

Hierarchical Phrase-Based Statistical Machine Translation System

Fethi BOUGARES

Maître de conférences
Laboratoire d'Informatique de l'Université du Maine

4 décembre 2020

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Motivations

SCFG : Synchronous Context Free Grammar

Système de traduction hiérarchique

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Translate a source sentence S (source) to target T

Translate \Rightarrow Find the best traduction among several hypothesis

$$\hat{t} = \operatorname{argmax}_{t \in T} P(t|s) = \operatorname{argmax}_{t \in T} P(s|t)P(t)$$

$P(s|t)$: Translation model

$P(t)$: language model

argmax : decoding algorithm

Translate a source sentence S (source) to target T

Translate \Rightarrow Find the best traduction among several hypothesis

$$\hat{t} = \operatorname{argmax}_{t \in T} P(t|s) = \operatorname{argmax}_{t \in T} P(s|t)P(t)$$

$P(s|t)$: *Hierarchical model*

$P(t)$: language model

argmax : decoding algorithm

Source : *A Hierarchical Phrase-Based Model for Statistical Machine Translation*, David Chiang

澳洲 是 与 北韩 有 邦交 的 少数 国家 之一 。
Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi .
Australia is with North Korea have dipl. rels. that few countries one of .

Australia is one of the few countries that have diplomatic relations with North Korea.

Limits of PBSMT (segment based translation) :

[Aozhou] [shi]₁ [yu Beihan]₂ [you] [bangjiao] [de shaoshu guojia zhiyi] [.]

[Australia] [has] [dipl. rels.] [with North Korea]₂ [is]₁ [one of the few countries] [.]

Incorrect translation because of the **distorsion limit**
(maximal distance of reordering) needed to avoid
computational explosion.

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Errors of translation related to the phrases reordering .

Solution : In order to reorder better

- ⇒ Replace words with sequences
- ⇒ Apply the reordering to sequences

Okay but how?

Example :

Using the following hierarchical sequences :

$$X \rightarrow \langle \text{yu } X_{[1]} \text{ you } X_{[2]}, \text{ have } X_{[2]} \text{ with } X_{[1]} \rangle \quad (6)$$

$$X \rightarrow \langle X_{[1]} \text{ de } X_{[2]}, \text{ the } X_{[2]} \text{ that } X_{[1]} \rangle \quad (7)$$

$$X \rightarrow \langle X_{[1]} \text{ zhiyi, one of } X_{[1]} \rangle \quad (8)$$

and the following conventional sequence of words :

$$X \rightarrow \langle \text{Aozhou, Australia} \rangle \quad (9)$$

$$X \rightarrow \langle \text{Beihan, North Korea} \rangle \quad (10)$$

$$X \rightarrow \langle \text{shi, is} \rangle \quad (11)$$

$$X \rightarrow \langle \text{bangjiao, diplomatic relations} \rangle \quad (12)$$

$$X \rightarrow \langle \text{shaoshu guojia, few countries} \rangle \quad (13)$$

and two rules :

$$S \rightarrow \langle S_{[1]} X_{[2]}, S_{[1]} X_{[2]} \rangle \quad (14)$$

$$S \rightarrow \langle X_{[1]}, X_{[1]} \rangle \quad (15)$$

Motivations

SCFG : Synchronous Context Free Grammar

Système de traduction hiérarchique

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

SynCFG is a kind of context free grammar that generates pair of strings.

Rules in a SynCFG are superficially similar to CFG rules :
→ except that they specify the structure of two phrases at the same time.

What is a grammar?

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

- ▶ A grammar consists of one or more variables that represent classes of strings
- ▶ There are rules that say how the strings in each class are constructed.
- ▶ Rules can use :
 1. symbols of the alphabet
 2. strings that are already known to be in one of the classes
 3. or both

Definition of Context-Free Grammar

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

A CFG, G is a tuple $G = (V, \Sigma, S, P)$ where

1. V is the (finite) set of variables (non terminals)
2. Σ is a finite set of terminals
3. S is the start symbol
4. P is a set of production rules, in the following form
$$V \rightarrow (V \cup \Sigma)^*$$

Each Production rules consists of :

1. A variable that is being defined by the production (called the head of the production)
2. the production symbol \rightarrow
3. A string of zero or more terminals and variables (called the body of the production)

Example of Context-Free Grammar

Syntactically correct algebraic expressions of x , y and z :

- ▶ Three terminal symbols " x ", " y " and " z "
- ▶ one nonterminal symbol S .
- ▶ The production rules are :

1. $S \rightarrow x$

2. $S \rightarrow y$

3. $S \rightarrow z$

4. $S \rightarrow S + S$

5. $S \rightarrow S - S$

6. $S \rightarrow S * S$

7. $S \rightarrow S / S$

8. $S \rightarrow (S)$

How this grammar can generate the string : $(x + y) * x - z$
 $* y / (x + x)$?

