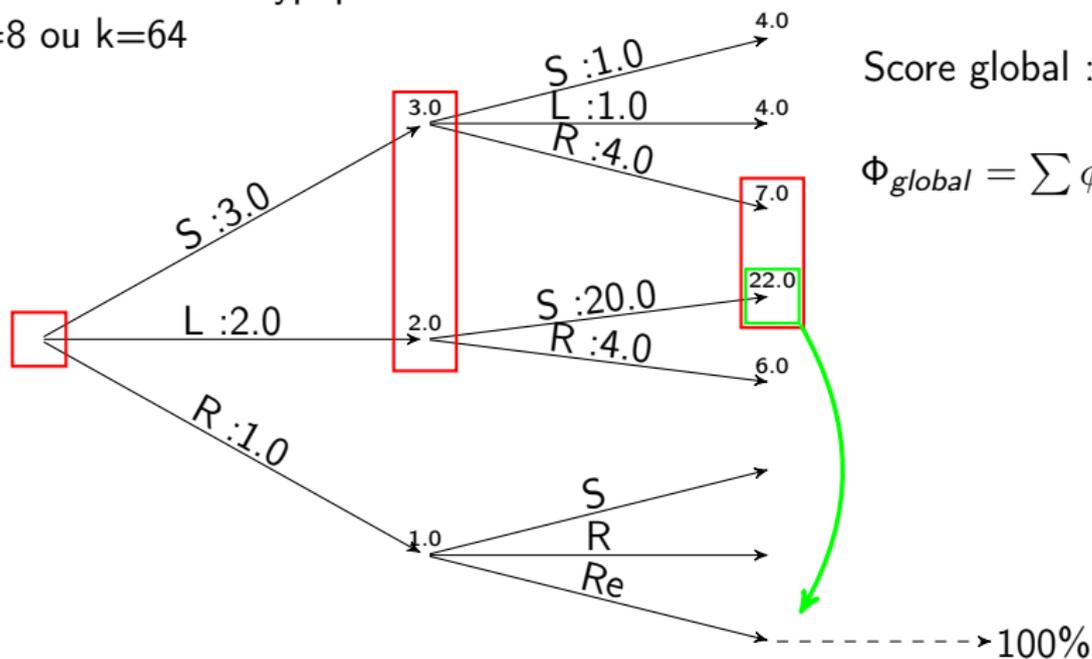


Score global :

$$\Phi_{global} = \sum \phi_{local}$$

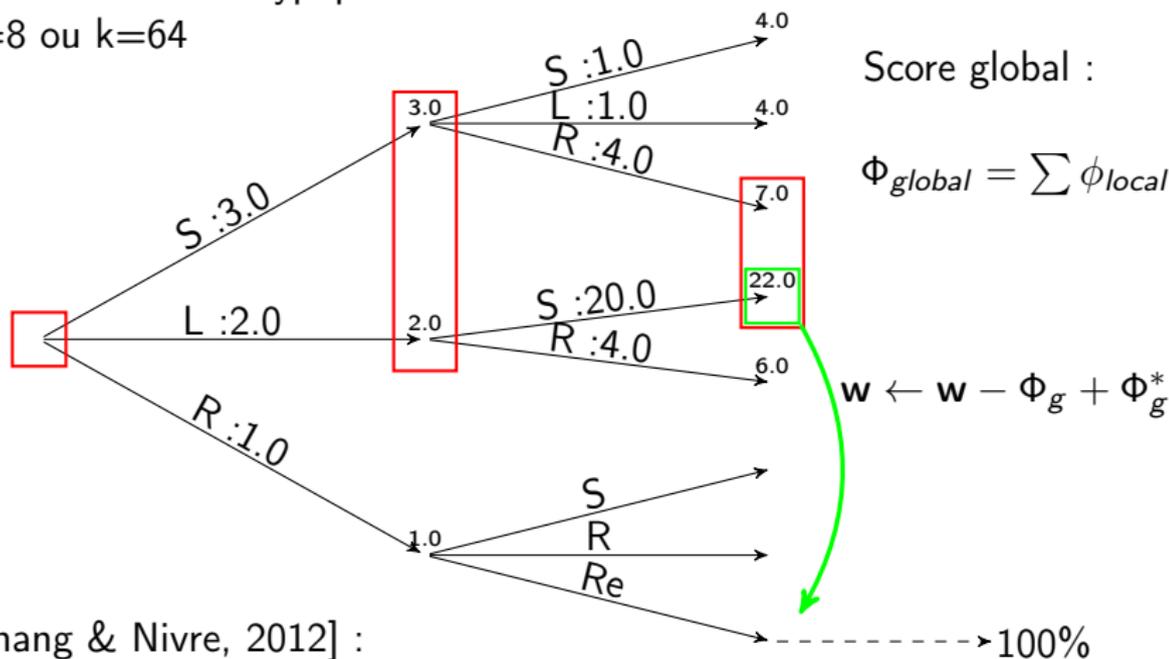
# Apprentissage global

Taille de faisceau typique :  
k=8 ou k=64



# Apprentissage global

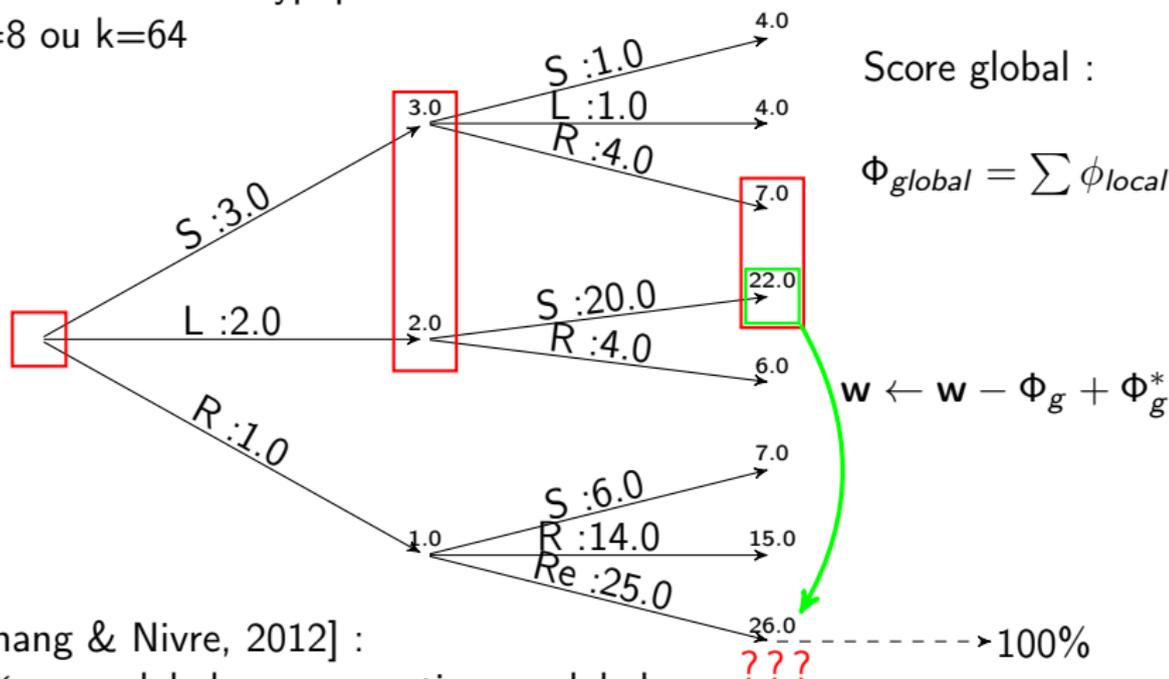
Taille de faisceau typique :  
k=8 ou k=64



[Zhang & Nivre, 2012] :  
inférence globale  $\Leftrightarrow$  apprentissage global

# Apprentissage global

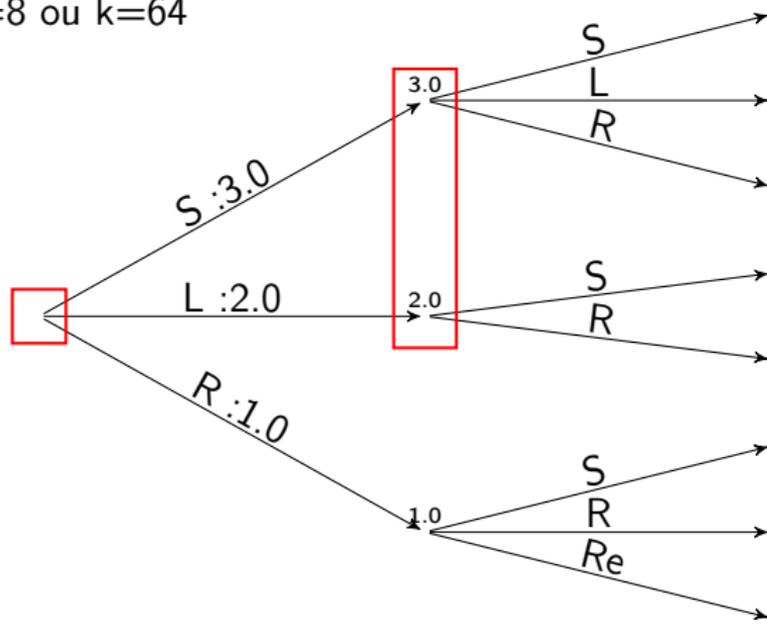
Taille de faisceau typique :  
k=8 ou k=64



[Zhang & Nivre, 2012] :  
inférence globale  $\Leftrightarrow$  apprentissage global

# Apprentissage global

Taille de faisceau typique :  
k=8 ou k=64



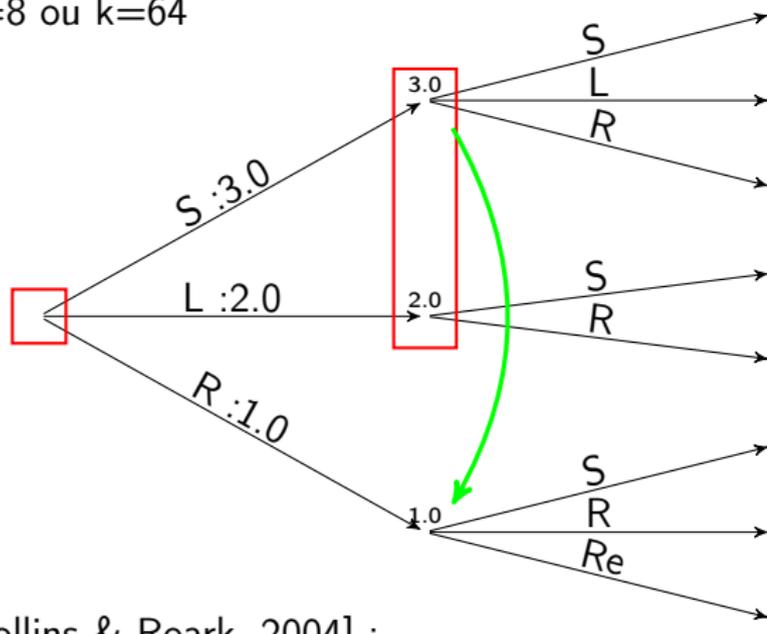
Score global :

$$\Phi_{global} = \sum \phi_{local}$$

$$\mathbf{w} \leftarrow \mathbf{w} - \Phi_g + \Phi_g^*$$

# Apprentissage global

Taille de faisceau typique :  
 $k=8$  ou  $k=64$



Score global :

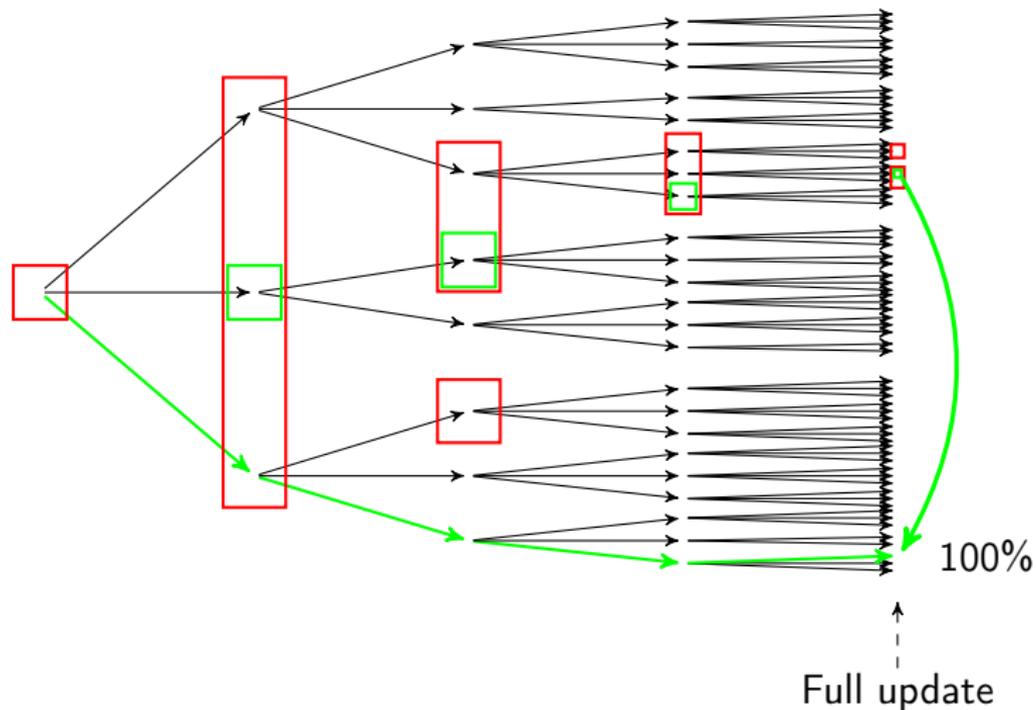
$$\Phi_{global} = \sum \phi_{local}$$

$$\mathbf{w} \leftarrow \mathbf{w} - \Phi_g + \Phi_g^*$$

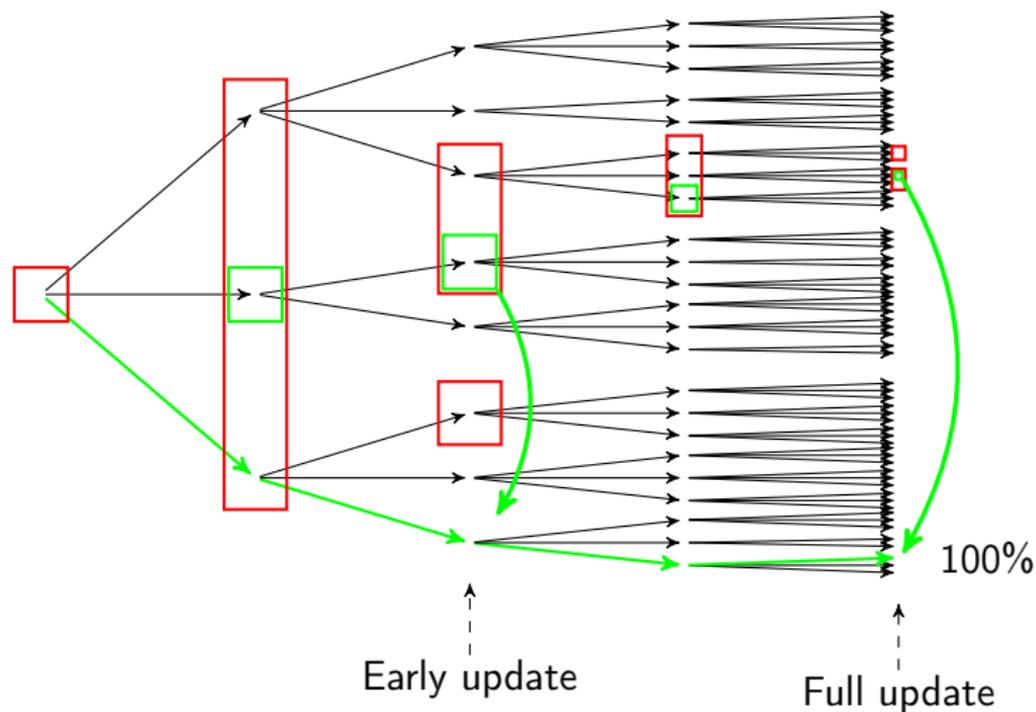
[Collins & Roark, 2004] :  
Early update



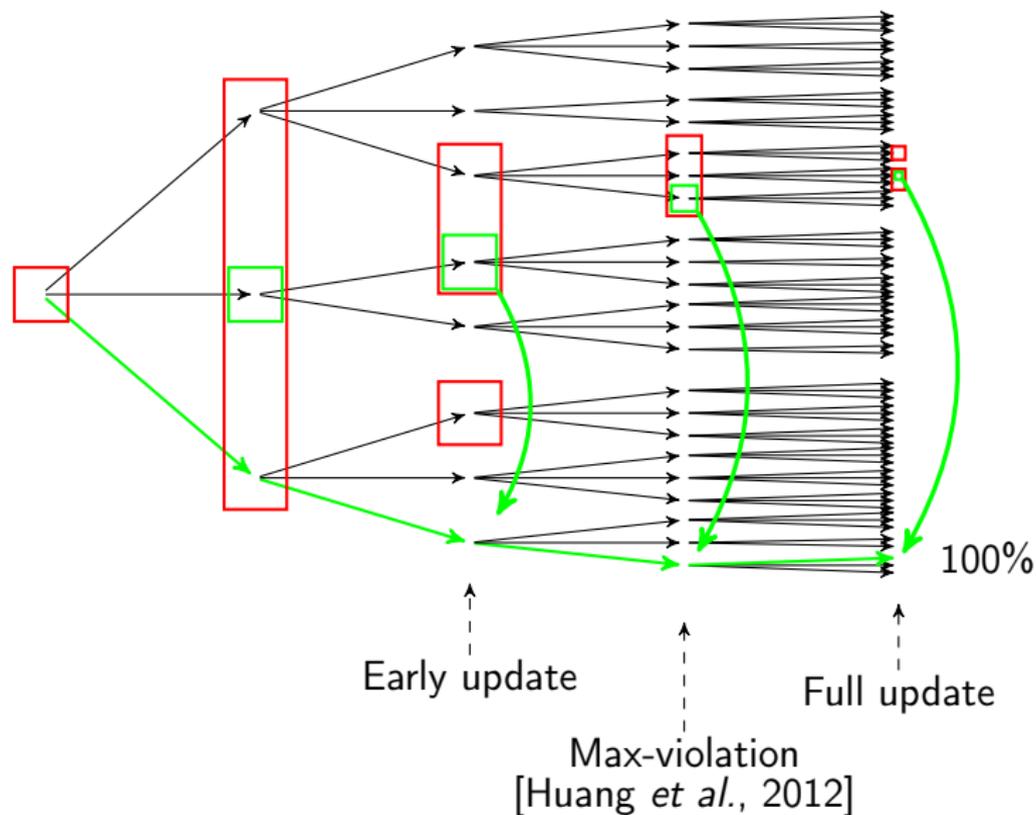
# Méthodes d'apprentissage global



# Méthodes d'apprentissage global

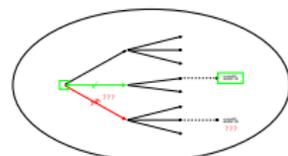


# Méthodes d'apprentissage global

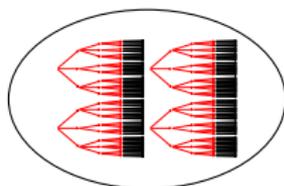


# Limites des méthodes d'apprentissage actuelles

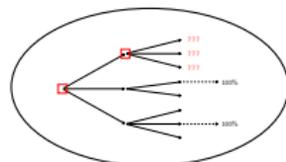
Identification de trois limites en apprentissage global :



Oracle déterministe



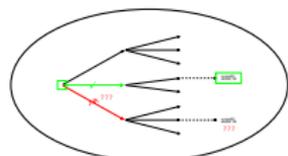
Biais vers les préfixes  
de dérivation



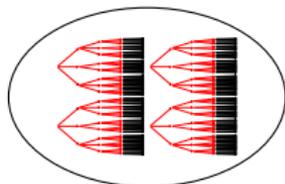
Pas de références imparfaites

# Limites des méthodes d'apprentissage actuelles

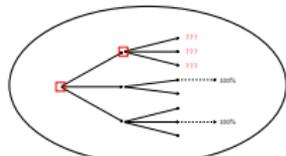
Identification de trois limites en apprentissage global :



Oracle déterministe

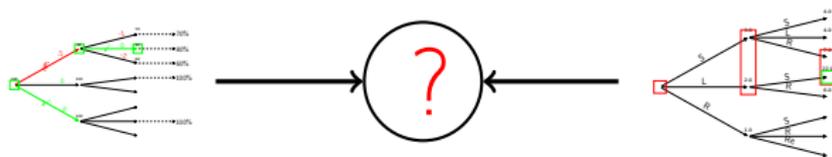


Biais vers les préfixes  
de dérivation



Pas de références imparfaites

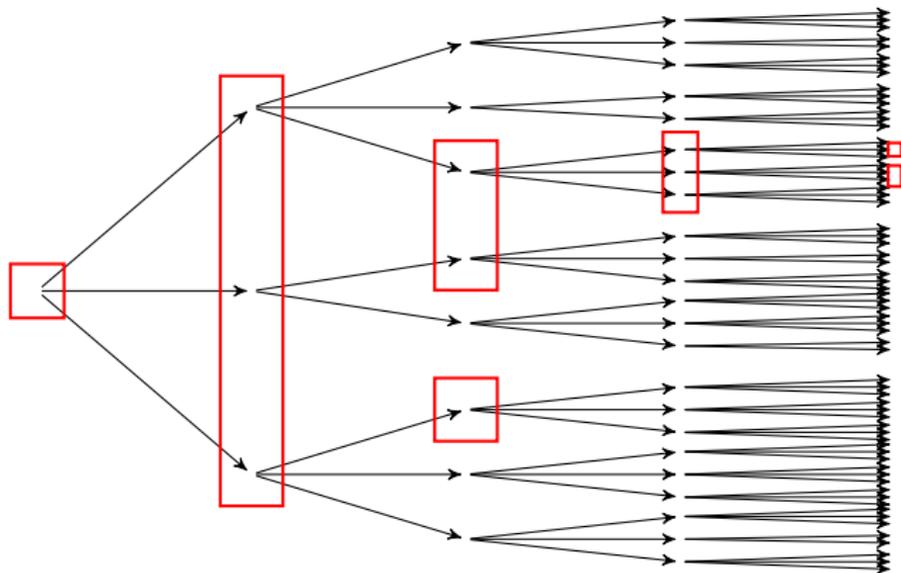
Peut-on combiner oracle dynamique et recherche en faisceau ?