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Example of Context-Free Grammar

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Parenthesis matching :

- ▶ two terminal symbols "a" and "b"
- ▶ one nonterminal symbol S.
- ▶ The production rules are :

$$S \rightarrow aSb$$

$$S \rightarrow ab$$

$$S \rightarrow \epsilon$$

This grammar generates the language $\{ a^n b^n \mid n \geq 0 \}$

This grammar can, for example, generate the string :
"aaabbb"

$$S \rightarrow aSb \rightarrow aaSbb \rightarrow aaaSbbb \rightarrow aaabbb$$

Les structures élémentaires sont des règles de réécriture de la forme :

$$X \rightarrow \langle \gamma, \alpha, \sim \rangle$$

X est non terminal

γ et α sont des séquences d'éléments terminaux et non terminaux

\sim est fonction bijective (*one-to-one*) entre éléments non terminaux de γ et éléments non terminaux de α

Un exemple d'une grammaire hors contexte synchrones :

$r_0 : S \rightarrow \langle X_1, X_1 \rangle$ (glue rule)

$r_1 : X \rightarrow \langle X_0 \text{ of the } X_1, X_0 \text{ des } X_1 \rangle$

$r_2 : X \rightarrow \langle \text{activity, activités} \rangle$

$r_3 : X \rightarrow \langle \text{chambers, chambres} \rangle$

À partir de cette grammaire on pourra générer d'une manière synchrone deux phrases dans deux langues différentes :

$\langle X_1, X_1 \rangle$

$\langle X_2 \text{ of the } X_3, X_2 \text{ des } X_3 \rangle$

$\langle \text{activity of the } X_3, \text{activités des } X_3 \rangle$

$\langle \text{activity of the chambers, activités des chambres} \rangle$

Les indices indiquent la relation entre les non terminaux liés par \sim

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

La dérivation à partir d'une grammaire hors contexte synchrone commence à partir de la symbole S (*Start*).

Glue rules :

$$S \rightarrow \langle X_1, X_1 \rangle$$

$$S \rightarrow \langle S_1 X_2, S_1 X_2 \rangle$$

À chaque étape les deux éléments non terminaux sont ré-écrits avec les deux éléments d'une seule règle

\Rightarrow Application sur l'exemple introductif

Exemple

Exemple :

Avec ces séquences de mots hiérarchiques :

$$X \rightarrow \langle \text{yu } X_{[1]} \text{ you } X_{[2]}, \text{ have } X_{[2]} \text{ with } X_{[1]} \rangle \quad (6)$$

$$X \rightarrow \langle X_{[1]} \text{ de } X_{[2]}, \text{ the } X_{[2]} \text{ that } X_{[1]} \rangle \quad (7)$$

$$X \rightarrow \langle X_{[1]} \text{ zhiyi, one of } X_{[1]} \rangle \quad (8)$$

et des séquences de mots conventionnelles :

$$X \rightarrow \langle \text{Aozhou, Australia} \rangle \quad (9)$$

$$X \rightarrow \langle \text{Beihan, North Korea} \rangle \quad (10)$$

$$X \rightarrow \langle \text{shi, is} \rangle \quad (11)$$

$$X \rightarrow \langle \text{bangjiao, diplomatic relations} \rangle \quad (12)$$

$$X \rightarrow \langle \text{shaoshu guojia, few countries} \rangle \quad (13)$$

et deux règles de production de départ :

$$S \rightarrow \langle S_{[1]} X_{[2]}, S_{[1]} X_{[2]} \rangle \quad (14)$$

$$S \rightarrow \langle X_{[1]}, X_{[1]} \rangle \quad (15)$$

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Exemple

$\langle S_1, S_1 \rangle$

$\xRightarrow{(14)} \langle S_2 X_3, S_2 X_3 \rangle$

$\xRightarrow{(14)} \langle S_4 X_5 X_3, S_4 X_5 X_3 \rangle$

$\xRightarrow{(15)} \langle X_6 X_5 X_3, X_6 X_5 X_3 \rangle$

$\xRightarrow{(9)} \langle \text{Aozhou } X_5 X_3, \text{Australia } X_5 X_3 \rangle$