# Plan

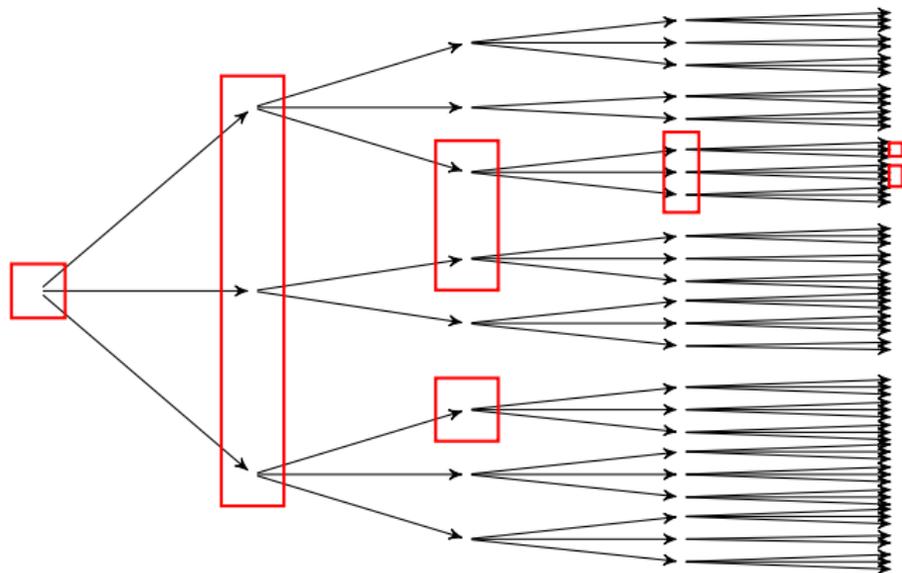
- 1 État de l'art
  - Analyse en dépendances par transition
  - Oracle dynamique glouton
  - Recherche en faisceau et apprentissage global
  - Limites des méthodes d'apprentissage actuelles
- 2 Améliorations proposées
  - Oracle dynamique global
  - Reprise d'apprentissage
  - Références dans l'espace sous-optimal
- 3 Validation expérimentale
  - Cadre expérimental
  - Résultats
  - Analyse de similarité
- 4 Conclusions

# Amélioration 1 : Oracle dynamique global



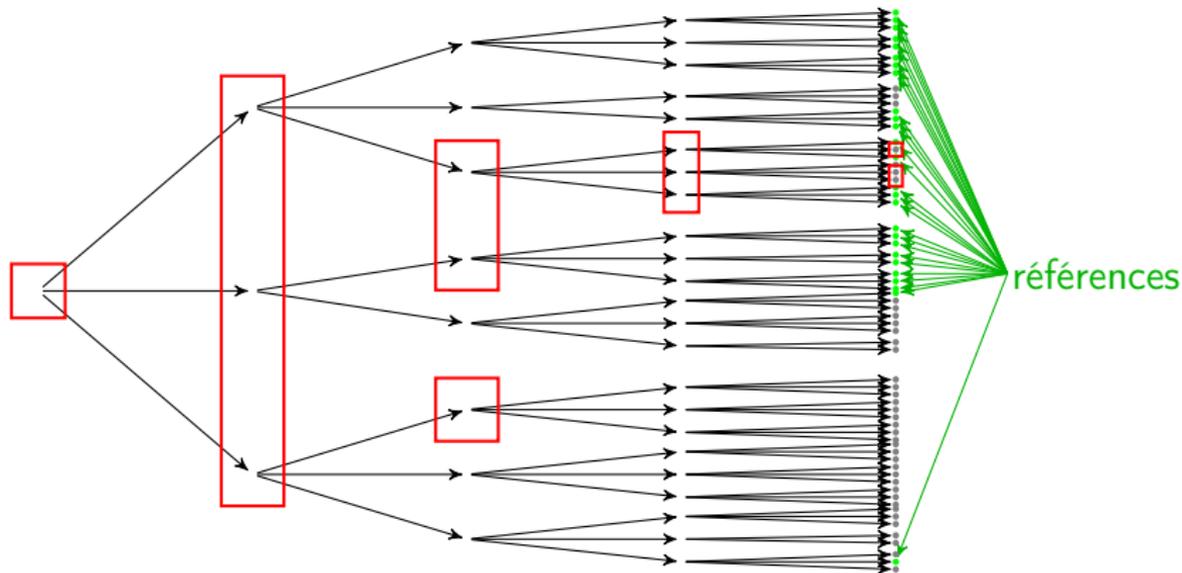
# Amélioration 1 : Oracle dynamique global

Oracle = détection des erreurs + choix de la référence



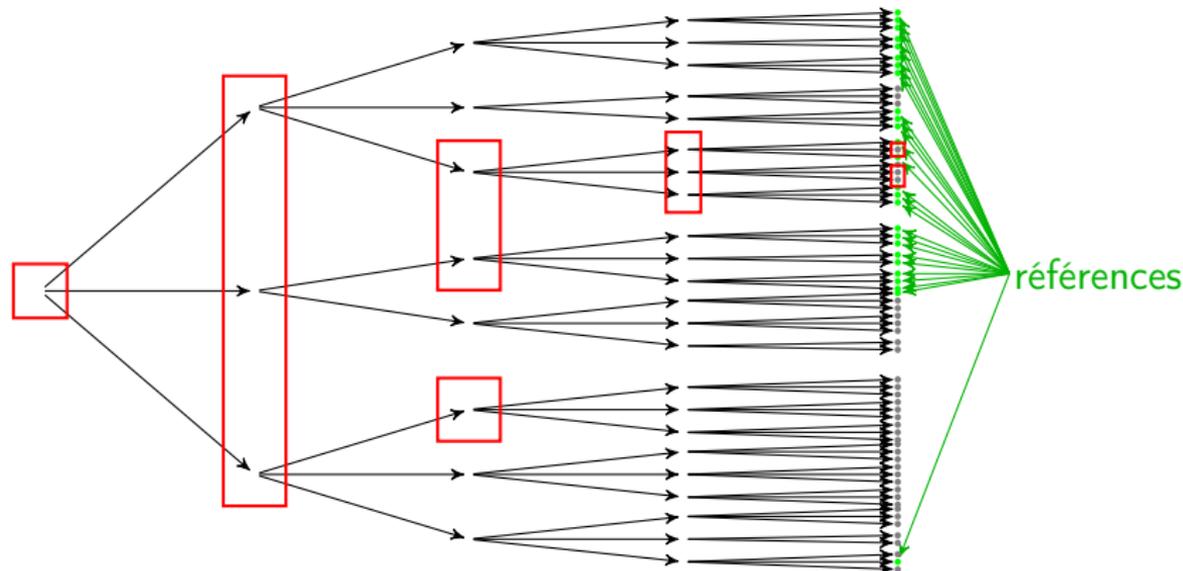
# Amélioration 1 : Oracle dynamique global

Oracle = détection des erreurs + choix de la référence



# Amélioration 1 : Oracle dynamique global

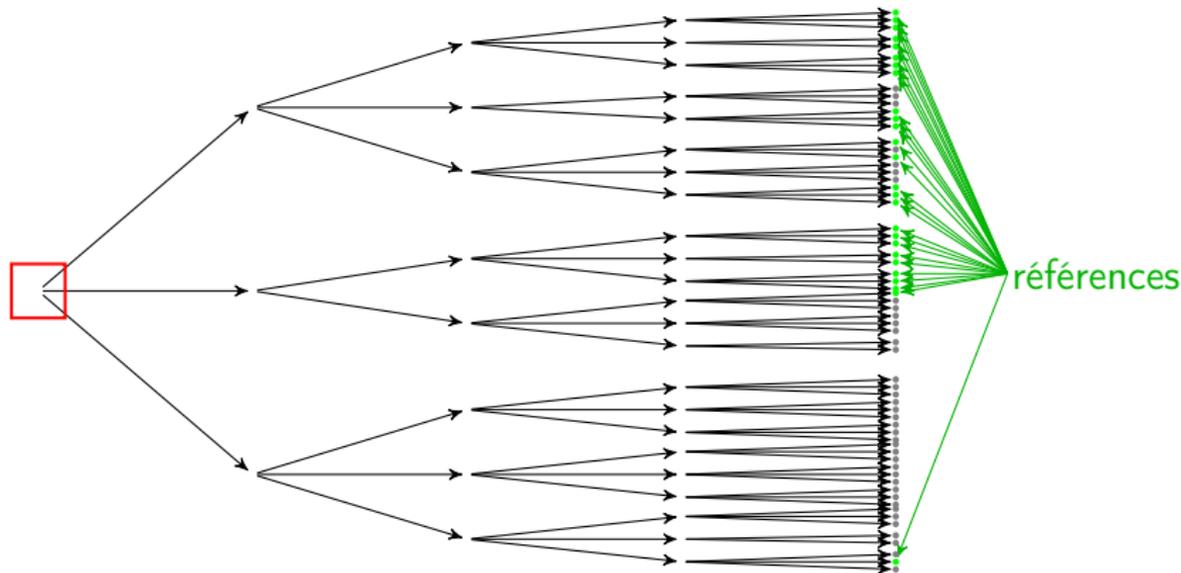
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

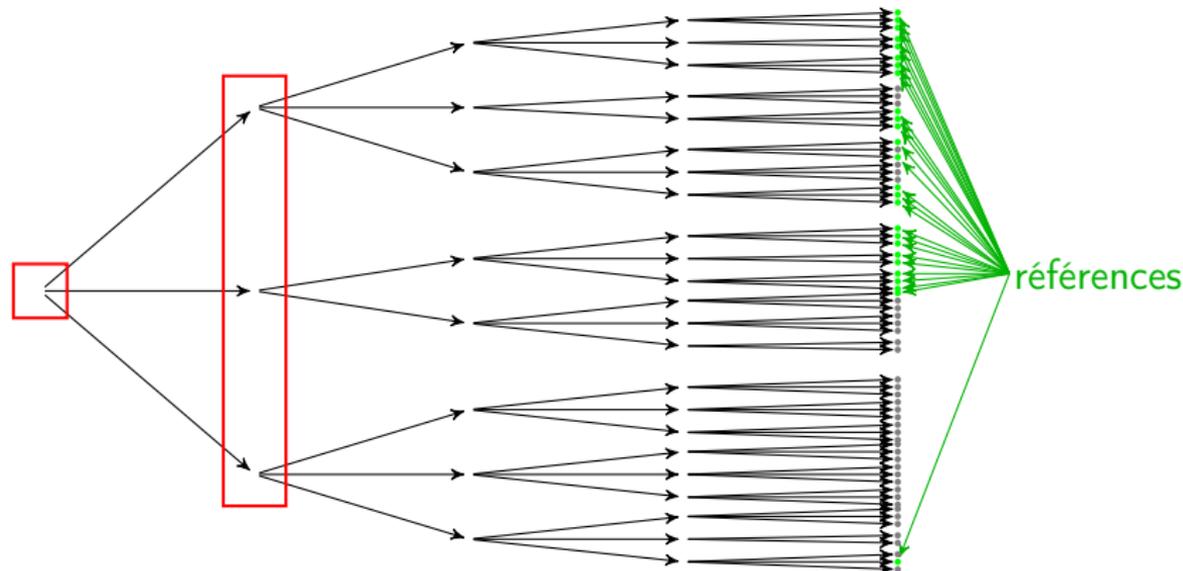
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

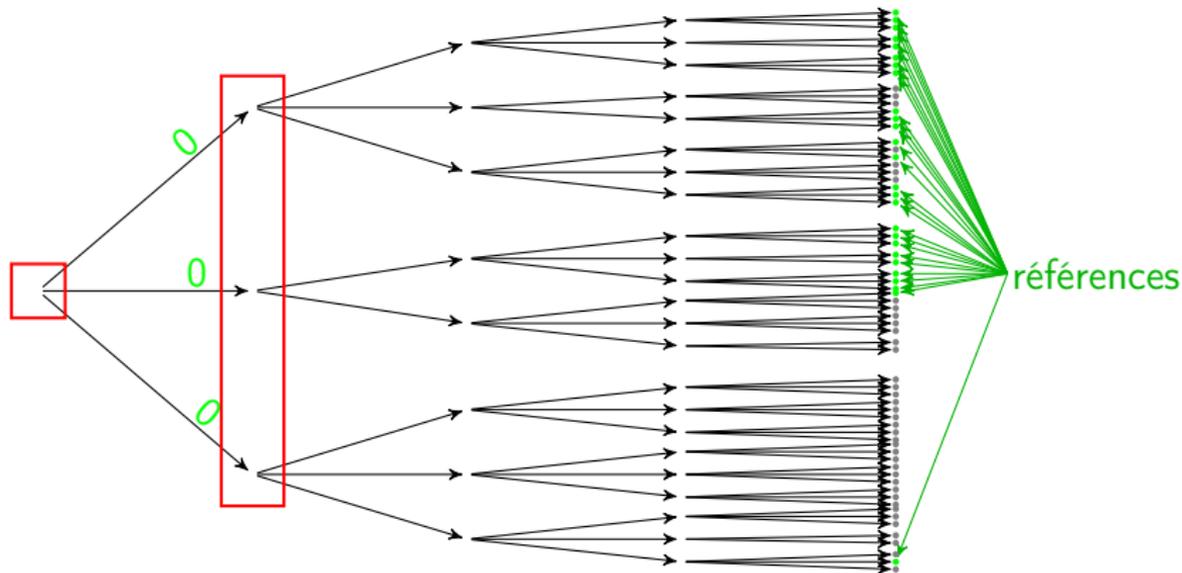
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

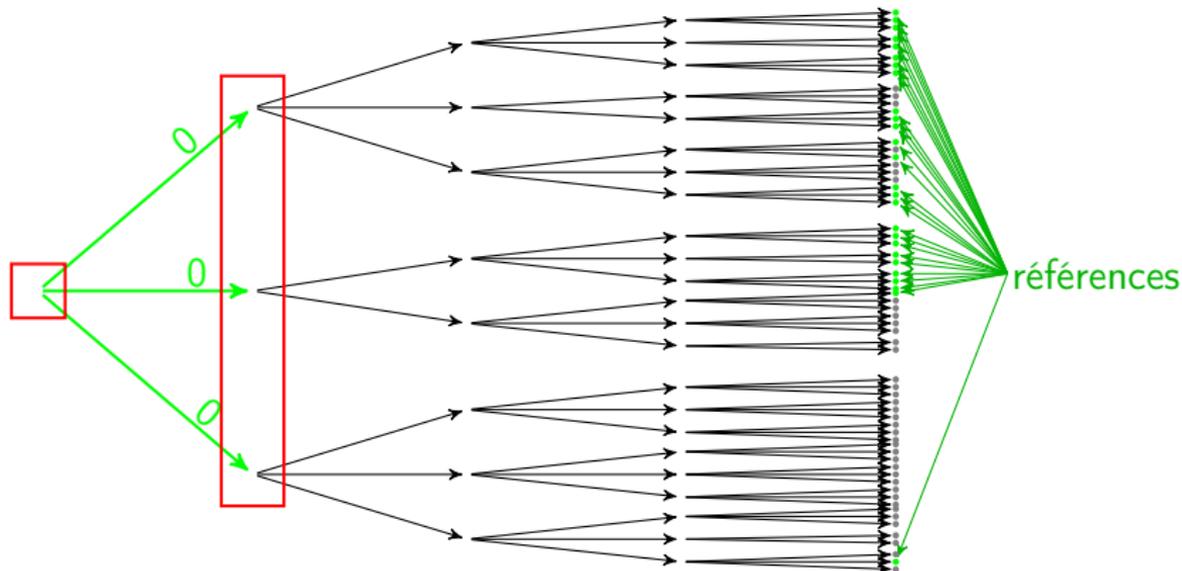
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

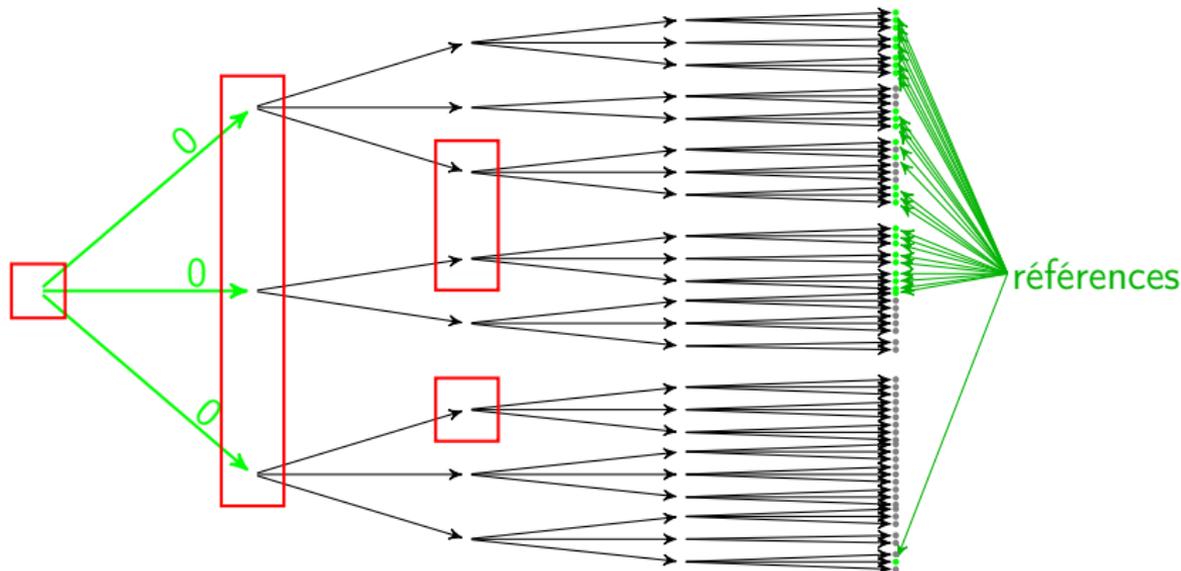
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

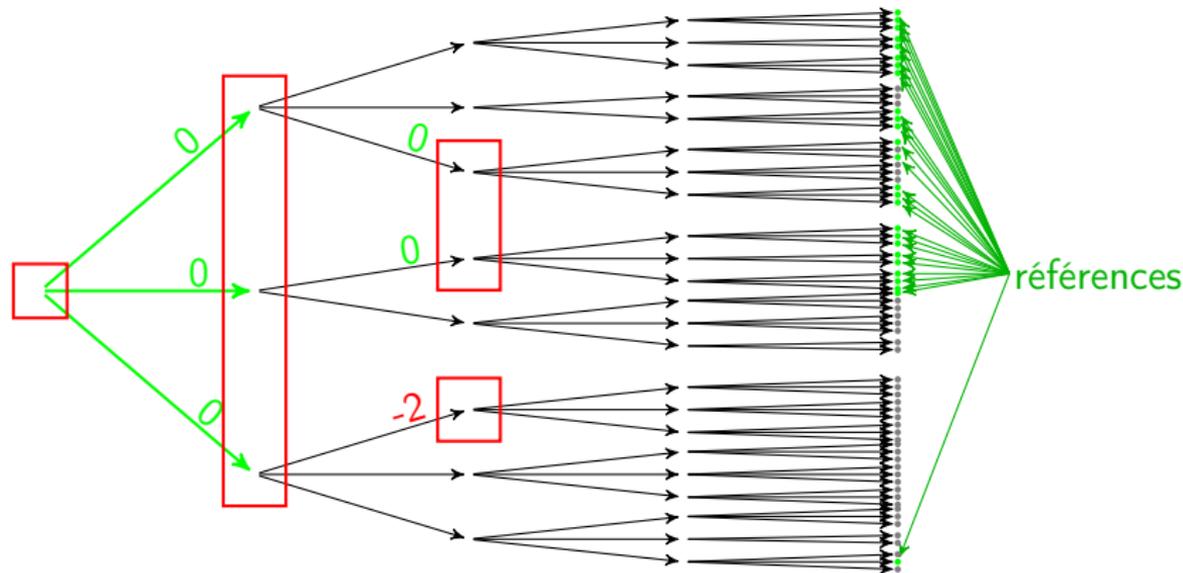
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