$\xRightarrow{(11)} \langle \text{Aozhou shi } X_3, \text{Australia is } X_3 \rangle$

$\xRightarrow{(8)} \langle \text{Aozhou shi } X_7 \text{ zhiyi, Australia is one of } X_7 \rangle$

$\xRightarrow{(7)} \langle \text{Aozhou shi } X_8 \text{ de } X_9 \text{ zhiyi, Australia is one of the } X_9 \text{ that } X_8 \rangle$

$\xRightarrow{(6)} \langle \text{Aozhou shi yu } X_1 \text{ you } X_2 \text{ de } X_9 \text{ zhiyi,}$
Australia is one of the X_9 that have X_2 with $X_1 \rangle$

$\xRightarrow{(10)} \langle \text{Aozhou shi yu Beihan you } X_2 \text{ de } X_9 \text{ zhiyi,}$
Australia is one of the X_9 that have X_2 with North Korea \rangle

$\xRightarrow{(12)} \langle \text{Aozhou shi yu Beihan you bangjiao de } X_9 \text{ zhiyi,}$
Australia is one of the X_9 that have diplomatic relations with North Korea \rangle

$\xRightarrow{(13)} \langle \text{Aozhou shi yu Beihan you bangjiao de shaoshu guojia zhiyi,}$
Australia is one of the few countries that have diplomatic relations with North Korea \rangle

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Motivations

SCFG : Synchronous Context Free Grammar

Système de traduction hiérarchique

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Un système de Traduction :

- ▶ Une table de traduction;
- ▶ Un modèle de langage;
- ▶ Et un algorithme de décodage

Table de traduction hiérarchique : apprentissage

- ▶ Extraction des règles sans annotation syntaxique
- ▶ Utilisation d'alignements mot par mot
- ▶ Première étape : définition des séquences de mots initiales comme dans un PBSMT
- ▶ Deuxième étape : extraction des règles : remplacer des sous-phrases par des symboles non-terminaux

Système hiérarchique

Première étape

Il doit y avoir **au moins un mot d'une séquence aligné** avec **un mot dans l'autre séquence**, et aucun mot dans une séquence doit être aligné avec un mot en dehors de l'autre séquence

	friendly	cooperation	over	the	last	30	years
30						■	
duonianlai			■	■	■		■
de							
youhao	■						
hezuo		■					

	friendly	cooperation	over	the	last	30	years
30						■	
duonianlai			■	■	■		■
de							
youhao	■						
hezuo		■					

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Système hiérarchique

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Deuxième étape

On cherche ensuite des séquences de mots qui sont contenues dans d'autres séquences de mots et on les remplace par des non terminaux

	friendly	cooperation	over	the	last	30	years
30							
duonianlai							
de							
youhao							
hezuo							

	X_2	over	the	last	X_1	years
X_1						
duonianlai						
de						
X_2						

L'ensemble des règles de $X \rightarrow \langle \gamma, \alpha, \sim \rangle$ est la plus petite ensemble qui satisfait :

1. Si $\langle \gamma_i^j, \alpha_i^{j'}, \sim \rangle$ est une paire de phrase, alors

$$X \rightarrow \langle \gamma_i^j, \alpha_i^{j'}, \sim \rangle$$

est une règle de $\langle \gamma, \alpha, \sim \rangle$

2. Si $X \rightarrow \langle \gamma_i^j, \alpha_i^{j'}, \sim \rangle$ est une règle de la grammaire telque : $\langle \gamma_1 \gamma_i^j \gamma_2, \alpha_1 \alpha_i^{j'} \alpha_2, \sim \rangle$, alors

$$X \rightarrow \langle \gamma_1 X_k \gamma_2, \alpha_1 X_k \alpha_2, \sim \rangle \text{ est ajouté à } \langle \gamma, \alpha, \sim \rangle$$

Le processus précédent génère un très grand nombre de règles.