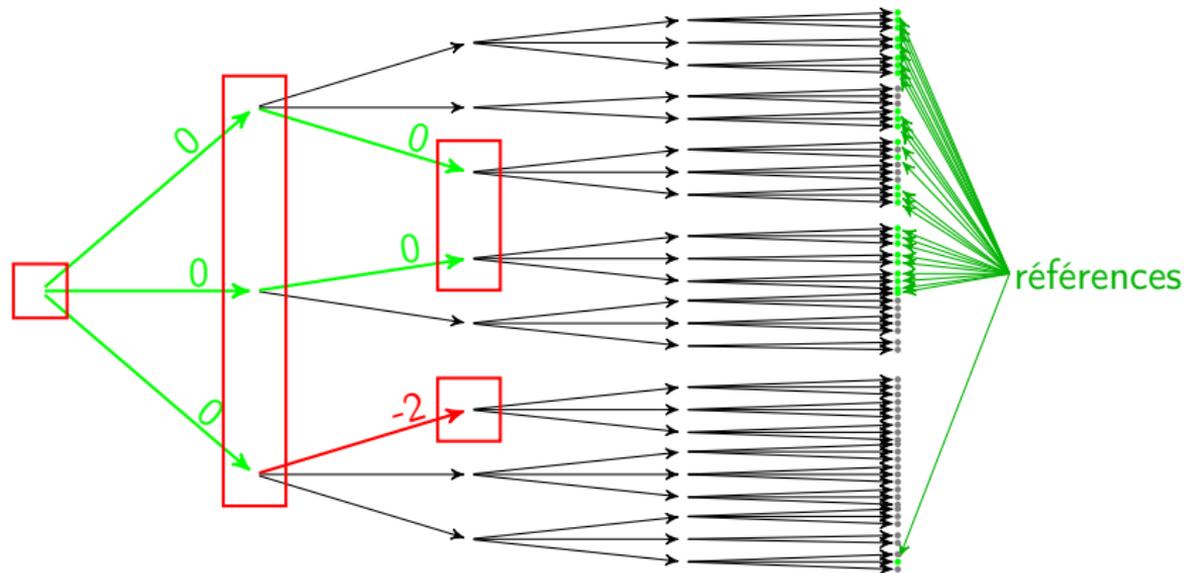
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

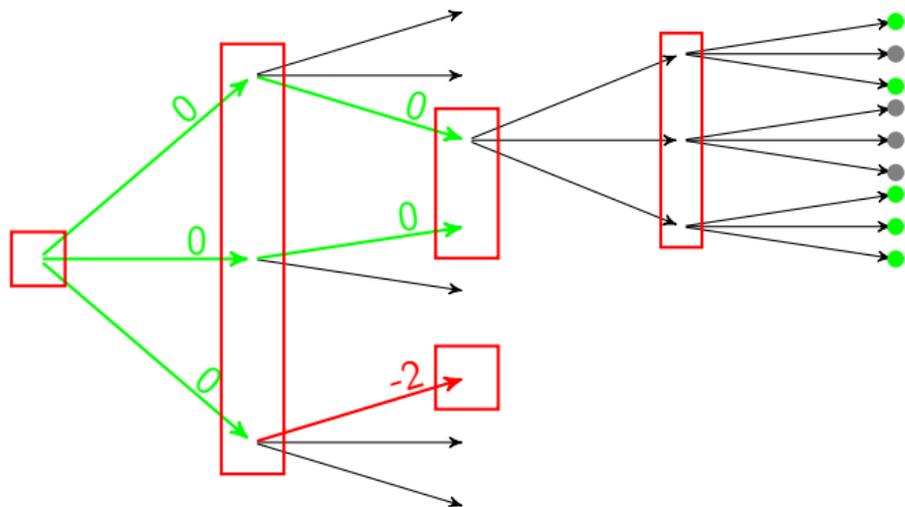
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

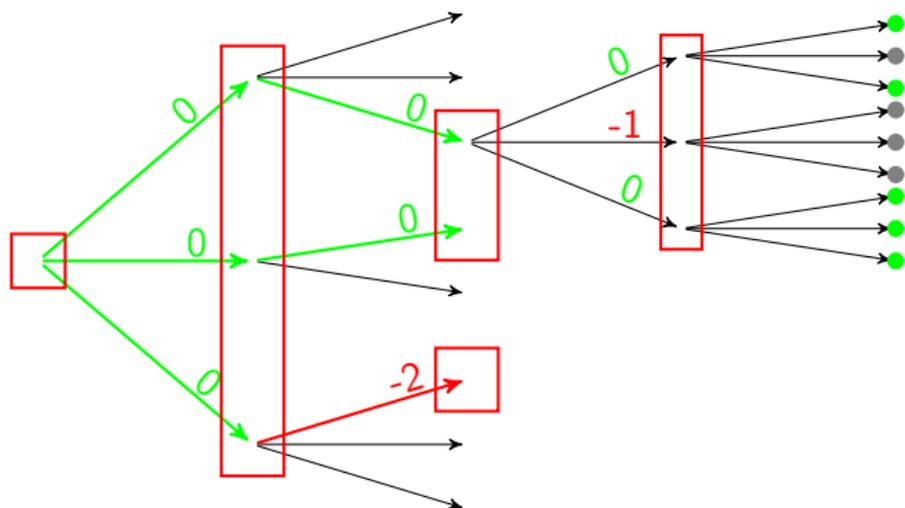
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

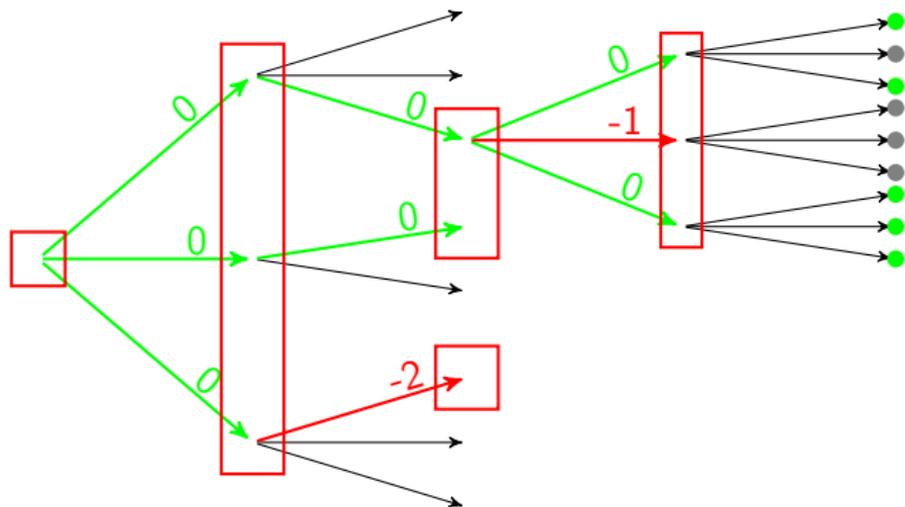
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

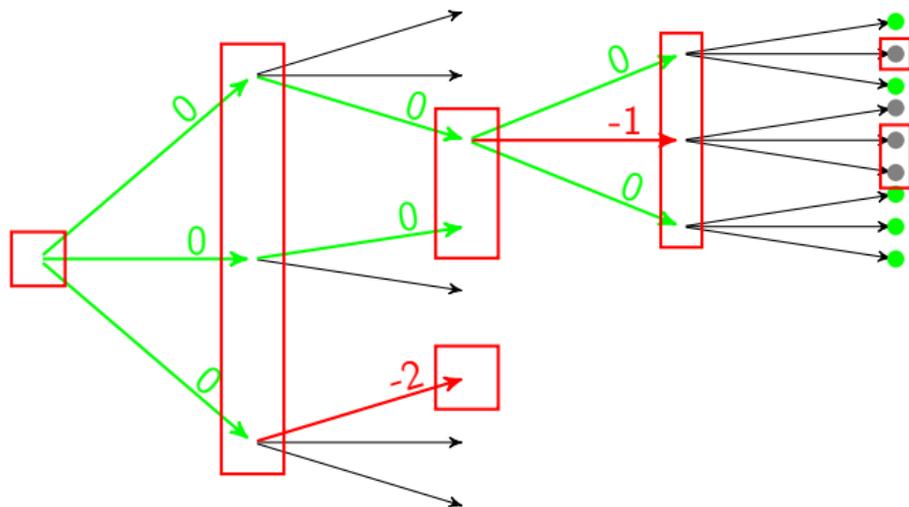
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

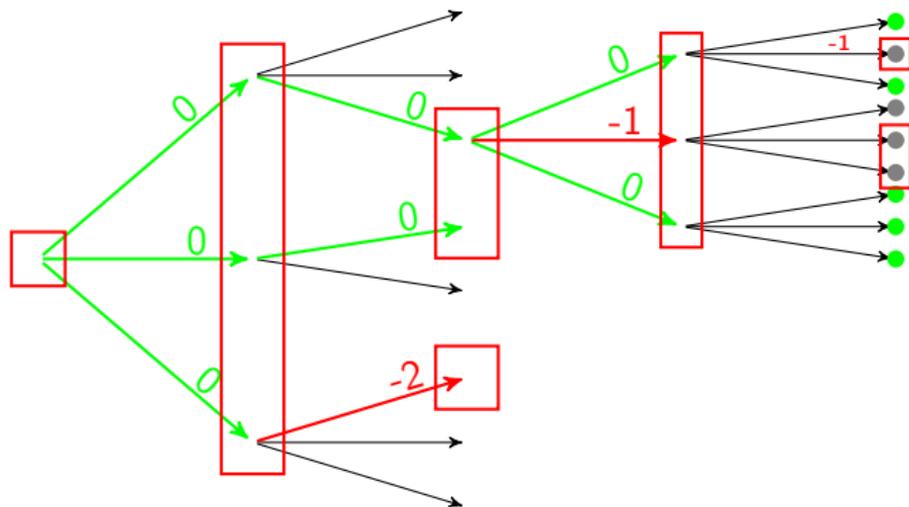
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

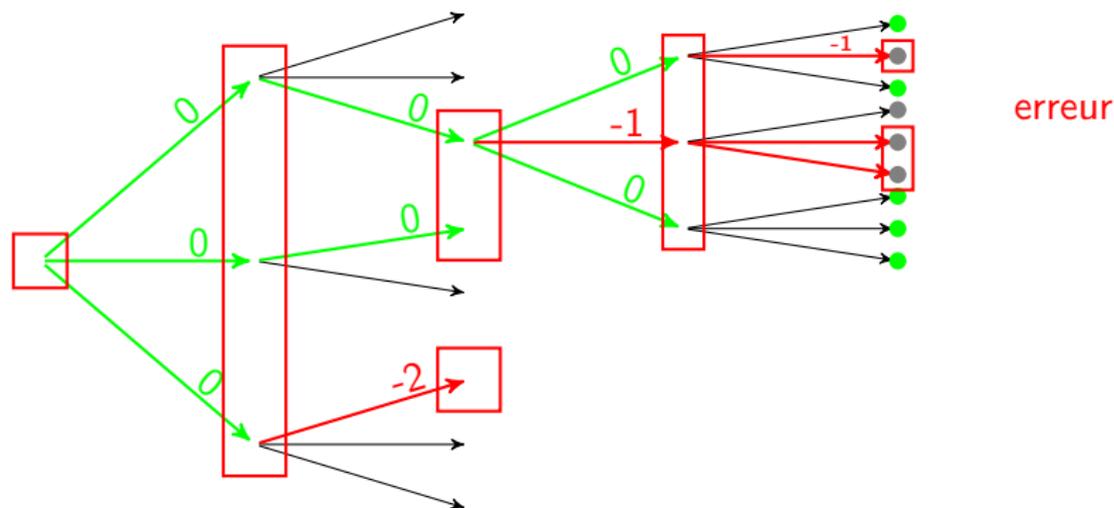
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