Un très grand nombre de règles est indésirable :

1. apprentissage très long ;
2. décodage très long ;
3. crée une ambiguïté ;
4. plusieurs dérivation génère la même traduction ;
5. n-best list avec peu de traduction distincte

Afin de limiter le nombre de règles extraites, des heuristiques sont appliquées et la grammaire est filtrée :

1. Les séquences initiales sont limités à 10.
2. Le nombre de non terminaux et terminaux de la langue source est limité à 5
3. les non terminaux adjacents dans la langue source sont interdits
4. ...

En plus des *Entity rules* sont ajoutées :

Règles permettant de modéliser des dates, des nombres

→ Peuvent être créées manuellement

Utilisation des *fonctions caractéristiques* analogue aux fonctions (*features*) de PBSMT.

Les fonctions extraites de la table de traduction sont :

- ▶ $P(\gamma|\alpha)$ et $P(\alpha|\gamma)$ probabilité de traduction des séquences (trad inv)
- ▶ $P_w(\gamma|\alpha)$ et $P_w(\alpha|\gamma)$ qui estime la qualité de la traduction α de γ
- ▶ une pénalité pour chaque utilisation de règle afin de choisir entre de courtes ou de longues dérivations

⇒ Ces scores sont combinés avec une l'optimisation par "minimization error rate" ou "Max BLEU"

Système hiérarchique : Décodage

Algorithme d'analyse grammaticale :

Top-down vs. bottom-up

Top-down (goal-driven) : à partir de symbole de début

bottom-up (data-driven) : à partir des règles

Naive vs. dynamic programming :

Naive : énumérer toute les possibilités

Backtracking : essayer et élaguer les mauvaises solutions

Programmation Dynamique : enregistrer les solutions partielles dans un tableau

CKY * : \rightarrow bottom-up ; Programmation Dynamique

* Cock-Younger-Kasami

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

CKY * : Exemple

On considère la grammaire suivante :

$$S \rightarrow \epsilon | AB | XB$$

$$T \rightarrow AB | XB$$

$$X \rightarrow AT$$

$$A \rightarrow a$$

$$B \rightarrow b$$

Question : Est-ce-que $w = aaabb$ est dans $L(G)$?

Système hiérarchique : *Parsing*

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

CKY est un algorithme *Bottom-up* (de bas en haut)

$$S \rightarrow \epsilon \mid AB \mid XB$$

$$T \rightarrow AB \mid XB$$

$$X \rightarrow AT$$

$$A \rightarrow a$$

$$B \rightarrow b$$

 $a \ a \ a \ b \ b$

Système hiérarchique : *Parsing*

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

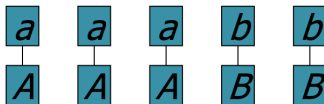
$$S \rightarrow \epsilon \mid AB \mid XB$$

$$T \rightarrow AB \mid XB$$

$$X \rightarrow AT$$

$$A \rightarrow a$$

$$B \rightarrow b$$



Système hiérarchique : *Parsing*

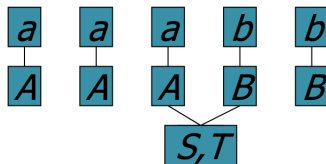
$$S \rightarrow \epsilon \mid AB \mid XB$$

$$T \rightarrow AB \mid XB$$

$$X \rightarrow AT$$

$$A \rightarrow a$$

$$B \rightarrow b$$



Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Système hiérarchique : *Parsing*

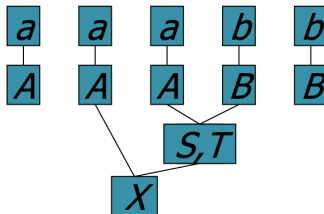
$$S \rightarrow \epsilon \mid AB \mid XB$$

$$T \rightarrow AB \mid XB$$

$$X \rightarrow AT$$

$$A \rightarrow a$$

$$B \rightarrow b$$



Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Système hiérarchique : *Parsing*

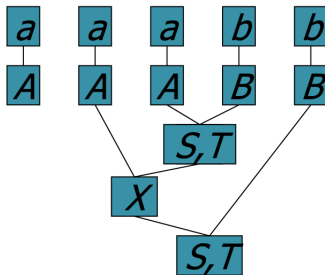
$S \rightarrow \epsilon \mid AB \mid XB$

$T \rightarrow AB \mid XB$

$X \rightarrow AT$

$A \rightarrow a$

$B \rightarrow b$



Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Système hiérarchique : *Parsing*

Vérifions pour aaabbb

$$S \rightarrow \epsilon \mid AB \mid XB$$

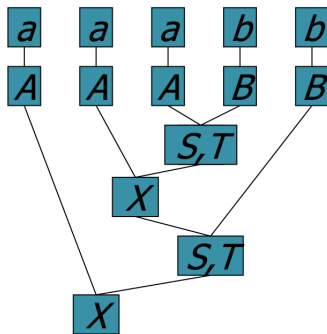
$$T \rightarrow AB \mid XB$$

$$X \rightarrow AT$$

$$A \rightarrow a$$

$$B \rightarrow b$$

aaabbb est rejetée



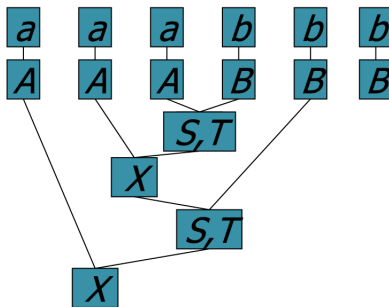
Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Systeme hiérarchique : *Parsing*

Et si $w = aqabbbb$

$$S \rightarrow \epsilon \mid AB \mid XB$$
$$T \rightarrow AB|XB$$
$$X \rightarrow AT$$
$$A \rightarrow a$$
$$B \rightarrow b$$


Système hiérarchique : *Parsing*

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

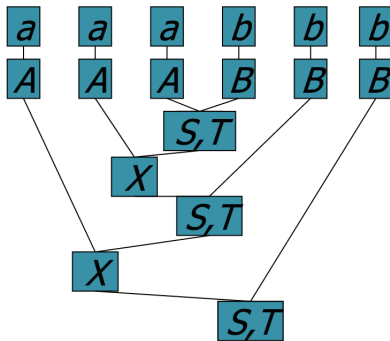
$S \rightarrow \epsilon \mid AB \mid XB$

$T \rightarrow AB \mid XB$

$X \rightarrow AT$

$A \rightarrow a$

$B \rightarrow b$



S est dans la racine $\Rightarrow aaabbb$ est accepté

Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Construction de la traduction durant le *Parsing*

Exemple (source *Chris Dyer @ CMU*) : (Japonais - Anglais)

S → NP VP : [1] [2] (monotonic)

VP → NP V : [2] [1] (inverted)

V → tabeta : *ate*

NP → jon-ga : *John*

NP → ringo-o : *an apple*

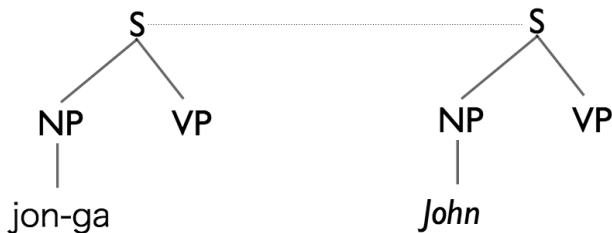
Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Construction de la traduction durant le *Parsing*



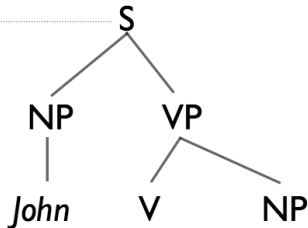
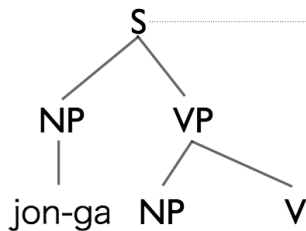
Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Construction de la traduction durant le *Parsing*



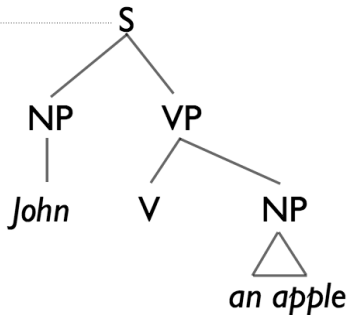
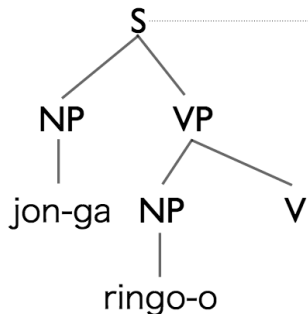
Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Construction de la traduction durant le *Parsing*