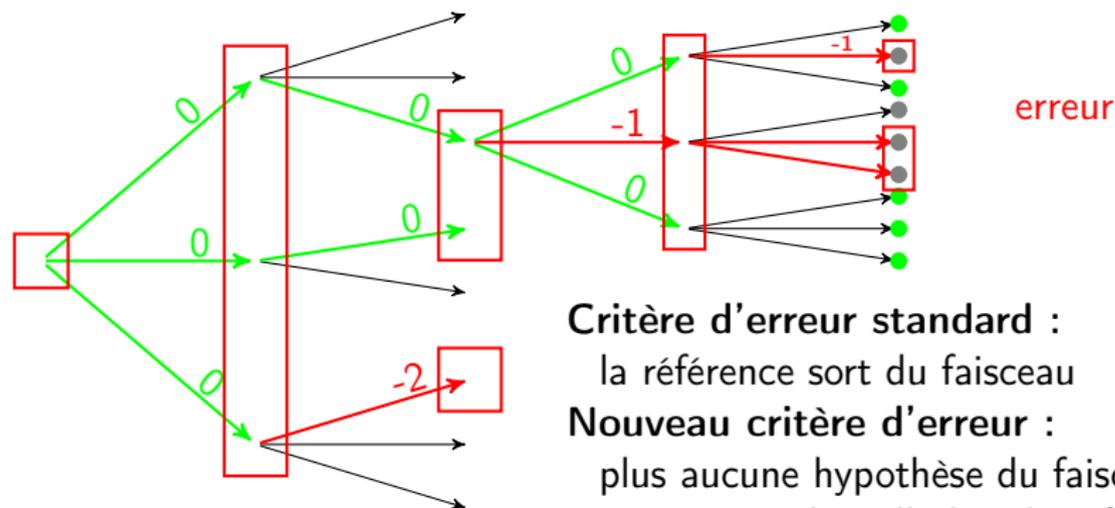
Oracle = détection des erreurs + choix de la référence



Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

Oracle = détection des erreurs + choix de la référence



**Critère d'erreur standard :**

la référence sort du faisceau

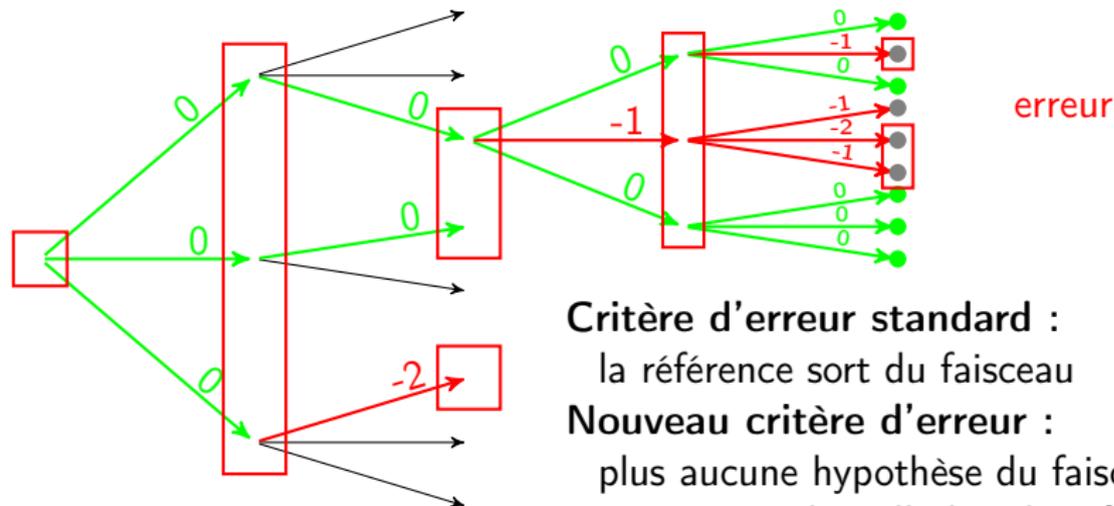
**Nouveau critère d'erreur :**

plus aucune hypothèse du faisceau  
ne peut produire l'arbre de référence

Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

# Amélioration 1 : Oracle dynamique global

Oracle = détection des erreurs + choix de la référence



**Critère d'erreur standard :**

la référence sort du faisceau

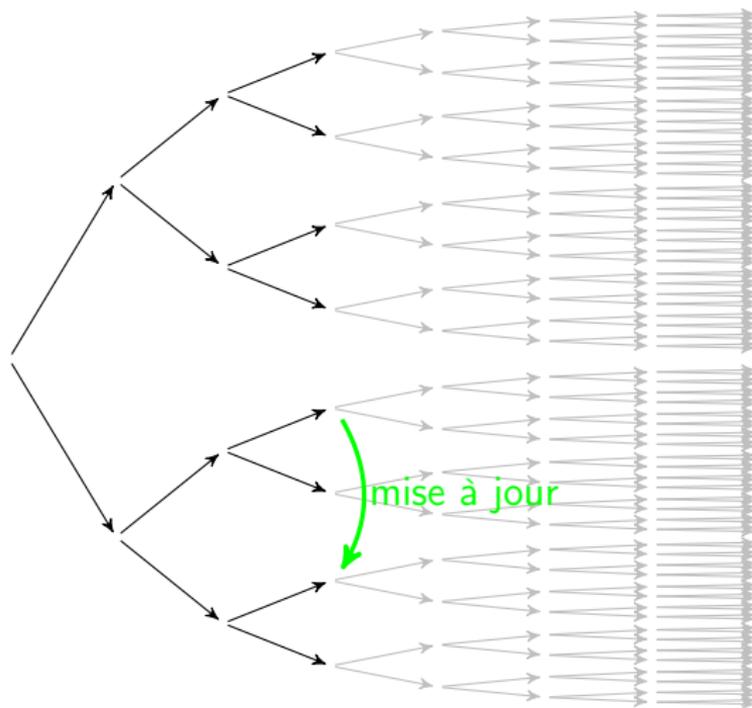
**Nouveau critère d'erreur :**

plus aucune hypothèse du faisceau  
ne peut produire l'arbre de référence

Sans construire de référence,  
peut-on déterminer si une séquence d'actions donnée est correcte ?

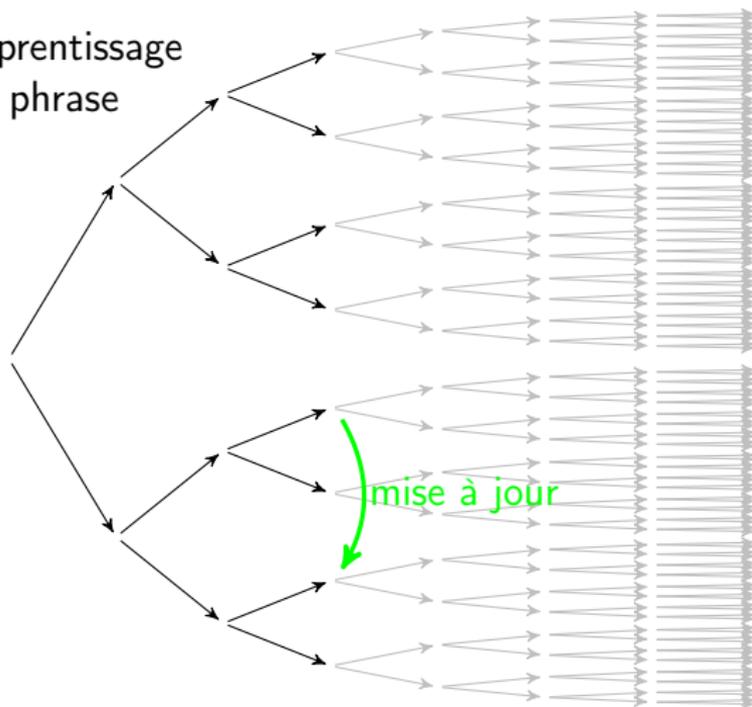


## Amélioration 2 : Reprise d'apprentissage

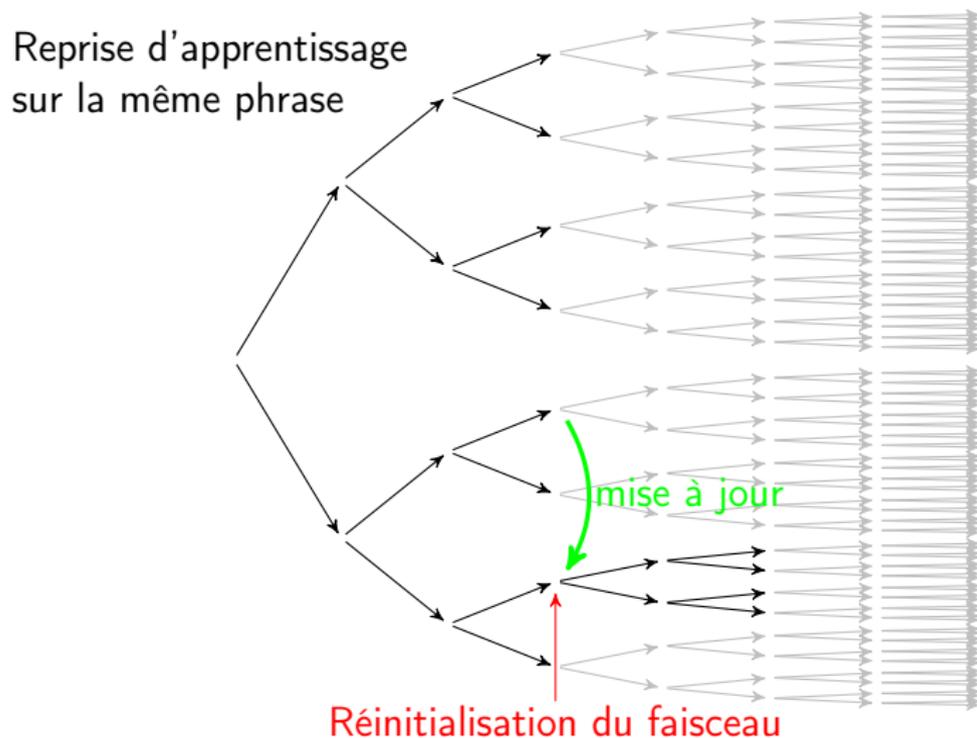


## Amélioration 2 : Reprise d'apprentissage

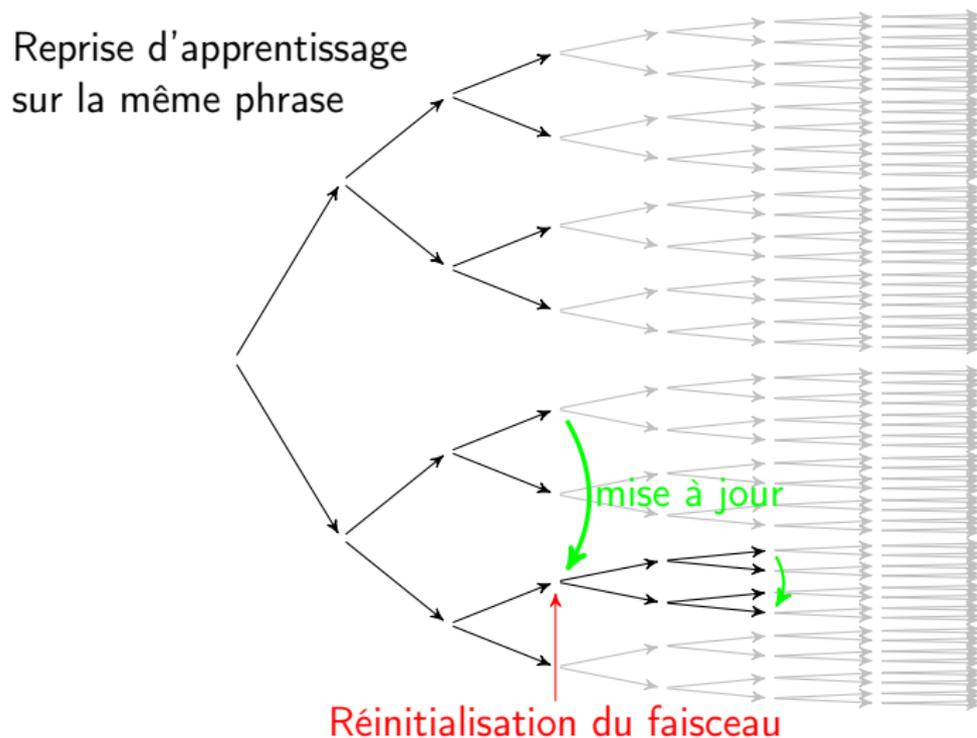
Reprise d'apprentissage  
sur la même phrase



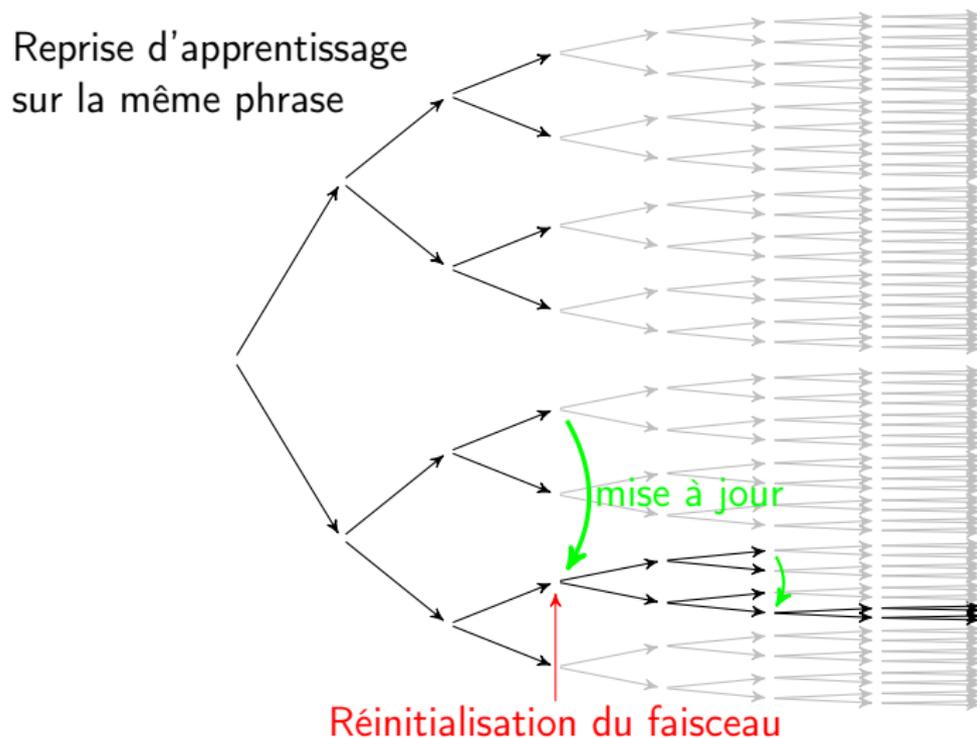
## Amélioration 2 : Reprise d'apprentissage



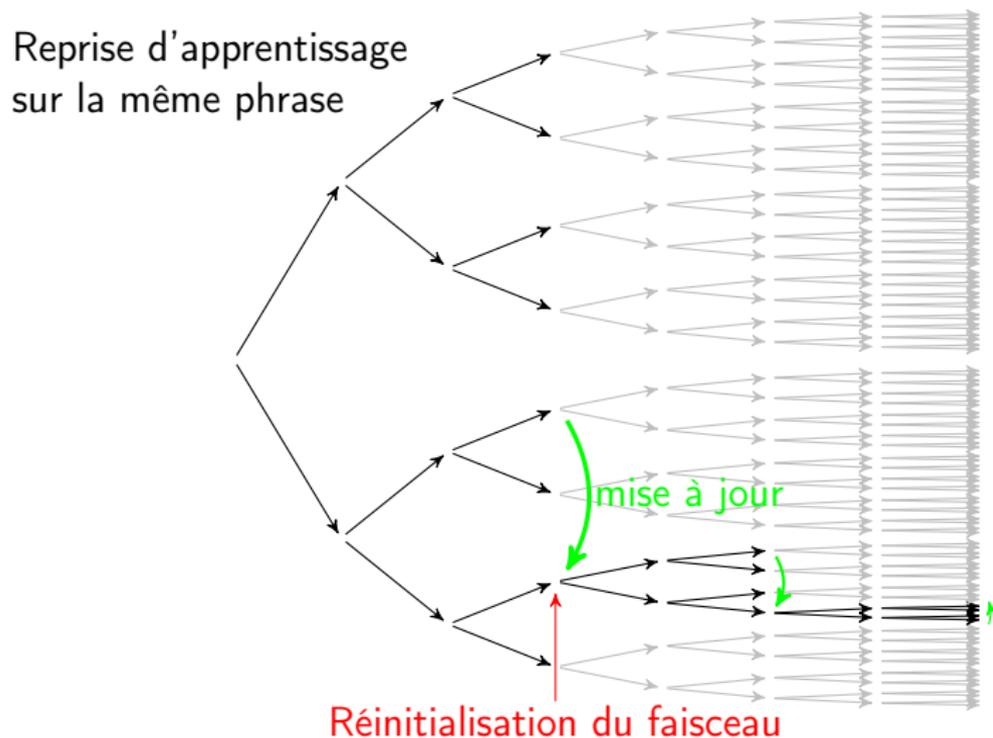
## Amélioration 2 : Reprise d'apprentissage



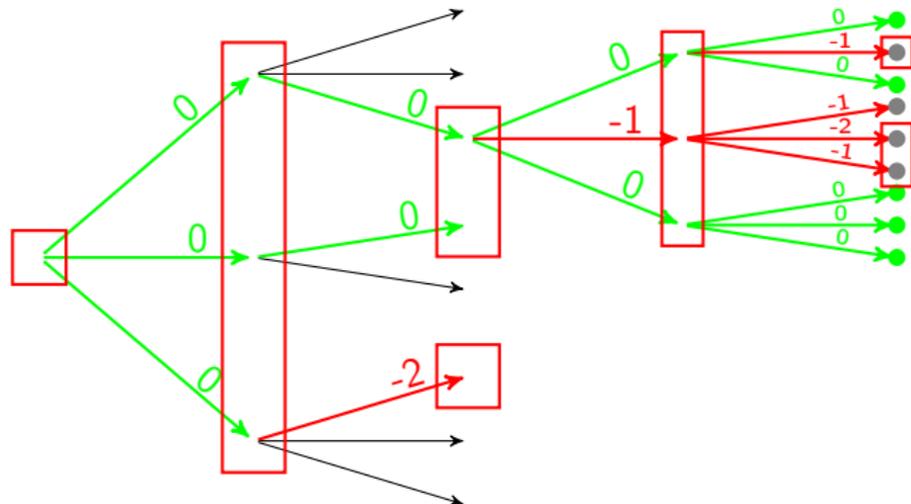
## Amélioration 2 : Reprise d'apprentissage



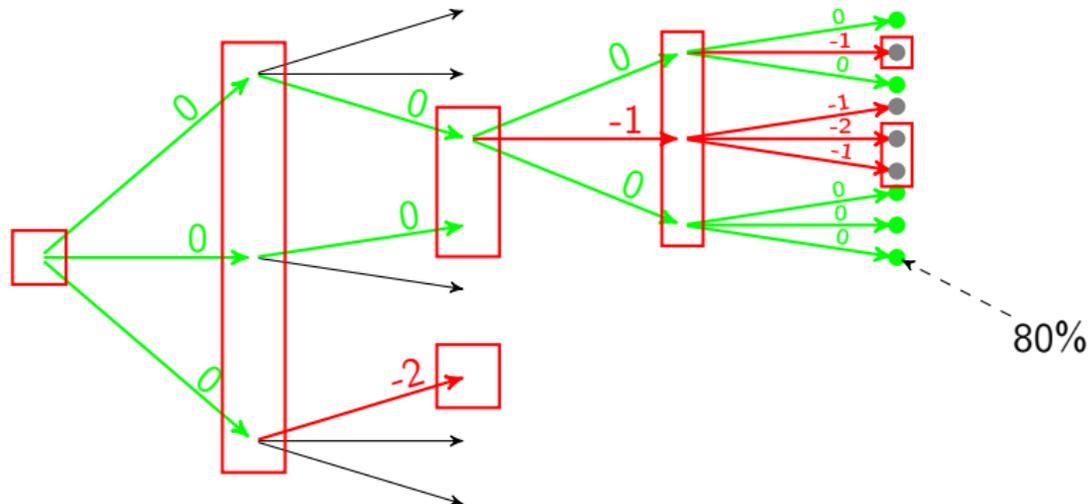
## Amélioration 2 : Reprise d'apprentissage