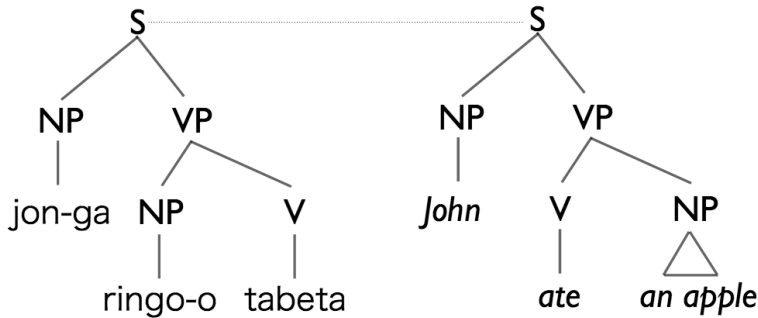
Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Construction de la traduction durant le *Parsing*



Traduction : (Jon-ga ringo-o tabeta; John ate an apple)

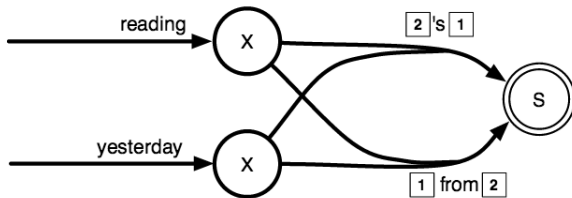
Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Construction de la traduction durant le *Parsing*
Question : Comment intégrer le LM?



(yesterday 's reading; reading of yesterday)

Comment on peut ajouter le score LM à chaque
hypothèse partielle de traduction (durant le parsing)?

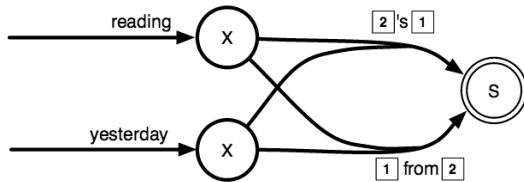
Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Question : Pourquoi c'est compliqué ?



1. On connaît pas le contenu des variables 1 et 2;
2. On connaît pas le context (à gauche).

⇒ Solution naïve :

1. Extraire tous les k-best traduction
2. Calculer le score LM de chaque phrase
3. Ordonner les phrases

Système hiérarchique : *Parsing* et traduction

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

⇒ Solution :

1. Restructurer le hyper-graphe de décodage
2. Sauvegarder dans chaque noeud suffisamment d'information pour le LM
3. Calculer le score LM
4. Recherche exhaustive impossible → élagage des hypothèses partielles

Système hiérarchique : Implémentations

Motivations

SCFG :
Synchronous
Context Free
Grammar

Système de
traduction
hiérarchique

Joshua : Joshua est un décodeur (LGPL) écrit en Java, implémentant l'approche par séquences hiérarchiques

Il est accompagné de l'ensemble des outils nécessaire à son fonctionnement :

- ▶ apprentissage des règles ;
- ▶ décodeur ;
- ▶ optimisation des poids ;
- ▶ minimisation d'erreur

Moses - Hiero : Extension de moteur de traduction Moses

[http://www.statmt.org/moses/?n=Moses.
SyntaxTutorial](http://www.statmt.org/moses/?n=Moses.SyntaxTutorial)

Résultats sur dev de IWSLT 09 en BLEU (Expériences réalisées par Y. Estève - LIUM) :

- ▶ Moses (PB) : 53,41
- ▶ Joshua : 54,00
- ▶ Moses (PB) avec CSLM 54,75
- ▶ Moses (PB) + Joshua : 55,54

Moses - Hiero : Extension de moteur de traduction Moses

[http://www.statmt.org/moses/?n=Moses.
SyntaxTutorial](http://www.statmt.org/moses/?n=Moses.SyntaxTutorial)

Hierarchical Phrase-Based Translation. *David Chiang*
University of Southern California, ACL 2007

Joshua : An Open Source Toolkit for Parsing-based
Machine Translation. *Zhifei Li, Chris Callison-Burch, Chris
Dyer, Juri Ganitkevitch, Sanjeev Khudanpur, Lane Schwartz,
Wren N. G. Thornton, Jonathan Weese and Omar F. Zaidan,*
WMT 2009