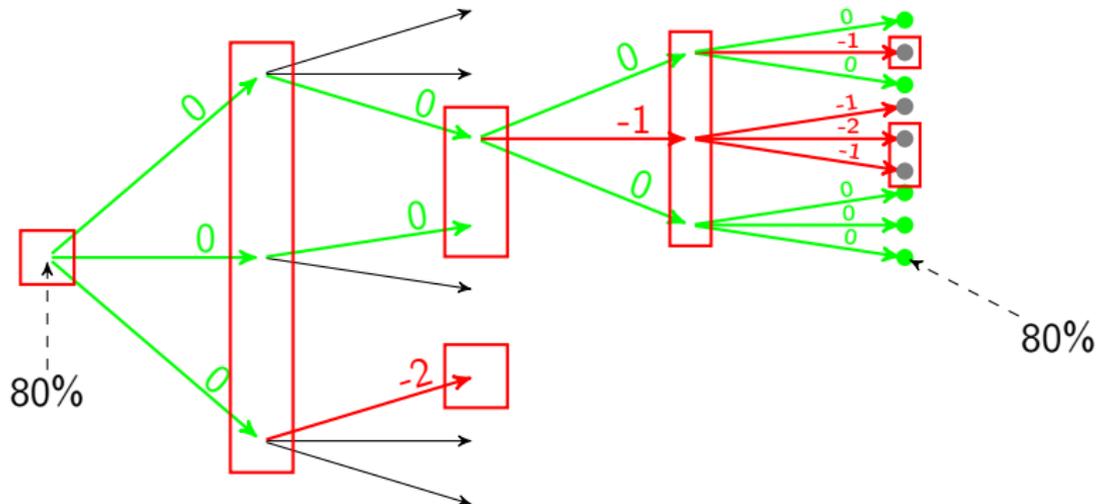
## Amélioration 3 : Références dans l'espace sous-optimal



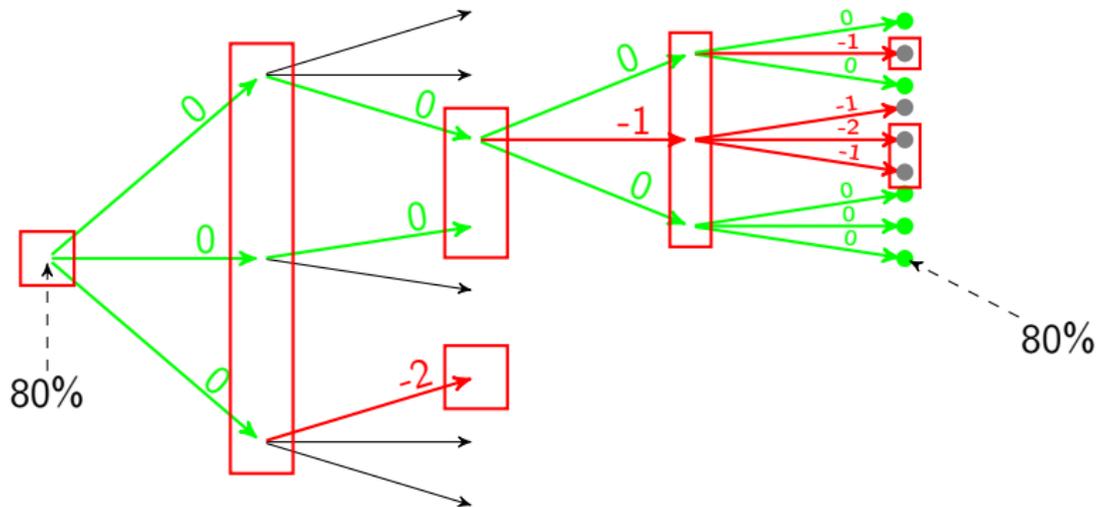
## Amélioration 3 : Références dans l'espace sous-optimal



## Amélioration 3 : Références dans l'espace sous-optimal

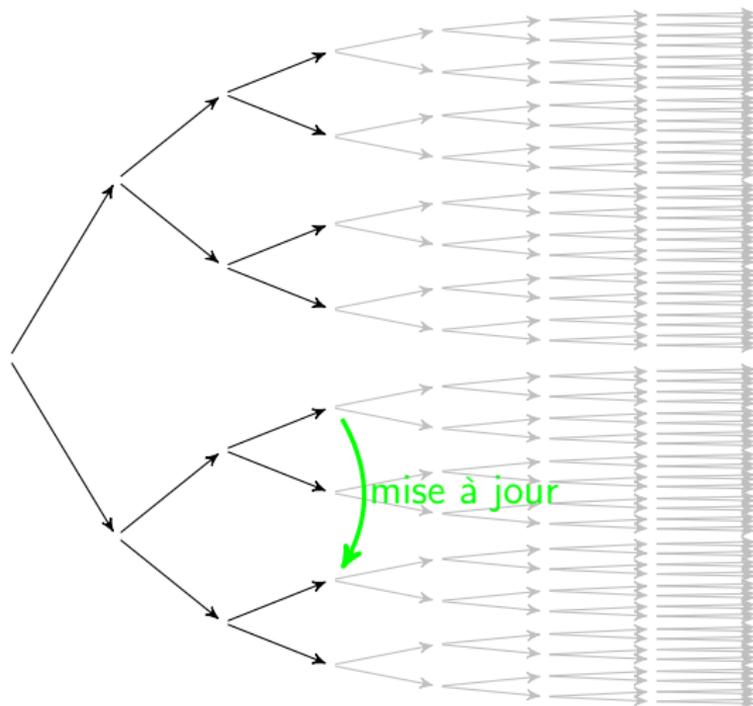


## Amélioration 3 : Références dans l'espace sous-optimal



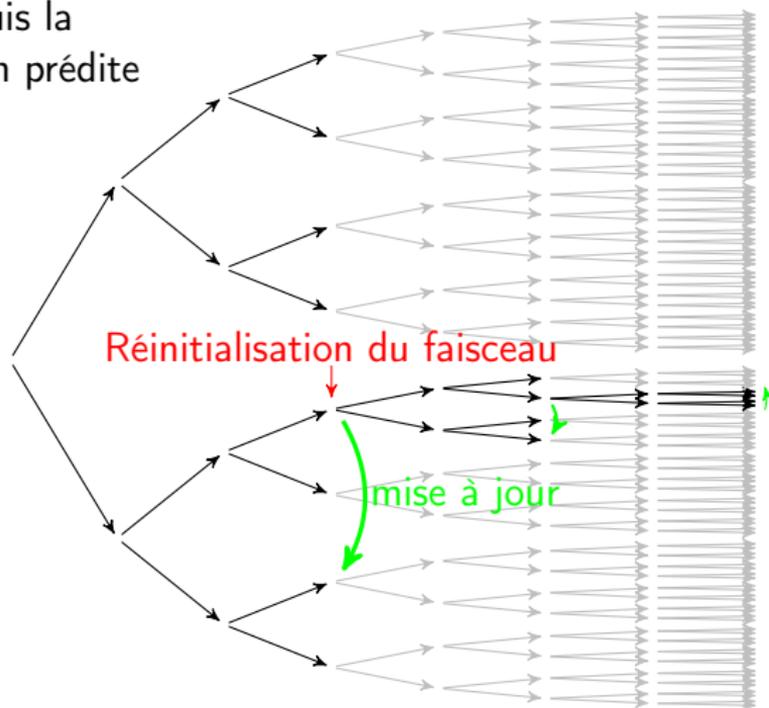
⇒ complétude de l'oracle dynamique

## Amélioration 3 : Références dans l'espace sous-optimal



## Amélioration 3 : Références dans l'espace sous-optimal

Reprise depuis la configuration prédite



# Plan

- 1 État de l'art
  - Analyse en dépendances par transition
  - Oracle dynamique glouton
  - Recherche en faisceau et apprentissage global
  - Limites des méthodes d'apprentissage actuelles
- 2 Améliorations proposées
  - Oracle dynamique global
  - Reprise d'apprentissage
  - Références dans l'espace sous-optimal
- 3 Validation expérimentale
  - Cadre expérimental
  - Résultats
  - Analyse de similarité
- 4 Conclusions

- Expériences sur 9 langues : corpus SPMRL
- Validation successive des trois améliorations proposées
- Comparaison avec les systèmes de l'état de l'art :
  - glouton dynamique
  - early update
  - max-violation

# Résultats (Unlabeled Attachment Score)

- 1 ND = oracle non-déterministe
- 2 + = reprise d'apprentissage
- 3 SO = références dans l'espace sous-optimal

Français	Classique	ND	+ ND	+ ND SO
Glouton dynamique				<b>84,23</b>
Early update	<b>86,02</b>	85,85	86,18	<b>86,26</b>
Max-violation	<b>86,10</b>	86,13	86,19	<b>86,26</b>

Gain moyen sur les autres langues : 0,2 UAS

Vitesse de convergence : au moins équivalente à max-violation

Divergence de Kullback-Leibler entre les distributions de vecteurs caractéristiques évalués :

Français	Classique	+ ND SO
Early update	0,350	0,280
Max-violation	0,357	0,277

Avec nos améliorations, les exemples d'apprentissage ressemblent davantage aux configurations rencontrées en test.

# Plan

- 1 État de l'art
  - Analyse en dépendances par transition
  - Oracle dynamique glouton
  - Recherche en faisceau et apprentissage global
  - Limites des méthodes d'apprentissage actuelles
- 2 Améliorations proposées
  - Oracle dynamique global
  - Reprise d'apprentissage
  - Références dans l'espace sous-optimal
- 3 Validation expérimentale
  - Cadre expérimental
  - Résultats
  - Analyse de similarité
- 4 Conclusions

# Conclusions

- Extension des stratégies classiques d'apprentissage : performances jamais dégradées, en moyenne 0,2 UAS meilleures, et surtout de meilleures propriétés durant l'apprentissage (et de nouvelles applications : [Lacroix *et al.*, 2016])
- Grande variabilité des algorithmes d'apprentissage : peu de combinaisons ont été explorées dans l'état de l'art
- Stratégies d'apprentissage souvent imprécises dans la littérature : nous proposons une formalisation commune des méthodes de l'état de l'art

Merci !

- Collins M. & Roark B. (2004). Incremental Parsing with the Perceptron Algorithm. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*.
- Goldberg Y. & Nivre J. (2012). A Dynamic Oracle for Arc-Eager Dependency Parsing. In *Proceedings of COLING 2012*, p. 959–976.
- Huang L., Fayong S. & Guo Y. (2012). Structured Perceptron with Inexact Search. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, p. 142–151.
- Lacroix O., Aufrant L., Wisniewski G. & Yvon F. (2016). Apprentissage d'analyseur en dépendances cross-lingue par projection partielle de dépendances. In *Actes de la 23e conférence sur le Traitement Automatique des Langues Naturelles*, p. 1–14.
- Zhang Y. & Nivre J. (2012). Analyzing the Effect of Global Learning and Beam-Search on Transition-Based Dependency Parsing. In *Proceedings of COLING 2012 : Posters*, p. 1391–1400.